Volume 3
Spring 2004

SURJ

Stanford Undergraduate Research Journal

SURJ.

STANFORD UNDERGRADUATE RESEARCH JOURNAL

Volume 3, Spring 2004

http://surj.stanford.edu

Stanford, California 94309

Journal Staff

The academic year of 2003-2004 marks the third issue of the *Stanford Undergraduate Research Journal (SURJ)*. Spotlighting exemplary undergraduate research and spanning multiple academic disciplines, *SURJ* strives to publicize the achievements of Stanford undergraduates, to facilitate the dissemination of information, and to connect students of all majors to the worlds of academia that exist beyond the classroom. This year, the Journal aims to take its objective to a broader and deeper level through its support and implementation of open access research publication and through the exchange of knowledge that inevitably follows.

The papers in this issue are a testament to the rigor of Stanford undergraduate research and to the pioneering spirit that fuels it. In addition, the newly expanded Special Features section showcases some of Stanford's efforts to expand the public's access to research findings. For instance, one Special Features article documents the growth of Stanford's Public Library of Science (PLoS), an organization dedicated to publishing scientific articles online with no charge or restrictions to subsequent use and redistribution. Meanwhile, as discussed in another Special Features article, the Stanford-founded HOPES website presents research on Huntington's Disease to families in over 47 countries who normally cannot access it fully. The time is ripe, both at Stanford and in America, for the push towards the free exchange of research findings. This year, as *SURJ* continues to champion the undergraduate's contributions to knowledge, it also joins in the effort to proliferate that knowledge.

Accordingly, *SURJ* is publishing its articles under a Creative Commons license, which enables one to freely copy and distribute the articles as long as the original author is attributed. Moreover, *SURJ* will distribute approximately 50% more print copies this year to local and national locations. While *SURJ* is already freely available online and widely distributed throughout the Stanford campus, the staff this year will also be working with Stanford administrators, professors, libraries, and the Office of Admissions to distribute copies of *SURJ* to research-oriented high school students, alumni, and academics across the nation. Inherent in the dissemination and exchange of research is the need to insure the integrity of research methodologies, findings, and reporting. Therefore, *SURJ* invited Stanford faculty to comment on the work of student editors this year. The input of faculty editors provided a double-check on the staff's decisions while still allowing the Journal to remain a peer-reviewed publication.

This year has marked the advent of several new developments in the communication and exchange of research. We at *SURJ* are excited to be a part of these developments and are continually striving to improve our efforts. We hope that the content of this issue, along with its far-reaching distribution, will vitalize interest in Stanford's undergraduate research and inspire others to take on research projects of their own.

On behalf of the 2003-2004 *SURJ* staff and authors, we would like to extend a special thank you to Vice Provost John Bravman and President John Hennessy, who contributed ideas and a great deal of support to our cause. Lastly, we thank you for your readership and hope that you find the articles in this issue as enduring as we do.

Sincerely,

Liz Williams Grace Wu

Elizabeth K. Williams & Grace W. Wu
Editors-in-Chief, 2003-2004

SURJ. Table of Contents

Computer Science & Linguistics

Building a Regionally Inclusive Dictionary for Speech Recognition

Speech Recognition (SR) is the automated conversion of speech into written text. Applications range from simple phone-based information services to commercial-grade automated customer service systems, such as those for airline phone reservations. While this process is complex in and of itself, it is further complicated by the fact that speakers (the users) from different parts of the country have varying accents and pronounce the same words differently. Our aim is to create a more speaker-independent SR system while maintaining speed and accuracy of transcription. This requires the construction of an SR dictionary that takes into account the existence of multiple pronunciations for the same word. However, the existence of too many alternate pronunciations overloads the system and is detrimental to accuracy and speed. By finding the optimal number of pronunciations per word, the percentage of words correctly identified by the SR system increased from 78% using the traditional technique to 85.7% using the improved method outlined in this article. This brings the technology 35% closer to the goal of complete recognition and the use of speech as the primary method of human-computer interaction.

Justin Burdick

Speech Recognition (SR) is a process that transcribes speech into text using a computer. Many repetitive phone tasks can be automated with speech recognition technology, saving businesses significant amounts of money. However, the transcription of everyday human speech is significantly more difficult than simply the recognition of a small set of words, as is the case in a SR-based phone information-retrieval system (e.g. 411 services).

A complete Speech Recognition system is fairly complex and consists of various subsystems that interact together to convert speech into written text, as shown in Figure 1. A typical SR system utilizes a common tool for pattern-matching called Hidden Markov Models (HMM) to model parts of speech known as “phonemes” (similar to syllables) [1]. The models are first trained using a set of training data, which consists of a large number of wave sound files with speech and the corresponding text transcriptions. Each speech file is converted into a series of observations known as “feature vectors." These are sets of numbers which describe the sound mathematically by extracting specific information from the waveform at 25-millisecond intervals. The computer then matches this numerical representation to the corresponding text transcription included in the training data. Applied over the entire training set, this process creates a series of models which later allow the SR system to translate speech into text. The training process results in the creation of two data files: one consisting of a model of properties for each phoneme and another of a list of pronunciations for each word. Phonemes, the most fundamental units of speech, are single sounds; each phoneme is modeled by its average sound, the variation of that sound across various speakers, and the transition probabilities between it and other phonemes. The list of pronunciations, termed the “pronouncing dictionary,” contains each of the words used during training along with the sequence of phonemes that compose that word.

Once the models are trained, the system can be used to perform recognition. In order to produce meaningful sentences, the SR system relies on a dictionary and a grammar model as well as the trained phoneme models from the training phase. Transcription is accomplished using the Viterbi algorithm. This algorithm divides the incoming speech into individual “observations” and processes one observation at a time by comparing it to every possible phoneme. When the algorithm moves to the next observation, it eliminates paths that have a low matching probability. By this elimination at each observation, only the most probable paths survive until the end of the observation sequence; the recognized word sequence can then be generated. The dictionary is important in this process. As the system processes each observation, it must check the dictionary to see if it has made a word yet. Once the word is recognized, the system uses the grammar file to influence the next word chosen, which is very valuable for making the sentence meaningful, similar to grammar check in a word-processing program.

Implementing an effective SR system requires the synthesis of concepts

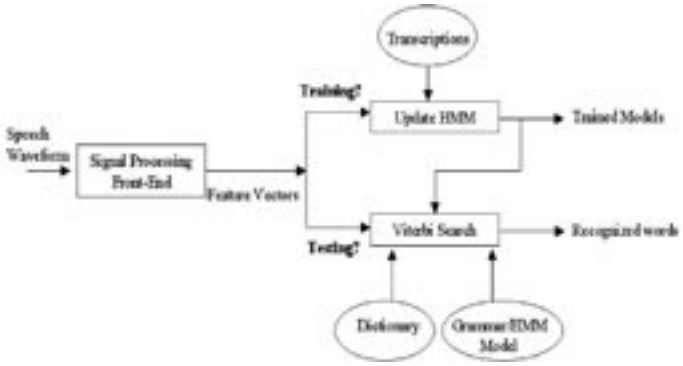


Figure 1. A simplified model of an HMM-based speech recognition system

from several fields, including:

- Signal Processing (for extracting important features from the sound waves)
- Probability and Statistics (for defining and training the models and recognition)
- Linguistics (for building the dictionary and the grammar model)
- Computer Science (for creating efficient search algorithms)

Improvements in any of the subsystems above could result in better overall performance. The following performance goals are among the main objectives of any SR system [2]:

1. Increased Accuracy

Accuracy is measured as a percentage of words (or sentences) that are correctly detected. To emphasize the importance of this value, consider a certain SR system that has an accuracy of 99%. This means that during the task of dictation, one word in every 100 words is identified incorrectly. For the average dictation, such an error rate would result in approximately six errors per single-spaced page. Searching for and correcting these typos is a tedious and time-consuming task. This shows that even an accuracy of 99% may not be sufficiently high for the task of dictation.

2. Increased Speed

Many speech recognition applications must run on small hand-held devices with limited CPU power. It is important for the SR system to be as computationally inexpensive as possible. This allows the system to transcribe speech in real-time, even on a low-performance system like a PDA or cell phone.

3. Speaker Independence

Most SR systems can operate with very high accuracy if they are trained to a specific speaker. However, problems can arise when the system attempts to transcribe the speech of a different person. The reasons for this include, but are not limited to:

a. Different speakers may pronounce the same phonemes (sub-words) differently, e.g. a speaker from Brooklyn may pronounce certain vowels differently than a speaker from California.
b. Different speakers may pronounce the same words differently, with a different sequence of phonemes, e.g. the word “maybe” could be pronounced as “m ey b iy” or as “m ey v iy.”

In this project, our goal is to improve recognition of speaker independent SR systems by taking into account that different speakers may pronounce similar words differently. This requires building a dictionary that has multiple pronunciations for any given word. The details of building such a dictionary are explained below. In order to test the effects of this approach, a set of carefully designed experiments were conducted, the results of which are in the Results and Discussion section.

Building the Dictionary

The TIMIT (Texas Instruments-Massachusetts Institute of Technology) database was used for both training and testing of the SR system. This database contains audio files of sample sentences as well as word-level and phoneme-level transcriptions. However, we needed to create our own "pronouncing dictionary" before we could do training and recognition. Each word in a pronouncing dictionary is associated with a series of phonemes that approximate the sound of that word. However, some words need multiple pronunciations in order to be best represented (see Table 1 for examples). When this sequence of phonemes is pronounced, it accurately approximates the sound of the word being spoken. All of the TIMIT data was used in the experiments. About 70% was used to train the models, and the rest was used for performing the testing. Table 2 summarizes key characteristics of the TIMIT data.

Method limitations

In order to provide a versatile training set, data was collected from speakers spanning eight different geographical areas to capture regional dialects [4] (shown in Figure 2). Unfortunately, these regional dialects introduce problems for the construction of a dictionary. Since many of the speakers pronounce words differently—especially in the case of short, common words—multiple transcriptions for the same word were common. In fact, the

Word	**Phoneme-level Transcription**
As	ae z
	ax z
Coauthors	kcl k ow ao th axr z
Coffee	k aa iy f
	k ao f iy

Table 1. *Sample dictionary entries. Multiple pronunciations can exist for the same word. For example, the first pronunciation of coffee is that of a typical New Yorker, while the second is that of someone from the old Northwest.*

Number of...	
Speakers	630 (70% male, 30% female)
Utterances	6300 (10 per speaker)
Distinct texts	2342
Words	6099
Utterances in the training set	4620
Utterances in the test set	1680
Male speakers	326 (training set) + 112 (test set)
Female speakers	136 (training set) + 56 (test set)

Table 2. *Key characteristics of the TIMIT data*

unedited dictionary contained an average of two pronunciations per word [1]. However, these pronunciations tended to cluster around certain words, with most words having only one pronunciation but with others having six or more. A certain level of multiple pronunciations can be helpful at capturing common divergences in pronunciation (e.g. potato/pot_to), as even the best hand-made dictionaries contain 1.1 to 1.2 pronunciations per word [5]. Too many pronunciations per word, however, could be misleading when the computer performs recognition, because a poorly recognized string of phonemes could cause a word mismatch that makes a sentence meaningless.

Methodology

In order to reduce the number of redundant definitions in the dictionary, two methods were investigated. The first, called "skimming," simply edits out and deletes low-frequency transcriptions. This method defined a threshold based on the most commonly occurring pronunciations. Anything below this threshold was removed from the dictionary. The second method, called "percentaging," encodes the frequency at which each pronunciation was encountered into the definitions themselves, which a speech recognizer can use to modify its recognition network.

As mentioned earlier, the dictionary was created by examining all the speech transcriptions during training mode. The original system just added all of the different combinations of word and phonemes into the system. The new method does the same thing, except that it keeps count of how many times each combination occurs. These counts can be used to perform either skimming (removal of low-occurrence pairs), or percentaging (placing percentage occurrence information with the data). This paper examines the effect of skimming alone, percentaging alone, and finally skimming followed by percentaging on SR system accuracy.

Results and Discussion

The two systems of dictionary making (skimming and percentaging) were tested independently and then with the 2-step method (skimming followed by percentaging). As shown in Figure 4, skimming examples of dictionaries after processing provided a much larger accuracy gain than did percentaging. When the two were used simultaneously, percentaging added a very slim increase in accuracy to the method of skimming. An unexpected bonus of using the skimming method is a reduction in required recognition time. Since there are fewer possible recognition sequences, the computer can process each speech fragment more rapidly. At optimal skimming (30%), the test took only two hours, rather than the full three, which is a 33% reduction in time. Furthermore, at this level of skimming, there was an average of 1.6 pronunciations per word, as opposed to 2.0 pronunciations without skimming. Thus, not only does skimming provide *more accurate* speech recognition, it also allows for faster recognition.

Conclusion

Since people speak in many differ-ent ways, it is expected that having a large dictionary, inclusive of *all* pronunciation and dialectic differences, would be a valuable aid to recognition. However, this paper demonstrates that the most effective SR system dictionary may be one with a small number of select alternate pronunciations per word. The dictionary must be built so that a balance is found between being inclusive of pronunciation variants and of conforming to the limitations of a computerized recognition system. This study observed that while a typical SR system contains on average 2.0 pronunciations per word, maximum

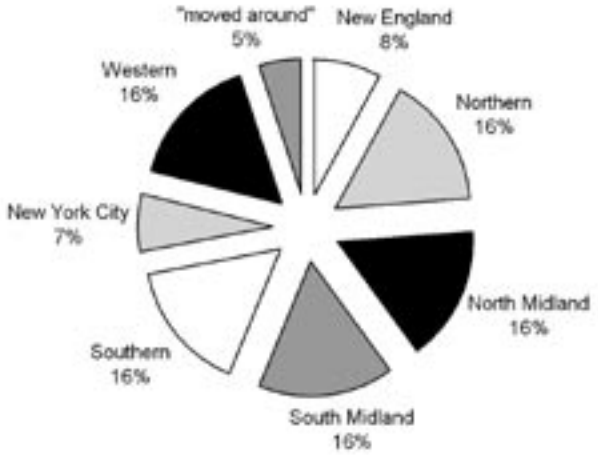


Figure 2: *Speaker Breakdown by Region (total number of speakers= 630).*

SR system accuracy is achieved at 1.6 pronunciations per word. Tailoring a dictionary to contain fewer pronunciations per word is best achieved manually; however, given the size of a typical SR system dictionary, (13,000 entries) a computerized method of deleting erroneous and infrequently used pronunciations is more practical. This study revealed that the method of skimming both effectively increases recognition accuracy and results in a 33% reduction in the time required for speech recognition tasks.

An example of an unaltered dictionary

Clearly, "about ax b aw t" is the most commonly given pronunciation for the word about, and so it should be given the highest weight, whereas definitions "bah" and "baw" should just be erased.

Original

1 about ax b ae t
2 about ax b ah
4 about ax b aw d
17 about ax b aw t
1 about b ah
1 about b aw

Examples of dictionaries after processing

After Skimming

4 about ax b aw d
17 about ax b aw t

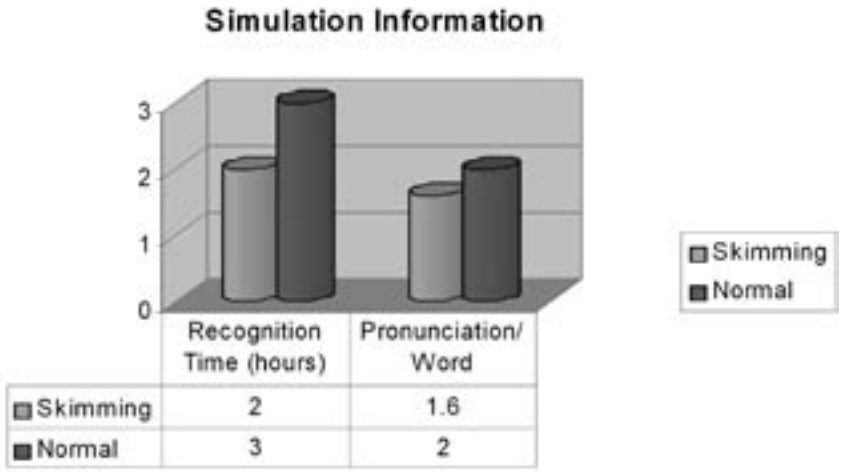
Here we have done skimming, deleting the least frequently occurring pronunciations.

After Percentaging

3.85% about ax b ae t
7.69% about ax b ah
15.38% about ax b aw d
65.38% about ax b aw t
3.85% about b ah
3.85% about b aw

Here we have done percentaging, inserting the frequency of each pronunciation as a percentage.

Figure 3: *The process of modifying the dictionary*



	Recognition Time (hours)	Pronunciation/ Word
Skimming	2	1.6
Normal	3	2

Table 3. *Simulation Test Results*

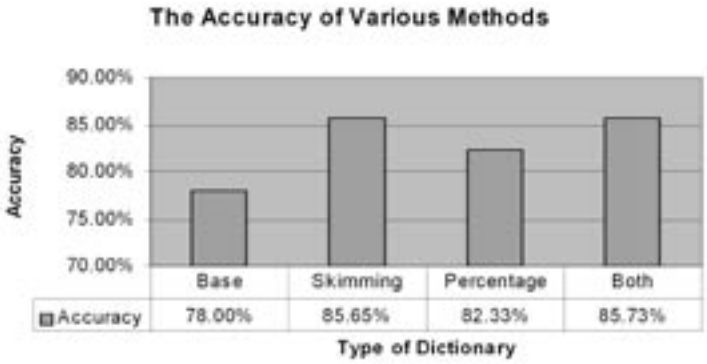


Figure 4. *Test Results*

References

1. Mohajer, K., Zhong-Min Hu, Pratt, V. Time Boundary Assisted speech recognition. *International Journal of Information Fusion* (Special Issue on Multi-Sensor Information Fusion For Speech Processing Applications).
2. Rabiner, Juang. *Fundamentals of Speech Recognition*, Chapter 6, Prentice Hall Signal Processing Series, New Jersey, 1993.
3. Young, Evermann, Kershaw, Moore, Odell, Ollason, Valtchev, Woodland. *The HTK Book*, Ch5, Cambridge University Engineering Department, 2001-2002.
4. Garofolo, Lamel, Fisher, Fiscus, Pallett, Dahlgren. TIMIT Printed Documentation, Ch5, U.S. Depart. Of Commerce, 1993.
5. Hain, T., Woodland, P.C., Evermann, G. et al. New features in the CU-HTK system for transcription of conversational telephone speech. *Acoustics, Speech, and Signal Processing*, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on, Vol.1, Iss., 2001. Pages:57-60 vol.1.
6. Culicover, Peter. Lecture: An Introduction to Language in the Humanities. Spring 2002. http://www.ling.ohio-state.edu/~culicove/H201/Language variation.pdf.

Justy Burdick

Justy Burdick is a sophomore majoring in Electrical Engineering and is considering a minor in Computer Science.

He would like to thank Professor Vaughan Pratt and the URP office for sponsoring this Research Experience for Undergraduates project. Furthermore, he would like to thank Keyvan Mohajer for the guidance and help he received at every step of the research process. Finally, he would like to thank the 2002 and 2003 speech group REU team members: Simon Hu, Melissa Mansur, Ryan Bickerstaff, Kenneth Lee, Gu Pan Grace.

Language Acquisition

The Search for the Object Begins at the Verb: Rapid Inferential Word-Learning by Two-year-old Children

Research on early word learning has focused primarily on how infants learn words through direct labeling. But as children grow older, most of their word learning becomes indirect as they begin to use contextual cues to infer the meanings of new words. This study examined how infants use information from a familiar verb to infer the referent of a novel noun. The goal was to teach children the names for two new objects that were never labeled directly. By monitoring children's eye movements we tracked the time course of orientation to the novel object that corresponded to each of the verbs. Within milliseconds of hearing the verb - and before hearing the novel noun - children inferred which picture was being talked about and oriented themselves toward it. Moreover, children learned and remembered the names for the novel objects. When tested later in sentences without cues from the verb, children reliably looked at the correct picture. This is the first study to show that inferential learning of new words by young children happens very rapidly as the sentence unfolds.

Tiffany Early, Ana Luz Portillo, Carolyn Quam

As adults we take learning words for granted. However, for young children acquiring their native language it is a difficult task. A simple example illustrates the complexity of the word learning issue. Suppose a British businesswoman travels to Italy, having forgotten her pocket translator. While seated in a restaurant, she overhears someone at a nearby table say *Limoncella!* Looking around, she sees people enjoying their meals, surrounded by objects like food, utensils, tables, and chairs. How can she figure out what *limoncella* refers to? In essence, this is the "word-learning problem." How can children listening to a stream of language make the correct sound-meaning correspondences? One possibility is that children use their existing linguistic knowledge to infer the meanings of new words. For example, assume our British businesswoman's bilingual friend then turns to her and comments *She's drinking limoncella.* Now when the businesswoman looks at the neighboring table and sees a woman drinking from a glass, she can use her knowledge of the verb "drink" to infer that "limoncella" probably refers to the beverage. As children begin to acquire new vocabulary at a faster rate around the age of two years, they must also use existing linguistic knowledge in a similar way to figure out the meanings of novel words.

In this study we investigated whether or not 28- to 30-month olds could learn the names of unfamiliar objects using their knowledge of familiar verbs and cues from the visual context.

A previous study by Goodman, McDonough, and Brown (1998) showed that 24- and 30-month-old children can use information from a familiar verb to learn the name for an unfamiliar object. In their picture book task, the children heard sentences such as *Mommy feeds the ferret. Show me the ferret*. as they looked at pictures of four novel objects. In this example, only one of the pictures was animate. Goodman et al. predicted that if children understood that the verb "feed" requires an animate recipient, they should infer that the novel word "ferret" refers to the animate picture, since none of the three inanimate distracter pictures was an appropriate match. They found that children used semantic information from the familiar verbs to identify the correct novel pictures.

While a picture book procedure can be used to demonstrate word learning, it offers little insight about how quickly children make such inferences. The experimental method used in the present study allowed us to pinpoint exactly when children made the connection between the novel word and its referent. We examined how children form sound-meaning correspondences as speech unfolds in a looking-while-listening task. This procedure let us identify at which point in the sentence children oriented themselves toward the correct picture.

Methods

Participants

We tested 37 children (18 males and 19 females) from 28 to 30 months of age. Subjects were healthy, full-term babies recruited through Stanford Hospital.

Procedure

During testing, the participant sat on her parent's lap in the center of the booth. A curtain was brought down in between parent and child so the parent would not bias the child's looking patterns. A video camera between the two monitors recorded the participant's eye movements. On each trial, two pictures appeared simultaneously on adjacent computer monitors (see Figure 1). Children saw the pictures for approximately four seconds and heard a natural, pre-recorded, female voice from the loudspeaker.

Experimental Design

Children participated in four verb comprehension trials, eight teaching trials, twelve testing trials, and seventeen filler trials. The filler trials contained only familiar pictures and were interspersed throughout the study to keep children engaged. Each child saw a total of 41 trials over a period of approximately six minutes.

The study began with four verb comprehension trials intended to confirm that children knew the verbs *drive* and *eat* and to familiarize them with our experimental task. The children were presented with pictures of familiar vehicles and foods and asked, *Which one do you drive?* or *Which one do you eat?*

The next section consisted of teaching trials in which children were introduced to two novel objects – the manju (a Japanese pastry) and the tempo (an Asian vehicle). In half of the trials, they saw a manju (the target picture) paired with a novel plastic fan (the distracter picture) and heard *You can eat the manju. Eat the manju*. We provided contextual cues such as a table, plate, and spoon to indicate that the manju was edible. In the other half of the trials, they saw a tempo (the target) paired with a novel rubber brush (the distracter) and heard *You can drive the tempo. Drive the tempo.* The tempo was depicted on a road to suggest that it was a drivable object. All four objects presented in the teaching trials were completely unfamiliar to the children, and neither of the target objects was explicitly paired with the novel word. Thus children had to infer from the verb information and visual context which of the novel objects the new word most likely referred to.

On testing trials the tempo and the manju were paired with each other, both on gray backgrounds to eliminate visual context as a cue. On these trials, children were asked *Where's the tempo?* or *Where's the manju?* with no preceding verbs to provide semantic cues.

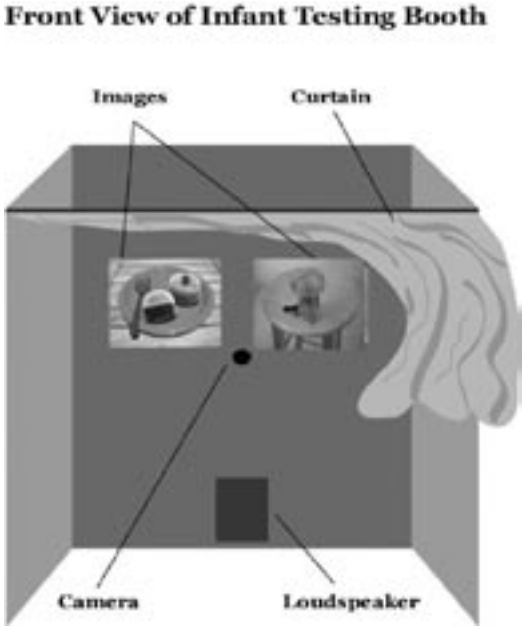


Figure 1. Experimental setup.

Stimuli

The visual stimuli were photographs taken with a 640x480 resolution camera and edited in Photoshop (to make the objects appear natural and to balance for salience). To counterbalance for target location, each target appeared equally often on the left and right monitors across trials.

The auditory stimuli were pre-recorded in a natural, female voice appealing to children. Recording was done using Peak 2.6 LE software with a digital microphone. We matched the target verbs and nouns, respectively, for duration and naturally stressed them relative to the other words in the sentences. Thus, children heard: *You can EAT the MANJU. EAT the MANJU; You can DRIVE the TEMPO. DRIVE the TEMPO; Where's the TEMPO;* and *Where's the MANJU.*

Coding

Trained coders analyzed the video recordings frame by frame with 33 millisecond resolution, noting whether the child's eyes were left, right, in between, or off task on each frame. Using custom software, coders marked picture and sound onset for each trial to align the eye movement data with the time course of the speech signals. All coders were blind to the

Figure 2. In testing trials, children saw the tempo and manju pictures paired together. After hearing the noun, without linguistic or visual contextual cues, children oriented to the correct picture with about 80% accuracy. The black vertical line corresponds to the end of the noun, and the black horizontal line at 0.5 indicates chance performance.

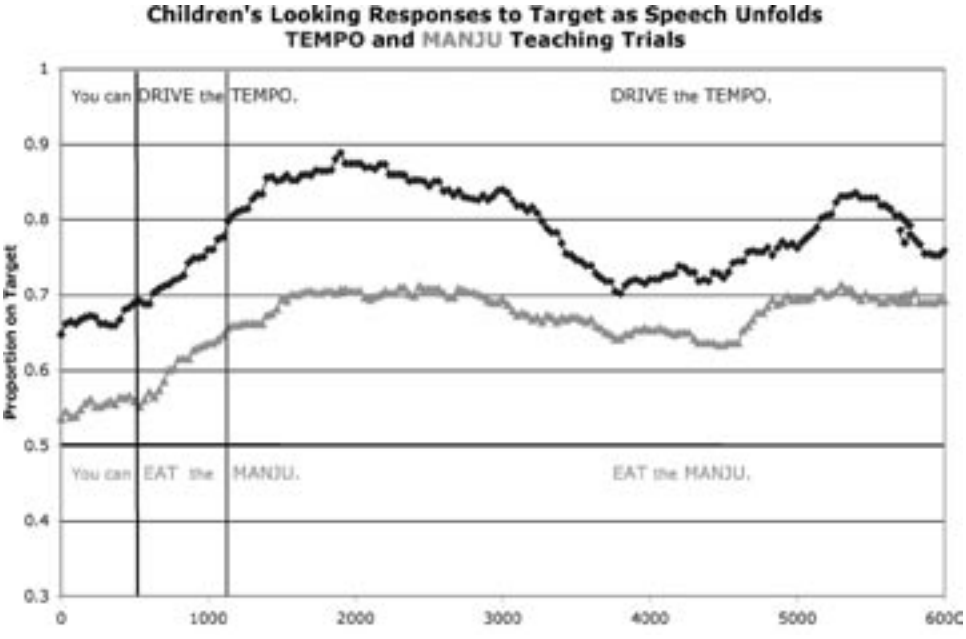


Figure 3. In the teaching trials, children looked at the appropriate picture within milliseconds of hearing the verb. The black vertical lines correspond to the respective onsets of verb and noun, and the black horizontal line at 0.5 indicates chance performance. Children reached about 85% accuracy in tempo trials, and about 70% accuracy in manju trials.

locations of the target and distracter pictures. The final analysis yielded a continuous record of the time course of children's eye movements throughout the experiment.

Results and Discussion

The goal of this study was to examine how 28- to 30-month-olds might use information from the linguistic context—specifically from semantically-constraining, familiar verbs—to learn new words. While a previous study by Goodman, et al. (1998) showed that children could use familiar verbs to learn novel nouns in a picture-book task, it was not clear how quickly children made these inferences. Our looking-while-listening procedure allowed us to view the time-course of children's responses as they inferred the appropriate referent for each novel word.

Even in the absence of visual and linguistic cues, children looked more reliably at the target picture referred to in the sentence, indicating they had learned the target words. In the testing trials, children were asked, *Where's the manju?* or *Where's the tempo?* and were highly accurate in orienting to the correct picture (see Figure 2).

To get a better idea of how and when children inferred the referents of the novel nouns, we examined their responses in the teaching section of the study. In the teaching trials, participants looked more reliably at each target than at its distracter picture. This supports our finding from the testing section that children made the correct inferences about the novel nouns' referents (see Figure 3). Furthermore, the graph shows that children use the information from the familiar verb to orient to the correct picture – even before hearing the label for the novel noun.

By the first teaching trial, children were looking at the target picture when they heard the noun label, allowing them to make the mapping between label and object. In this regard, the verb pointed

the child to the correct object, much as ostensive labeling does in a natural context. Although two-year-olds may be speaking only a few hundred words, they are already drawing on contextual information to rapidly learn new ones. This finding supports the current focus in language acquisition research on inferential rather than ostensive word learning.

Further Research

Our findings suggest an important role for contextual cues in language acquisition. Future studies could extend our results to a wider range of semantically constraining verbs and could also explore the development of this learning strategy as children's vocabularies grow. We are currently investigating children's initial associations between objects and actions as well as the robustness of these representations. Further research should continue examining the roles of syntactic, semantic, and social-pragmatic cues in guiding early word learning.

Reference

Goodman, J., McDonough, L., and Brown, N. (1998). The Role of Semantic Context and Memory in the Acquisition of Novel Nouns. *Child Development*, 69, 1330-1344.

Tiffany Early, Ana Luz Portillo, and Carolyn Quam

A mutual interest in language acquisition brought Tiffany Early, Ana Luz Portillo, and Carolyn Quam to Professor Anne Fernald's research lab, the Center for Infant Studies, in the fall of 2002. The following summer, they conducted this study as participants in Stanford University's Summer Research College and presented their findings at the 2nd Annual Symposium for Undergraduate Research in Progress. In March 2004, Fernald and Zangl also presented this research at the 17th Annual CUNY Conference on Human Sentence Processing at the University of Maryland.

Tiffany Early is a junior majoring in Human Biology with an area of concentration in neural and cognitive development. Ana Luz Portillo is a junior in Human Biology as well, focusing on language and cognitive development. She is also pursuing a minor in Spanish and Portuguese. Carolyn Quam is a senior honors student in Psychology minoring in music. Next year, she will continue studying language acquisition as a PhD candidate at the University of Pennsylvania. They would like to thank their mentors Anne Fernald and Renate Zangl.

French Music

Gaspard de la nuit: Horror and Elegance

This paper examines Gaspard de la nuit*, a composition for solo piano written by Maurice Ravel in 1908. The research concerns those aspects of* Gaspard *that locate and solidify it in the history of music, and make it a valuable and unique composition: its virtuosity and technical inventiveness; its basis on an earlier French poem by Aloysius Bertrand; and its function as a window into Ravel the composer and the man. One can marvel at the ways in which Ravel pushed technique to its then-limits. One can discover the amazing ways in which Bertrand's prose poetry became Ravel's highly imagistic and structured music that virtually speaks through notes. One can understand the style, working process, and personality of a brilliant and complicated man. The paper has been written for novice and knowledgeable musician alike, though if one wishes for a more in-depth experience, the author recommends a copy of the musical score to accompany the reading. Recordings of* Gaspard de la nuit *by pianists Jean-Yves Thibaudet or Arturo Benedetti Michelangeli are also recommended. This paper does not seek to prove anything new; it is merely a passionate exploration that, hopefully, inspires appreciation for, and sheds extensive light on,* Gaspard de la nuit.

Alexander Eccles

"Gaspard has been a devil in coming, but that is only logical since it was he who is the author of the poems."

"My ambition is to say with notes what a poet expresses with words."

-Maurice Ravel (Dubbiosi).

When pianists think of *Gaspard de la nuit*, their hearts stop. Ravel's masterpiece, as famous for its depiction of Aloysius Bertrand's poems as it is infamous for stunning virtuosity, is like a realistic dream, a lucid world of darkness and terror evoked through refinement and detail. Ravel brings to life Bertrand's poems using a new and forward-looking technique, a classical form, and a vast knowledge of keyboard composition to convey the likes of mermaids, monsters and corpses. Though daring technical devices and harmonies were an integral part of Ravel's realization of the poems, the sheer technical accomplishment of the work is in itself an outstanding feat. The music extends the capabilities of the ten-fingered, two-handed pianist, and in the process makes *Gaspard de la nuit* one of the most challenging and virtuosic compositions ever written. Yet, for all of its remarkable innovations, *Gaspard de la nuit* is confined to a strict, classical structure that is in many ways like a conventional sonata: an opening movement of moderate speed, a central slow movement, and a fast, impressive finale (Perlemuter, 1988). However, this exacting approach to compositional form did not hamper Ravel's realization of the poetry; in fact, it was with a classical form that Ravel could follow Bertrand's work like a blueprint, thematically reconstructing every dancing demon and burst of water. Ravel was a master of using the piano-- everything from scales and arpeggios to rhythms and articulations unique to the instrument-- to create specific objects and emotions. Detailed, almost tangible images make *Gaspard de la nuit* unforgettable, defining its immortality as a masterpiece which is both a brilliant and ground-breaking technical study, and a hauntingly evocative musical recreation of Bertrand's poetry.

A Brief History Of The Composer

Ravel was born on March 7, 1875 in the Basque fishing village of Ciboure, but when he was very young the family moved to Paris. Ravel joined the Paris Conservatory at age 12, but he made most of his musical friendships outside the school. For example, he joined "Les Apaches," a group of artists that included Manuel de Falla. One of his closest friends, however, was the pianist Ricardo Vinès, who was an important figure in Ravel's life; he even gave the first performance of *Gaspard de la nuit*. While Paris was a stimulating musical environment, it was not the only source of Ravel's inspiration. As a child, he was taken to a factory and was fascinated by the engineered precision of the machines; Ravel would remember "their clicking and roaring, which, along with the Spanish folk songs sung to me at night-time as a berceuse by my mother, formed my first instruction in music!" (Larner, 1996). Even in Ravel's childhood there is evidence of

two predominant qualities in his music: a love of melody and imagery, and attention to mechanisms and details. *Gaspard de la nuit* is the pinnacle of this two-sided style, and its unparalleled capacity to evoke images is actually enhanced by a precise structure.

Igor Stravinsky called Ravel the "Swiss clockmaker" (Bricard), an appropriate nickname for a man who sought perfection and flawlessness in his art. Ravel treated his compositions as if they were fine jewelry or handmade ornaments; one can imagine him carving each note into the page, using a magnifying glass to make sure every inch of music was ideal. He also had a flair for elegance and formality in his compositions as well as in his personal life; almost all known photographs show a man who has groomed himself for the picture. He was "meticulous about his personal appearance... as unlikely to present himself as he was a composition to the world in an 'unfinished' state" (Hopkins, 1986).With this meticulous nature, Ravel's works are picture-perfect. Though in the case of *Gaspard de la nuit* it is not obvious how one can create a "perfect" representation of a swinging corpse, it is safe to say that the refined, glossy music makes such depictions more realistic.

To bring his works closer to perfection, Ravel borrowed from other artists, employing their styles and techniques more with pride than embarrassment. He once said, "To develop your craftsmanship, you must learn from others," you need "a model or source of inspiration" that would improve the quality of his music (Dubbiosi, 1967). For example, Ravel wrote *Gaspard de la nuit* intending to compose three poems of "transcendental" virtuosity -- an allusion to Liszt's Transcendental Etudes. According to Roland-Manuel, "Transcendental takes the word in its exact meaning, because the virtuosity transcends its domain and makes *Gaspard de la nuit* one of the principal successes of Ravel and one of the summits of his art" (Bricard, 9).

Before analyzing each of the movements, we must examine two aspects essential to the entire piece: relationships between music and poetry, and technical accomplishments.

The Work As A Whole

Ravel and Bertrand seem to have been destined to meet at some point. Both shared a talent for precision and imagery in their art; Bertrand was a "goldsmith of words" (Perlemuter, 1988). *Gaspard de la nuit* was first published in 1842, a year after Bertrand died, but a 1908 reprint of the poems in the *Mercure de France* was Ravel's probable source of inspiration (Larner, 1996). Bertrand's gothic prose poems were perfect models for Ravel. In each of the three poems Ravel chose from the set-- respectively *Ondine, Le Gibet* and *Scarbo*-- the poetry is dark and hallucinatory. Even the title seems fantastic: Gaspard derives from a Persian name for one who guards treasure (Bruhn, 1997).

Although classical in structure, with three movements of proportionate dimensions and organized sections where distinct motifs are introduced and developed, *Gaspard de la nuit* is hardly a Mozart sonata. *Ondine, Le Gibet,* and *Scarbo* are polyrhythmic and polytonal, with complex rhythmic patterns and vague key signatures. There are also post-impressionist harmonies and even elements of jazz (Dubbiosi, 1967).

Ravel had an amazing talent for evoking physical and visible things-- what may be referred to as "musical onomatopoeia" (Dubbiosi, 1967). He created descriptive objects throughout some of his compositions, from Spanish jesters to sad birds, dead princesses to hundreds of chiming clocks (as in *Alborada del gracioso, Oiseaux tristes, Pavane pour une infante défunte*, and *La Vallée des cloches*). He had composed works based on poems before, too: *Jeux d'eau*, about water, and *Noctuelles*, or "Night Moths" (Dubbiosi, 1967). *Gaspard de la nuit* is the pinnacle of this representative style-- a veritable treasure chest of images-- because Ravel captures every minute detail from the poems: a dark sky, a lake, a mermaid-princess, a hanged corpse, a sunset, an empty city, a dancing dwarf-monster, nightmarish shadows, and, in an even more impressive display of his control of keyboard effects, such nebulous entities as mystified horror, resigned sorrow, death, lust, and nervous laughter. How on earth, one might ask, does one use a piano to mimic the sentiments of a lustful mermaid shrouded in sprinkling water? Ravel accomplished this feat by using, with superb imagination, an enormous range of sonorities, articulations, and rhythms that actually replicate Bertrand's words. For example, broken octaves and dotted rhythms may be said to represent a creature who moves "like the spindle fallen from the wand of a sorceress" (from Bertrand's Scarbo). Or, for another of many possible examples, quiet and slowly dissipating chords, synesthetically, actually sound like a "setting sun" (from Bertrand's Le Gibet). Ravel used essentially everything, prompting legendary French pianist Alfred Cortot to say that "these three poems enrich the piano repertoire of our era by one of the most astonishing examples of instrumental resourcefulness that I have ever witnessed in the work of composers" (Bricard).

However, its most impressive and valuable aspect is a comprehensive and revolutionary study of piano technique. Not only does it explore a range of sounds and articulations, but all three movements demand a remarkable virtuoso technique. Ravel wanted to advance piano technique; he was quoted as wanting to compose "something more difficult than *Islamey*" (Larner, 1996), which is widely regarded as one of the most challenging pieces in the repertoire. *Gaspard de la nuit*'s difficulty lies not only in the notes themselves, but, more importantly, in the

way the notes should be played. There are frenzied staccato chords and languid scale passages, sonorities ranging from *fff* to *ppp*, and inner melodies that must sing out over varying accompanimental voices. The pianist, in short, must concentrate during every second of the approximately 23 minutes it takes to perform this piece.

Ravel focused on particular pianistic challenges in his "transcendental" writing; "[t]he solution of technical problems (problème de métier) was [his] primary purpose in composing *Gaspard de la nuit*" (Dubbiosi, 1967). For example, parts of *Ondine* consist of lightning-fast arpeggios with each hand in contrary motion on different tonal keys (creating dissonance). Other technical demands include the interaction between the hands, glissandi, repeated notes, and functions of the thumb.

The division of melodies and chords between hands had been used before by Liszt's rival Sigismond Thalberg, who created a three-voiced effect by having an inner melody surrounded by two accompaniments. Ravel created more range through "doubling the melody in octaves and the placement of the harmonic accompaniment within the limits of the melody voices" (Dubbiosi, 1967). When one hand plays the melody, the other joins in the melody or plays accompanying notes in such proximity that specific hands playing certain notes are no longer obvious.

Glissandi are used primarily for imagery in Ondine. The conventional glissando is a sweeping motion with one finger over the white keys of the piano (Liszt used glissandi in the "classic" form in many of his virtuoso showpieces). However, Ravel bids the pianist use the black keys, which is not only very difficult, but also creates tonalities less prominent in Western music.

The use of repeated notes "was to become a hallmark of Ravel's style" (Hopkins, 1986). Repeated notes are divided between octaves, shared by both hands, played unimaginably fast, and, famously, form the integral structure of Le Gibet.

Of particular importance is the thumb which, as Roland-Manuel claims, "is very remarkable in the works for piano of Ravel. It is the thumb that commands, particularly in *Gaspard de la nuit*, one of the works most characteristic of his transcendental technique" (Bricard, 9), extending the functions of the hands, the capabilities of pianists, and the range of modern piano-playing.

Although we have examined the style in which *Gaspard de la nuit* was composed as a whole, it is necessary to analyze each movement in detail.

Ondine

Ondine is arguably the most beautiful movement of *Gaspard de la nuit*; it is certainly the most colorful and sensuous. In the poem, Ondine is a mermaid who sings to a man, describing her fantastic world and trying at length to seduce him. The man, however, is married to a mortal, and when he explains this to Ondine she cries, laughs, then disappears as quickly as she first appeared. While this basic story forms the skeletal structure of Ravel's music, there is far more to the poetry than a simple narrative. Words and phrases create impressions of darkness, mystery, and a magical world of water. Ravel's incorporates these themes as he follows the progression of the story, introducing and developing the setting and characters.

Just as the poem begins with "Listen! Listen!" and does not provide any introduction to explain what is going on, the music evolves out of nothingness, like something from a dream. Ravel begins the piece in *ppp* with a very fast and quiet repeating series of chords that becomes the central and driving rhythmic element of the work. These initial bars establish a quiet atmosphere of mystery, metrical elegance, and anticipation. The shimmering notes immediately draw one into "the beautiful starry night and the beautiful sleeping lake."

The actual poetry begins in the third measure, where Ondine is first heard in a dreamy, song-like melody. When Ravel introduces her playfully amorous character, the melodic line conjures images of the mermaid calling "Listen! It is I, it is Ondine who brushes drops of water on the resonant panes of your window" from outside. The music "follows the actions and thoughts of the sprite-princess and sustain[s] the least inflection of her voice" (Long, 1973), and later becomes more complicated as Ondine's song builds in strength and sensuality. For example, when Ravel repeats the melody, it is in broken chords rather than single notes, with more agitated accompaniment. In the third stanza of the poem, Ondine again calls "Listen! Listen!"-- the repetition creates urgency. Incidentally, the first three stanzas of the poem are in ABA form, which is also applicable to the musical organization: a theme (A), which goes through a development section (B), and the initial motifs are then repeated (A) (Bruhn, 1997).

Ravel changes the key and melodic line as Ondine offers her love to a married man and tension mounts. The music is more sinister, and for the first time the melody is sung in the lower, darker regions of the keyboard. There is unmistakable momentum as the music becomes more dance-like, and soon the playfulness disappears. A stormy romantic passage builds as she offers her hand in marriage. When the man rejects Ondine, water seems to crash everywhere as the pianist pounds over the keyboard. In the wake of this terrible climax, the music is softer and more fragile, eventually disintegrating into nothing more than a timid reflection of the melody. This is the nadir of the piece, showing Ondine alone with a mixture of childish disappointment and adult sadness-- "she wept some tears." However, as she "uttered a burst of laughter, and vanished in a shower," the music blossoms into a coda of rapid arpeggios before trailing off like a

dream.

Ondine is difficult to play because it requires a good deal of strength to maintain a light touch. The pianist must create a supple and brilliant atmosphere, with subtly different touches for water that ripples, shines, shimmers, and cascades. At one point, the hands traverse practically the entire keyboard in a matter of seconds, yet even in this loud and fearsome passage the music must be flowing, like water. In addition, the rhythmic motif of chords is hard simply to play fast, but near impossible to play fast and quietly. While the pianist sweats over this technical challenge, the melody must sing out, which becomes difficult to maintain when the hands begin leaping around the keyboard. Still, despite the difficultly, the atmosphere must remain constant and vivid. The pianist must create a sparkling, nighttime atmosphere, like a prism suspended in rain and darkness.

Le Gibet

Le Gibet is a picture of desolation and misery. It is a musical landscape of the singular, breath-taking image of a lonely corpse "reddened by the setting sun." Bertrand's poem is a series of contemplations of this image, all made with gloomy, hopeless resignation. Unlike *Ondine*, there is no story being told. Instead, the poem consists of five gruesome questions and a horrific answer. *Le Gibet* opens with a picture of a man who is "the hanged one," yet who is still alive and "utters a sigh." By the end of the poem, after the relentless series of disturbing questions, the man is certainly dead-- his emotions have expired, his spirit is extinguished, he is now a "corpse." It is as though we, as readers, have witnessed "the unfolding of that moment between almost-no-life and definite death" (Bruhn, 1997).

Ravel devotes almost the entire focus of the composition to atmosphere. He does not try to recreate specific descriptions like "scarab beetles" and "sterile ivy." Those images are incorporated into a larger theme based on the evocation of emotions rather than objects. The character of the poetry is introspective and cyclical, more observational than explanatory. Like the poetry, the music is an odyssey withone mournful, extended image. The whole movement consists of three relentlessly repeating motifs: a constant B-flat, a melodic chord progression, and a second, more singing melody. Pianist Vlado Perlemuter remarked that "you must not be afraid of making it sound monotonous"; in fact, monotony is an integral part of the music.

The forlorn atmosphere is maintained by a B-flat that sounds throughout the entire piece, like a perpetual sigh or ringing bell. Bells play a key role in *Le Gibet*. The musical landscape is reminiscent of Edgar Allen Poe, whose poetry Ravel admired; in fact, Poe's poem The Bells may have influenced this music (Larner, 1996). Siglind Bruhn finds that the B-flat bell is important because the "tolling constitutes the link among the different aspects embedded in the poem." Though this B-flat theme is the most famous aspect of *Le Gibet*, all three motifs overlap and build off one another to create an indelible image.

Technical difficulties of *Le Gibet* are in the management of sound quality and voicing. Marguerite Long points out that independence of the hands and discrete differences between *ppp* and *pp* make the piece extremely challenging. The pianist must not only play quietly, but also be aware of subtle changes in sound and texture. Henri Gil-Marchex believes 27 different touches are necessary in *Le Gibet* (Bricard). A particularly difficult task is the maintenance of each musical line. For example, the B-flat is sometimes divided in octaves between hands, but it still must sound exactly the same as when played by only one hand. At other moments, different melodic lines are played simultaneously and must sound as completely independent voices. Another technical difficultly is the playing of extended chords. All the notes must be played in the same way and at the same time, and, because the music is slow and quiet, any notes depressed a half-second too late are obvious mistakes. Performing *Le Gibet* is like standing perfectly still so that even your breathing is imperceptible, only you must also always be looking at a single ghastly scene. When the music at last dies out, that ceaseless B-flat is the only sound that remains. It is like the corpse left hanging from the rope, and then the scene fades to black.

Scarbo

Scarbo: this is arguably the most famous movement from *Gaspard de la nuit*, the piece people talk about long after they have first heard the nightmarish music. *Scarbo* is truly the work that represents transcendental virtuosity. The music is unbelievably difficult and seems very advanced and dissonant for Ravel's time. Literally every key of the piano is used. Furthermore, the virtuoso elements become a vehicle for conveying the poetry, which like the music is frenetic and bizarre, almost drugged-out. However, as Bertrand never abandons structure, Ravel manages to fit the shadows and dwarfs and hallucinatory images into a "tightly knit form [that] is comparable to the Chopin *scherzi* and *ballades*" (Dubbiosi 1967). The result is an intense and relentless journey that lies somewhere between a technical study and a psychotic episode.

In the poem, the narrator describes his fear of Scarbo, an evil dwarf who comes in the dead of night. Scarbo plays with the narrator's mind: sometimes he dances, other times he hides and only makes noises, and then sometimes he appears and "grows between the moon and me like the belfry of a gothic cathedral." At the end, as the insanity becomes unbearable, Scarbo disappears, "his face pales like the wax of a candle end--and suddenly he is extinguished." Although the poem is short, there is grandeur in its horror, "a sense of awe, embodied in the image of

the majestically high spire of a Gothic cathedral" (Bruhn, 1997).

Although the music is surreal, it is relatively simple to follow its relationship to the poetry. *Scarbo* consists of several motifs, each representing a theme or image from the poem. The piece begins with three ascending notes, uttered softly and deeply from the lower registers of the piano. This is like a "sneak preview," like the trailer to a horror film (Bruhn, 1997). The three-note motif varies from a whisper to a romantic sweep, and representing the emotions of the narrator. At the end of the piece, this is played one last time in assertive left hand octaves before the music abruptly trails off, representing a sort of relief as Scarbo disappears. Likewise, a staccato and rhythmic theme is frantic and grotesque, like the dwarf's jagged, uneven dancing. Another motif consists of broken octaves and is more mysterious; this represents suspense, like Scarbo hanging from the ceiling, unseen. Finally, there is a very intense, rhythmic progression of chords that interrupts the broken octaves and twice builds up to a climax, like a scream of awe and terror. Ravel, a brilliant orchestrator, wanted the piano to sound like "kettle drums" at this point, which creates a frightening effect (Perlemuter, 1988). Ravel also integrates and interchanges these motifs. For example, there is a strangely calm section in the middle of the piece, but its melody is actually a slower version of *Scarbo*'s "dancing" theme. In this way, Ravel creates anticipation, hinting that the dwarf is hiding, but not gone-- "Do I think him vanished then?"

There are three areas of technical concern in *Scarbo*: playing the notes, developing specific articulations for each motif, and, possibly more difficult than anything, the juxtaposition of very different articulations. Simply learning the notes is an accomplishment, for *Scarbo* demands speed, precision and agility. The pianist must cross and interlock hands, play multiple notes with the thumb, and perform similar acrobatics at a very high speed. However, the pianist cannot simply hit the right notes; the pianist must hit the right notes in the right way. For example, the initial sequence of repeated notes is very fast and therefore challenging to play. However, as Bruhn insightfully points out, "[t]he musical metaphor is that of trembling."

Still, the most challenging aspect of *Scarbo* is adjusting to accommodate different articulations in virtually no time. As Marguerite Long said, "the performer must learn to cope with continued neuro-muscular equilibrium in the fingers." At one moment, for instance in one of the "jumping chord" passages, the playing must be light and quick, with measured staccato. Then, a half-second later, the pianist must entirely adjust the volume and articulation to play pedaled broken octaves. Likewise, the atmosphere changes from frenzy to mystery. If the pianist can somehow master the transcendental virtuosity of *Scarbo*, the result is a captivating musical experience.

It is interesting to note that *Scarbo* ends much like *Ondine* begins, with fast and quiet notes played in the higher registers of the piano that seem to shimmer. This is further evidence of the cohesiveness of *Gaspard de la nuit. Ondine* and *Scarbo* have a mixture of darkness and color, a glittery effect. *Ondine* opens with a "beautiful starry night," and *Scarbo* with "the moon glitters in the sky like a silver shield on an azure banner strewn with golden bees." The only movement that does not sparkle in the darkness is *Le Gibet*, but this is appropriate since it is the slow movement.

Ravel even provided explicit instructions for the work's emotional interpretation: "With *'Scarbo'* and '*Ondine*'... it is fitting to bestow them with the sentimentality of Liszt and Chopin, whereas *'Le Gibet*' should be played uniformly throughout, implacably, terrifying by its even simplicity" (Dubbiosi, 1967). *Gaspard de la nuit* fits together brilliantly, and each movement is an impressive enough rendering of Bertrand's prose that the poetry may as well have been based on the music, and not the other way around.

Coda

"*Gaspard de la nuit* is celebrated everywhere as one of the all-time masterpieces of piano writing" (Larner, 1996). Having examined the genius behind this piece, Larner's sentiment seems truthful. It is certainly a widely held belief that this piece is a monumental accomplishment, a work that uses piano technique as a study in its own right and as a means of recreating poetry. No other piano work covers such a range of technical study or is as unforgettably vivid. *Gaspard de la nuit* is musical poetry; every page expresses Bertrand's words in Ravel's own style. To define his style, one must remember that Ravel was an innovator, a perfectionist, a witty romantic at heart, and a man who learned from centuries of music and composers, from Mozart to Liszt to Satie. This style, then, incorporates elements ranging from the classical era into the twentieth century-- elements such as a sonata-like form, romantic images, and a vast range of both classic and modern technical problems. *Gaspard de la nuit* is a compendium of the piano literature, and a perfectly crafted imprint of horror and elegance.

References

Bricard, Nancy. Foreword. Gaspard de la Nuit. By Maurice Ravel. Van Nuys: Alfred Publishing Co., Inc. pp.4-14.
Bruhn, Siglind. Images And Ideas In Modern French Piano Music. Aesthetics In Music no.6. Ed. Edward Lippman. Stuyvesan: Pendragon Press, 1997.
Dubbiosi, Stelio. The Piano Music Of Maurice Ravel: An Analysis of the Technical and Interpretative Problems Inherent in the Pianistic Style of Maurice Ravel. Diss. N.Y.U. Ann Arbor: University Microfilms International, 1967.
Fisk, Josiah, ed. Composers On Music: Eight Centuries of Writings. Boston: Northeastern University Press, 1997.
Hopkins, G.W. "Maurice Ravel." The New Grove: Twentieth-Century French Masters. 1986.
Larner, Gerald. Maurice Ravel. Regent's Wharf: Phaidon Press Limited, 1996.
Long, Marguerite. At the piano with Ravel. Ed. Pierre Laumonier. Trans. Olive Senior-Ellis. London: J. M. Dent and Sons Ltd., 1973.
Perlemuter, Vlado and Hélene Jourdan-Morhange. Ravel According To Ravel. Ed. Harold Taylor. Trans. Frances Tanner. London: Kahn & Averill, 1988.
Thibaudet, Jean-Yves. "Gaspard de la nuit." Comp. Maurice Ravel. Beautiful Starry Night. CD: cover, "Maurice Ravel, Piano Works." London, 1995.

Alexander Eccles

Alexander Eccles is a senior from Tiburon, California. He is majoring in History and English with a Creative Writing emphasis. He loves music and is a pianist, as wellas a golfer and actor. After college he would enjoy teaching, though he also hopes to explore the interests listed above in more depth. He would like to thank virtually everyone, and particularly those who helped in the editing of this paper and those who helped in teaching the author how to become a better writer.

Development Economics

Who Wants Credit?
Explaining the Demand for Land Titling in Mexico

This paper examines the link between land titling and credit in the Mexican land titling program - dominio pleno. Land titling programs assign secure property rights on land assets and potentially increase credit with titled assets being used as collateral to obtain loans. These programs can bring new economic opportunities to the poor who often lack formal ownership over their assets. The empirical evidence on the effect of titling on credit access, however, is mixed, and many authors have suggested the need to account for non-randomness in allocation of titles. This study addresses non-randomness of titling by examining the demand for dominio pleno, arguing that only those who expect titling to bring higher profits would adopt the program. Regression analysis shows that individuals obtain titles to sell their land for urbanization rather than to ask for credit. Therefore, land titling may not increase credit because individuals are not acquiring titles to obtain credit in the first place. The main implication of this study is that land titling effects are, in part, endogenous to the decision to acquire titling.

Fernando Galeana

Don Ramon looks at the maize growing in his parcel and thinks of the day, 30 years ago, when he received from the Mexican government the right to cultivate these two hectares of land. That day, Don Ramon became an ejidatario – a member of the social tenure system known as the ejido. Nevertheless, it was not until two years ago that Don Ramon received an individual certificate from the government program PROCEDE that detailed his parcel's boundaries. Today, Don Ramon is going to the Asamblea Ejidal to discuss with the other ejidatarios the option to adopt domino pleno or full dominion over their parcels. Ejidatarios were not allowed to sell, rent, or mortgage their parcels until 1992, when the government changed the law to give them the option to adopt domino pleno and receive individual titles. With titling, Don Ramon would be able to sell his parcel or use it as collateral for a loan. Don Ramon knows he needs money to buy inputs, like fertilizer, pesticides, and seeds, for next season. He also knows that a real estate developer is interested in buying his land to build a housing project. Don Ramon looks at the maize sprouting in his parcel and thinks…

Today, farmers like Don Ramon are at the center of a widespread debate over land titling. Governments around the world have implemented land titling programs to bring tenure security to informal landholders. The groundbreaking work of economist Hernando De Soto has been an important vehicle to spread this new interest on land titling. De Soto (2000) argues that titled property creates capital because landholders can use their assets as collateral for loans. Therefore, land titling programs can foster economic growth among the poor who lack formal ownership over their assets. I examine the link between land titling and credit access in the Mexican ejido sector. Rather than assuming that ejidatarios would acquire titles to increase credit access, I ask what factors would drive the demand for dominio pleno. I find that ejidatarios are more likely to acquire titling to sell their land for urbanization than to ask for credit. Therefore, land titling may not increase credit access in the ejido sector. The main contribution of this analysis is that the effect of land titling programs is endogenous to the participants' decision to acquire title.

Throughout much of Mexican history, land was concentrated in large, centrally-managed estates known as *haciendas*. After the Revolution of 1910, the government committed to carry out a land reform to redistribute many of these estates among landless farmers, resulting in the ejido system. Ejidatarios are granted the right to enjoy the *usufruct*, agricultural output, but they cannot transfer their assets through selling, renting, or mortgaging. Although the ejido fulfills the original objective of providing ejidatarios with land to cultivate, as times change, its restrictions creates constraints on other efficient uses of the land. The restriction on land mortgaging makes ejidatarios reliant on government development

banks to receive credit. The restriction on land sale forces ejidatarios into illegal markets to sell their land at lower prices than the legal real estate market. These restrictions became more inefficient since the 1980s as development banks decreased their financing and the demand to transform agricultural land into urban land increased.

In 1992, the government decided to lift these restrictions by giving ejidos the option to adopt domino pleno and receive individual titles to their parcels. For ejidatarios, dominio pleno is an additional step to the land certification program PROCEDE that also began in 1992 to grant individual ejido certificates. Domino pleno, however, represents organizational and monetary costs to ejidatarios. A quorum of ejidatarios must vote in favor of the program in the Asamblea Ejidal, the governing unit of the ejido. After dominio pleno has been accepted, ejidatarios are responsible for requesting their individual titles and paying taxes at the higher private property rates. The question that I address in my research is what would make ejidatarios want to adopt domino pleno and, specifically, whether ejidatarios acquire titles to use their land as collateral or to sell it for urbanization. I argue that land titling is endogenous in the sense that its effects depend on the reasons why individuals decided to acquire titles.

The rest of the paper is organized as follows: first, I address the literature that has examined the link between land titling and credit access; second, I describe the economic and econometric models I use to explain demand for dominio pleno; third, I show the results from the empirical analysis; and finally I conclude with some remarks about how to study land titling programs.

Literature Review

One of the main methodological challenges in empirical studies is the endogeneity of land titling. When the allocation of titles across households is not random, individuals choose to join these programs based on various economic, political, and social reasons. Therefore, studies that ignore land titling endogeneity tend to have biased estimates of the effects on credit access due to the selection bias in the titled sample. Some studies (Field 2002; Galiani and Schargrodsky 2003) solve the problem of endogeneity by selecting natural experiments where title allocation is random. Most land titling programs, however, are not natural experiments and empirical studies have found contradictory conclusions.[1] These mixed results are often attributed to the problem of addressing the endogeneity of titling (Field 2002).

In Mexico, empirical studies have found no significant impact of the land certification program PROCEDE on increasing credit access. Deininger, Lavadenz, Bresciani, and Diaz (2001) argue that the 1995 financial crisis in Mexico and the banks' unwillingness to accept the mortgaging of agricultural output are the main reasons land certification is not able to increase credit. Johnson (2002) tests the hypothesis that farmers face asset-based credit constraints due to a lack of collateral. The analysis does not support this hypothesis, and Johnson concludes that PROCEDE may not have a significant direct impact on ejidatarios' decisions regarding credit and capital use.

I address the issue of endogeneity by examining the factors that determine allocation of titles. I introduce the concept of urban constraint, adding to the credit constraint described by Johnson, to describe ejidatarios that would like to transfer their land to real estate developers but are hindered by legal restrictions. Also, I depart from previous studies about Mexico by using data from dominio pleno rather than the land certification program.

Economic Model

I assume that the two factors driving adoption of dominio pleno are demand for agricultural credit and demand to engage in the market for urban land. In relation to demand for agricultural credit, Feder and Feeny (1993) determine the existence of a collateral-constrained regime in which borrowing is exhausted and can only improve with an increase in the amount of collaterizable assets. Given that titling increases the amount of collaterizable assets, Johnson (2002) assumes that in a collateral-constrained regime adoption of dominio pleno must be driven by desire to expand credit. Similarly, the ejido is in an urban-constrained regime if the amount of land available for urbanization is less than the amount demanded by the market, including ejidatarios and public or private real estate developers. Given the legal constraints to urbanize land, in an urban-constrained regime, adoption for dominio pleno must be driven by desire to participate in the market for urban land. Whether the ejido is collateral- or urban-constrained, there are potential benefits available from dominio pleno.

Nevertheless, not all ejidos that present either constraint are going to participate, since dominio pleno represents a cost to the ejido. Therefore, an ejido will decide to acquire titles if the benefits outweigh the cost of titling. The benefits are the profits generated by either credit-enabled agricultural production or land sale for urbanization. Profits are the result of the revenue and costs involved in each activity.[2] Thus, even if some ejidos are collateral or urban constrained, they will not acquire titling if the cost outweighs the benefits.

To determine the demand for dominio pleno, we can analyze *ex-post* the titled ejidos and extrapolate which activity generated higher expected profits. Since the ejidos had to conduct a cost-benefit analysis *ex-ante* to determine whether benefits outweighed costs, then we can examine the characteristics of titled ejidos – like spatial location and economic activity – to infer whether agricultural production or urbanization was more profitable. For

Table 1 - Descriptive Statistics					
Morelos					
	N	**Min**	**Max**	**Mean**	**S.D.**
Land Titling	180	0	1	0.067	0.250
Urban Population 50,000 - 500,000	180	0	1	0.050	0.219
High Level of Urbanization	180	0	1	0.122	0.328
Population working in urban activities (%)	180	0.018	0.283	0.136	0.049
Ejidos Reporting Nonagricultural Activity (%)	180	0	0.556	0.276	0.164
Population working in agriculture, cattle, and fishing (%)	180	0.015	0.739	0.279	0.162
Ejidos that Have Agricultural Machinery (%)	180	0	1	0.672	0.268
Irrigated Land in Municipio	180	0	1	0.211	0.409
Ejidos that Received Credit form Banrural (%)	180	0	0.667	0.161	0.195
Households with sewage and electricity (%)	180	0.132	0.930	0.701	0.167
Parceled Hectares per Member	180	0	50.453	3.804	4.251
Land that is Forested in Municipality (%)	180	0	0.450	0.032	0.091
Number of Ejidatarios	180	14	880	176.072	156.654
National					
	N	**Min**	**Max**	**Mean**	**S.D.**
Titled Ejidos in the Municipality (%)	2201	0	1	0.048	0.135
Urban Population more than 50,000	2201	0	1	0.058	0.233
High Level of Urbanization	2201	0	1	0.001	0.030
Population Working in Urban Activities (%)	2201	0.001	1	0.122	0.111
Ejidos Reporting Nonagricultural Activity (%)	2201	0	1	0.162	0.252
Population working in agriculture, cattle, and fishing (%)	2201	0	0.998	0.553	0.242
Ejidos that Have Agricultural Machinery (%)	2201	0	1	0.410	0.360
Irrigated Land in Municipality (%)	2201	0	1	0.051	0.169
Ejidos that Received Credit from Banrural (%)	2201	0	1	0.158	0.236
Households with sewage and electricity (%)	2201	0	0.972	0.322	0.246
Parceled Hectares per Member	2201	0.13	120.5	7.572	8.631
Land that is Forested in the Municipality (%)	2201	0	0.99	0.102	0.180
Average Number of Ejidatarios	2201	9	4322	210.377	316.228

example, if the titled ejido is situated in an irrigated area, then agricultural productivity is likely to be high and ejidatarios will adopt dominio pleno in order to demand credit. If the ejido is instead located near an urban zone, then the price of urban land is likely to be high and ejidatarios adopt dominio pleno to sell land to urban developers. To conduct the analysis, I examine the association of land titling with groups of variables that signal demand of agricultural credit and demand for urbanization.

Data

The data to generate the two sets of variables is obtained from three main sources: the 1991 VII National Ejido Census, the 1990 Population and Household National Census, and 2003 Dominio Pleno results published by the National Agrarian Register (RAN). Data are available for all thirty-two states in Mexico at the municipio level. A municipio is the equivalent of a county in the United States. A municipio may have from anywhere from 1 to 206 ejidos within its boundaries. Since the decision to adopt dominio pleno is conducted at the ejido level, it would be ideal to obtain observations at this level. This disaggregated data is not easily accessible and could only be found for the state of Morelos. Therefore, the analysis includes one dataset for Morelos, with 180 observations, and another dataset for the nation, with 2201 observations. Morelos is not necessarily considered a representative state, but it allows the estimation of the demand at a more desirable scale.

The endogenous variable of land titling is calculated using 2003 Dominio Pleno results. For the Morelos dataset, this variable is a binary (0,1) and takes the value of zero if the ejido is not titled and one is the ejidos is titled. Out of 180 ejidos, only twelve are titled. For the national dataset, the land titling variable is the proportion of the ejidos titled in the municipio. Out of 2201 municipalities, 507 have at least one titled ejido.

The exogenous variables are selected from the 1991 VII National Ejido Census and 2000 National Census, and are assigned to a vector representing urban demand, another representing demand for agricultural credit, and finally one representing control variables. The descriptive statistics for these variables are in Table 1.

In the urban ejidos, high demand from real estate developers is expected to transform agricultural land into urban land. The urban vector includes four variables: a population variable that assigns a value of 1 if there is community in the ejido or municipio with more than 50,000 inhabitants and 0 if otherwise; a spatial variable that assigns a value of 1 if the ejido is geographically located inside one of the Morelos's three main urban zones[3] and 0 otherwise, or – in the case of the nation – 1 if the municipio is inside a city with more than 1 million inhabitants and 0 otherwise; a variable indicating the proportion of the population working in urban economic activities, including manufacturing, financial, and professional services; and a variable showing the proportion of ejidos reporting nonagricultural activities.

In the agricultural ejidos, demand for agricultural credit is expected to continue investment of that credit in productivity. This agricultural credit vector includes four variables: the proportion of population in the municipio working in agriculture, cattle ranching, or fishing; the proportion of ejidos in the municipio that own agricultural machinery; the proportion of irrigated land available in the municipio for cultivation; and the proportion of farmers that received credit from Banrural, the government development bank for agriculture.

The variables included in the control group are factors present to partial out any other possible characteristics that may affect titling and could be correlated with either agriculture or urbanization. These variables are: the proportion of households in the municipio that enjoy the basic services of sewage and electricity, the average parceled hectares per member, the proportion of land forested in the municipality, and the number of ejidatarios in the ejido.

Econometric Technique

The structural model to determine demand factors has land titling as an endogenous variable and includes the exogenous variables for the urban, agricultural credit, and control clusters.

Land Titling (1,0) =
β_0 +
β_1 (urban vector) +
β_2 (agricultural credit vector) +
β_3 (control group) +
ε

For the Morelos dataset, OLS and logistic regression techniques are used. Since the dependent variable is a binary variable, the linear probability model has several well-known shortcomings, notably that the predicted probabilities may not lie between 0 and 1. In this case, the binary logistic regression model is used to estimate the probability that an event occurs. For the national dataset only OLS regression is used.

Results

The regression results show there is clear evidence of association between the urban vector and dominio pleno in Morelos and more mixed, but still significant, association in all of Mexico. In column 1 of Table 2, the OLS results for the structural model for Morelos show that high level of urbanization is positive and significant at the 5% level. In column 2, the logistic results show that in the urban cluster the variable for urban populations larger than 50,000 is negative and significant at 5% level and high level of urbanization is positive and significant at the 10 % level. In the agricultural credit cluster, the variable of rural employment is negative and significant at the 5%, but the credit variable is positive and significant at the 10% level. In column 3, the OLS results for the structural model for the national dataset show that urban populations larger than 50,000 and urban employment are positive and significant at the 5% level in the urban cluster. In the agricultural credit cluster, rural employment is negative and significant at the 5% level, but agricultural machinery and credit variables are positive and significant at the 5% level.

Conclusion

I examine whether ejidatarios adopt dominio pleno either to demand agricultural credit or participate in the market for urban land. The statistical analysis suggests that demand to participate in the market for urban land is the main incentive to acquire titling. The significance of this analysis is to show that land titling effects are endogenous to the demand for titling. For example, if an empirical study finds no effect of dominio pleno on credit access, we now know that part of the reason is that ejidatarios do not acquire titles to increase their credit in the first place. As a result, estimates of access to credit as a function of land titling might easily find no result in areas where titling is demanded mostly for urbanization purposes. Even if land titling programs provide landowners with titled assets, this is a necessary, but not sufficient, condition for this collateral to be able to be used for credit. Owners must first demand credit, and then other institutional features, such as a good legal structure, will determine whether credit is supplied on the basis of this collateral.

Table 2 - Regression Results			
Morelos Dependent Variable = Land Titling (0,1)			
National Dependent Variable = Land Titling (0-1)	Morelos	Morelos	National
	(1)	**(2)**	**(3)**
Urbanization			
Urban Population more than 50,000	-0.120	-3.029**	0.091**
	(0.079)	(1.548)	(0.013)
High Level of Urbanization	0.256**	1.532*	0.071
	(0.069)	(0.916)	(0.087)
Population working in urban activities (%)	0.011	5.967	0.084**
	(0.590)	(13.545)	(0.035)
Ejidos Reporting Nonagricultural Activity (%)	0.029	0.194	-0.010
	(0.143)	(3.360)	(0.011)
Agricultural Credit			
Population working in agriculture, cattle, and fishing (%)	-0.431	-20.044**	-0.078**
	(0.276)	(10.031)	(0.021)
Ejidos that Have Agricultural Machinery (%)	0.051	6.357	0.032**
	(0.087)	(4.540)	(0.008)
Irrigated Land in Municipio (%)	-0.325	-15.973	0.004
	(0.528)	(13.400)	(0.016)
Ejidos that Received Credit form Banrural (%)	0.167	11.296*	0.030**
	(0.133)	(6.948)	(0.012)
Control Variables			
Households with sewage and electricity (%)	-0.145	-8.216	0.057**
	(0.221)	(6.533)	(0.016)
Parceled Hectares per Member	-0.003	-0.516	0.000
	(0.004)	(0.360)	(0.000)
Land that is Forested in Municipality (%)	0.230	19.435	0.008
	(0.254)	(12.734)	(0.015)
Number of Ejidatarios (x10,000)	0.779	44.729	-0.174**
	(0.000)	(0.003)	(0.000)
Constant	0.236	1.523	0.043**
	(0.260)	(5.468)	(0.018)
R-squared	0.185	-	0.178
Adjusted R-square	0.126	-	0.174
Akaike Information Criteria	-2.826	-	-4.196
Schwarz BIC	-2.593	-	-4.163
Number of Observations	178	178	2205
* p-values < .10			
** p-values < .05			

Author Notes

[1]Some studies in Brazil and Honduras find a positive impact of titling on credit (Alston, Libecap, and Muller 1999; Lopez 1997). Other empirical studies done in some rural areas of Kenya, Ghana, Rwanda, Somalia, and Paraguay, however, find that land titling has no significant effects on economic development (Migot-Adholla, Hazell, Blarel, and Place 1991; Carter, Wiebe, and Blarel 1991; Roth, Unruh, and Barrows 1994; Place and Migot-Adholla 1998; Carter and Olinto 2002).

[2]Agricultural credit profits = _ * P^f * $f(K_f) – K_f – (1 + _ *r)*B$, where (_) is the risk to agricultural income, (P^f) is the relative price for agricultural output, (f) is agricultural production which is a function of capital (K_f), (B) is credit available at interest rate (r), and (_) is the transaction cost that makes the market interest rate different from the effective rate (Johnson 2002). Urbanization profits = _ * P^u * $u(K^u) – K^u – \Omega$, where (_) is the probability of finding a real estate developer, (P^u) is the relative price for urban land, (u) is urban production which is a function of capital (K^u), and (Ω) is transaction cost associated with closing a deal.

[3] The three main urban zones in Morelos are Cuernavaca, Cuautla, and Jojutla.

References

Alston, Lee J., Libecap G.D., and B. Mueller. 1999. *Titles and Land Use: The Development of Property Rights on the Brazilian Amazon.* University of Michigan Press.

Carter, Michael, Keith Wiebe, and Benoit Blarel. 1991. "Tenure security for whom? Differential Impacts of Land Policy in Kenya." Land Tenure Center Research Paper No. 106, University of Wisconsin-Madison.

Carter, Michael, and Pedro Olinto. 2002. "Getting Institutions Right for Whom? Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment." *American Journal of Agricultural Economics* vol.85(1):173-86.

De Soto, Hernando. 2000. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* New York: Basic Books.

Deininger, Klaus, Isabel Lavadenz, Fabrizio Bresciani, and Manuel Diaz. 2001. *Mexico's "second agrarian reform:" Implementation and impact*. Washington DC: World Bank.

Dominio Pleno. 2003. Last update September 30, 2003. Available at http://www.ran.gob.mx Accessed 15 December, 2004

Feder, Gershon, and David Feeny. 1993. "The Theory of Land Tenure and Property Rights." In *The Economics of Rural Organization: Theory, Practice, and Policy*, Edited by Karla Hoff, Avishay Braverman, and Joseph Stiglitz. New York: Oxford University Press.

Field, Erica. 2002. "Entitled to Work: Urban Property Rights and Labor Supply in Peru." Working Paper.

Galiani, Sebastian, and Ernesto Schargrodsky. 2003. "Effect of Land Titling". Working Paper. INEGI. VII Censo Ejidal. 1991. Available at http://www.inegi.gob.mx Accessed 15 December, 2003

INEGI. XII Censo General de Población y Vivienda. 2002. Available at http://www.inegi.gob.mx Accessed 26 January, 2004

Johnson, Nancy L. 2001. "Tierra y Libertad: Will Tenure Reform Improve Productivity in Mexico's Ejido Agriculture?" *Economic Development and Cultural Change* v.49(2):291-309.

Lopez, R. and C. Romano. 1997. "Rural Poverty in Honduras: asset distribution and Liquidity constraints." Working Paper.

Migot-Adholla, S.E., Peter Hazell, Benoit Blarel, and Frank Place. 1991. "Indigenous land rights systems in Sub-Saharan Africa: A constraint on productivity?" *World Bank Economic Review* v.5(1):155-175.

Place, Frank and S.E. Migot-Adholla. 1998. "The economic effects of land registration for smallholder farms in Kenya: Evidence from Nyeri and Kakamega Districts." *Land Economics* v.74(3):360-73.

Roth, M., J. Unruh, and R. Barrows. 1994. "Land registration, tenure security, credit use, and Investment in the Shebelle Region of Somalia." Pp. 1999-230 in *Searching for Security of Tenure in Africa*, edited by Bruce and Mighot-Adholl. Dubuque, IA: Kendall-Hunt.

Seyde, Federico. 2000. "La incorporación de tierras de propiedad social al desarrollo Urbano en México: marco legal y estrategia de la política pública." Pp. 71-86 in *Los Pobres de la Ciudad y la Tierra*, edited by Alfonso Iracheta and Martim Smolka. Mexico: El Colegio Mexiquense and Lincon Institute of Land Policy.

Fernando Galeana

Fernando Galeana is a senior from Cuernavaca, Mexico. He is an Economics major with a focus in development economics. Since coming to Stanford, he has interned with the World Bank, studied the impacts of ecotourism in an Ese-Eja community in the Peruvian Amazon, and conducted field research in Mexico for his honors thesis in Latin American Studies. Fernando would like to thank Professors David McKenzie and Avner Greif for their invaluable advice; Alexandra Gerbasi, Sabrina Lee, and Joanne Yoong for their support; and the Center for Latin American Studies and Undergraduate Research Opportunities for their sponsorship. He would like to thank his mentor David McKenzie.

Anthropological Sciences

Examining the Extinction of the Pleistocene Megafauna

During the Pleistocene, the world saw a dramatic number of extinctions of very large terrestrial species. The losses of these megafauna have been attributed to either of two different hypotheses. One hypothesis states that global climate changes occurring during the Pleistocene caused environmental pressures that forced the extinction of several megafaunal species. The second hypothesis proposes that the global spread of Homo sapiens *and hunter-gatherer subsistence practices were responsible for these deaths. Examining these two theories reveals that neither climate changes nor human overkill were likely to be individually responsible for the Pleistocene extinctions. Instead, a synthetic model that includes both hypotheses appears to be the most plausible explanation for the Pleistocene losses.*

Robin Gibbons

There are two schools of thought regarding to the underlying cause of the Pleistocene extinctions. Both attempt to explain the loss of nineteen genera and over fifty species in Australia and more than seventy genera in the Americas—almost all terrestrial animals, and virtually all large in size (Martin 1984). The North American fauna saw the extinction of "elephants, horses, camels, ground sloths, all but one pronghorn, several ovibovids, peccaries, and the giant beaver" (Guilday 1967:125). South America also lost elephants and sloths, in addition to horses, some camels, glyptodonts, and a rhinoceros-sized giant rodent species. Extinctions in Australia included all herbivorous marsupials larger than modern day kangaroos (Guilday 1967). The first theory states that global climatic changes at the end of the Pleistocene were responsible for the demise of so many large terrestrial animals. Conversely, the overkill hypothesis proposes that the extinction of large land animals was concordant with arrival of modern *Homo sapiens* in the New World and Australia and that man-the-hunter brought about these deaths. However, neither theory has managed to dominate the debate over the cause of the megafaunal extinctions. Instead, the possibility exists that both explanations are simultaneously plausible. Therefore, it follows that studies should move beyond these individual arguments and start formulating multi-causal models that incorporate both climate and overkill into their reasoning.

Climate Changes

During the late Pleistocene, the retreat of the Wisconsin ice sheet caused global climate changes and changes in local environments. Temperatures became less homogenous, as winters became colder and summers became hotter. Essentially, seasonality increased. In addition, rainfall became more variable depending on the season, with distinctions between wet and dry seasons (Kiltie 1984). During the deglaciation, many streams in the glacial floodplains experienced net degradation and incision of their channels, and the water tables lowered, causing low order streams to become sporadic and transient and springs to dry up or significantly reduce discharge (Haynes 1984).

As a result of this climatic shift, several changes occurred. Some primary habitats were eliminated, while others that may have been only marginal during the existence of the Wisconsin ice sheet expanded (Guilday 1967). Thus, some animals experienced range expansion while others saw their habitats shrink or disappear. In addition, new competition arose between formerly noncompetitive herbivores as a consequence of the changing environments (Guilday 1967). Selective evolution will favor one species in a direct competition and drive the other to extinction. Because of their large size, competition affected the megafauna more than the smaller existing species. Larger animals generally require a larger primary habitat, more forage, and more water to maintain themselves than a smaller species. Therefore, there are usually only a few of them in a given area. As Grayson says, "the inverse relationship between body size and population size plays a powerful role in increasing the risk of extinction faced by larger animals" (Grayson 1991: 214). Following Guilday's logic, shrinking habitats forced large animals into cohabitating smaller areas, containing insufficient resources to support them all. The result was the elimination of some species due to competition.

There are several possible reasons underlying extinction in the face of environmental pressure brought about by

climate change. Overspecialization (inability to adjust to a new environment), competition with a superior species, inability to migrate to a more suitable habitat, and extremely rapid environmental change (impossibility of evolutionary adaptation) will all cause extinction of a species (Guilday 1967). These problems are magnified in shrinking habitats and will force the extinction of larger species first. This is not to say that smaller species are insensitive to climatic and environmental change. Rather, smaller specimens require smaller amounts of resources and can exist in smaller habitats. Therefore, they will not feel the effects of a shrinking habitat until well after the larger species have started to deal with the environmental pressures. The selective pressures brought about by environmental changes caused the extinction of several large species of land animals. Additionally, the minor halts and readvances of the Wisconsin ice sheet during this period of extinction (between 18,000 and 11,000 years ago), were insufficient to restore the previous environmental balance and merely temporarily slowed the inexorable extinction of several species under selective stress (Hester 1967).

A corollary argument to the climate hypothesis concerns gestation length. Slaughter (1967) proposes that the inability of certain species to adjust their mating habits to the changing climates of the late Pleistocene brought about their extinctions. These species tended to be larger, with longer gestation periods. Different species mate at different times of the year depending on either direct cues or correlated cues (Kiltie 1984). Those species that possess a gestation period shorter than the favorable season for producing offspring determine when to reproduce using direct cues—conditions that make birth favorable, such as temperature and food availability. Conversely, those species with gestation periods longer than the favorable season for producing offspring initiate mating using correlated cues. Correlated cues do not affect reproductive success, but rather, correlate with conditions that directly influence the survival of the mother and her offspring, such as temperature, rainfall, or photoperiod (Kiltie 1984). Larger species generally use correlated cues because their increased size lends itself to increased gestation periods. Correlated cues influence the time of mating such that the offspring will arrive during the favorable season, given that the seasonality does not change.

However, during the late Pleistocene, seasonality increased. Unpredictable environment skews the correlated cues such that the reproductive rates of larger species decreases. Therefore, the increasing seasonality of the late Pleistocene would increase the risk of extinction of these larger species due to their decreased birth rates. Kiltie (1984) also states that uncertainty in the length of gestation period and timing of birth with respect to season can reduce fitness. This uncertainty reduces the reproductive success in seasonal environments because the favorable season is more limited and survival of the offspring is directly related to whether or not birth occurred during the favorable season. Larger species are more affected by uncertainty because of their increased gestational period and decreased favorable season (offspring need sufficient time to reach a favorable size for upcoming harsh weather). Additionally, Kiltie (1984) says that reproductive rates decreased during the increasing seasonality of the late Pleistocene because time was lost synchronizing reproductive cycles to the changing environmental cycles. The "lost" time should have been used to produce offspring, but was instead wasted waiting for favorable breeding cues. This problem is magnified in large species with long gestation periods because more time is lost. Reduced reproductive success in these large species precipitated their extinction because their intrinsically smaller populations are much more sensitive to decreasing numbers of offspring. If insufficient offspring are produced to carry a population, it will disappear.

Another climate-related argument was made by Gingerich (1984). He states that the late Pleistocene extinctions are merely a response to a high rate of species origination in the early Pleistocene because "origination and extinction are in dynamic equilibrium constrained by environmental diversity" (Gingerich 1984:216). During the early Pleistocene, originations far outnumbered extinctions, thus suggesting a high level of environmental diversity. The glacial period exhibited high environmental heterogeneity, and therefore, supported diverse fauna (Gingerich 1984). However, during the retreat of the Wisconsin ice sheet and the deglaciation of the landscape, environmental heterogeneity decreased, rendering a high level of faunal diversity unsupportable. Thus, Gingerich (1984) makes the argument that the high rate of extinction during the late Pleistocene was a natural equilibration response to the previous high rate of origination during the early Pleistocene.

Human Overkill

In addition to the climatic changes occurring in the late Pleistocene, modern *Homo sapiens* hunter-gatherers were migrating across the Berengia land bridge into North and South America. The dates of the extinctions of the megafauna coincide with the Clovis culture in the Americas. Thus, the argument has been made that human hunting caused the Pleistocene extinctions.

According to Martin (1967), the extinctions of the megafauna are not consistent with climatic change because the majority of losses were in North and South America and Australia. In addition, the extinctions were largely terrestrial animals with very few marine or plant extinctions. Martin (1967) also points out that the previous glaciation cycles were roughly equivalent in magnitude to that of the Wisconsin ice sheet in terms of environmental stress, but did not experience nearly the same scale of extinctions. Many of the spe-

cies that became extinct during the late Pleistocene survived the previous glacial cycles, necessitating another explanation. The only unique thing about the Wisconsin glacial cycle is the global spread of *Homo sapiens*.

The extinctions of the late Pleistocene occurred relatively rapidly, within a few thousand years. Additionally, there are very few kill or processing sites in the archaeological record connecting *Homo sapiens* to the lost fauna. Grayson (1984) says that this lack of associations between extinct Pleistocene mammals and humans in North America refutes the overkill hypothesis, but Martin (1984) states that this paucity supports a blitzkrieg model. Blitzkrieg is a "special case of faunal overkill that maximizes speed and intensity of human impact and minimizes time of overlap between the first human invader and the disappearance of native fauna" (Martin 1984: 396). The blitzkrieg model explains the lack of kill sites by reasoning that the extinction of these animals occurred too rapidly to have left much, if any, evidence. Therefore, the uniqueness of *Homo sapiens* in the New World and Australia, coupled with the lack of kill sites in the archaeological record, can be taken as evidence of blitzkrieg human overkill causing the late Pleistocene extinctions.

During the deglaciation process, the general ambient temperature increased along with the net usable moisture over the land. Equability of temperatures and seasonality most likely decreased in the northern and central portions of North America, but remained stable in the coastal areas, southern North America, and the North American peninsula (McDonald 1984). Essentially, areas existed where temperatures and seasonality remained stable. Therefore, animals unable to cope with the climate changes in their habitats might have migrated. In addition, deglaciation caused an increase in continental land area due to the retreat of the ice sheet. Land area positively correlates with species diversity, thus the increased land area could have supported a greater number of species (McDonald 1984). Overall, the climate change was favorable to terrestrial animals due to increased primary productivity of the land and cannot account for extinction. Following this argument, *Homo sapiens* were the only disruptive element in the existing environment of the late Pleistocene. The disruption caused by this new, highly successful predator was enough to exterminate several species of large land animals due to lowered population sizes (McDonald 1984). Given the low reproductive rate of large animals, a diminished population size can lead to inadequate fitness to sustain the population as a whole.

During migration across the Berengian land bridge, *Homo sapiens* brought with them a meat-based diet and new bacteria and parasites. A meat-based subsistence pattern evolved during the Upper Paleolithic in Europe as a response to the frigid environment (Jelinek 1967). This same adaptation allowed *Homo sapiens* to survive during migration across the tundra environment of Asia and Berengia. This predilection for animal protein as the mainstay of diet combined with the inexperience of local prey (large animals in the Americas) with *Homo sapiens* hunting practices led to profligate extermination of these species (Jelinek 1967). Since the indigenous species of the New World had not evolved in the presence of *Homo sapiens*, they also had not developed the natural wariness exhibited by similarly large species in the Old World. Thus, the megafaunal species were easy hunting targets. In addition, *Homo sapiens* served as a transport method for foreign bacteria and parasites during migration across Berengia. The diffusion of new diseases into a population of herbivores unadapted to exposure decimated the indigenous large animal species (Edwards 1967). While disease alone is probably not sufficient to cause extinction alone, the resultant reduction in population size combined with *Homo sapiens* hunting practices and low reproductive rates common to megafauna most certainly caused some populations to shrink to numbers impossible to maintain.

Modeling studies done by Whittington and Dyke (1984) and, more recently, Alroy (2001) lend further credence to the human overkill hypothesis. Whittington and Dyke (1984) utilized the Mosimann and Martin model positing that megafaunal extinction in North America could be caused by a small number of *Homo sapiens* migrating in a "front" across the landscape, leaving little archaeological evidence behind. A "front" occurred when the hunting of the human population caused the extinction of megafaunal species in the inhabited area. The model was run using baseline parameters including initial prey biomass, prey carrying capacity, prey biomass growth rate, initial human population size, human population growth rate, human carrying capacity, and minimal kill, defined as the minimum value of prey destruction rate required to cause simulated megafaunal extinctions (Whittington and Dyke 1984). By assigning values to these parameters based on empirical archaeological evidence, Whittington and Dyke (1984) found that overkill is a plausible explanation for the late Pleistocene extinctions. Alroy (2001) used a computer simulation to model *Homo sapiens* and large herbivore population dynamics in the North American end-Pleistocene. He assumed slow *Homo sapiens* population growth, random hunting that was nonselective and limited, and low maximum hunting efforts combined with parameters culled from published values. Simulations found that *Homo sapiens* growth rate and hunting ability almost always led to mass extinctions, with hunting ability being the most important of all parameters (Alroy 2001). Alroy (2001) concluded that the overkill hypothesis was not only credible, but "an anthropogenic extinction was unavoidable given the facts of ecology and the fossil record—even assuming that human predation was limited and nonselective" (Alroy 2001: 1896).

Utilizing newer dating methods, Roberts (2001) looked at extinctions in Australia in attempt to determine if *Homo sapiens* or climate changes were the underlying cause. He rejected dates calculated by radiocarbon (^{14}C) as unreliable. Using optical dating, which measures the time elapsed since sediments were last exposed to light (luminescence method), and $^{230}Th/^{234}U$ dating, which measures crystallization age of flowstone, he determined the burial ages of 28 sites. From these measurements, he found that extinctions occurred around 46,400 years ago across Australia, thus ruling out any climatic impacts from the late Pleistocene as the underlying cause (Roberts 2001). Since *Homo sapiens* arrived in Australia at least 40,000 years ago (Horton 1984), Roberts' (2001) data concurs with the human overkill hypothesis of extinction.

While hunting large game seems to counter expectations generated by optimal-foraging and diet-breadth models, Hildebrandt and McGuire (2002) see it as an important social aspect in hunter-gatherer society apart from group provisioning. Large-game hunting served as an activity that had the added bonus of increased fitness due to increased mating opportunities and favorable treatment for offspring, in addition to greater access to social networks, political alliances, and political authority (Hildebrandt and McGuire 2002). This view is corroborated by Kaplan and Hill (1985), who contend that sharing behaviors had their origins in increased fitness benefits accrued from certain activities. Given that large game is most conducive to sharing outside the immediate family, it stands that *Homo sapiens* would pursue these animals, even after their numbers had dwindled, because of the perceived benefits. This is an example of "show off" behavior. These aspects of late Pleistocene *Homo sapiens* hunter-gatherer societies would be evident in the archaeological record in the form as specialized tools for hunting big game. An example of these tools includes the fluted points found in Clovis sites in North America. Hildebrandt and McGuire (2002) confirm that the relationship between hunter and hunting weaponry is fundamental to the high-risk, yet high-reward, behavior of hunting large animals. Thus, it is possible to argue that human overkill of megafauna was motivated by the pursuit of fitness benefits associated with hunting large game animals and resulted in extinction.

Discussion

In looking at these arguments, it becomes apparent that they are not mutually exclusive. It is quite possible that climate worked cooperatively with *Homo sapiens* to produce the extinction effect that claimed so many terrestrial animals in the New World and Australia. The changing climate, complete with rising temperatures, desiccation of the environment, and retreat of the Wisconsin ice sheet could have pushed large terrestrial animal populations into dangerously low numbers due to the environmental pressures. Then, the hunting practices of *Homo sapiens* groups, unaware of the imminent dangers of small population sizes for megafauna, could have pushed these animals over the brink, onto the inexorable path towards extinction. In this model of megafaunal extinction, both arguments play a role, yet neither is dominant enough to assume sole responsibility for the late-Pleistocene extinctions. If *Homo sapiens* groups were not present, population numbers might have recovered after an unstable period of adjusting to the new climate and habitat. Conversely, if climate had not already reduced the population sizes of large animals, the relatively small number of migrating *Homo sapiens* arriving in the New World and Australia may not have been sufficient to decimate entire species of megafauna.

The models proposed by Whittington and Dyke (1984) and Alroy (2001) are used to further the hypothesis that human overkill caused the late-Pleistocene extinctions; however, there are some aspects that they do not address. Neither model takes climate change into account, which is negligent considering the widespread acknowledgement that climate was changing during this time period. Given ethnographic data collected from modern hunter-gatherer groups, hunting behavior varies depending on environment. The Whittington and Dyke (1984) model assumed a rapid rate of migration as a "front" that left extinction in its wake. It ascribes the lack of archaeological data for this front to the rapidity with which *Homo sapiens* migrated across the continent, following Martin's (1984) blitzkrieg model. Attributing the lack of associated sites to being caused by a short time frame of extinction and migration is somewhat questionable. Sites might be sparse and lack high densities of accumulated material because of the short time frame, but that is not to say they do not exist. Simply, these sites have yet to be uncovered. If they were uncovered, this would not refute the blitzkrieg model unless they showed evidence of long inhabitance periods (i.e.: high densities of accumulated debris). In addition, Alroy's model (2001) sets the minimum number of people entering a given region at 100. Modern hunter-gatherer group size tends to hover around 25 individuals, especially if they exhibit high mobility like the overkill hypothesis assumes. It is unknown from Alroy's (2001) data what effect reducing group size to 25 individuals would have on the outcome of his model.

It is possible to combine the arguments presented by some of these studies, such that they would support a multifaceted model of extinction. For example, Guilday (1967) proposes that the changing climate in the late Pleistocene caused many habitats to shrink. This reduction in suitable habitat size forced a higher number of large terrestrial animals to inhabit a given area than normally would. While Guilday (1967) argues that the concentration of megafauna in these habitats caused competition such that several species became extinct, this accumulation also served

to concentrate prey for hunter-gatherer groups. A high number of prey in a small area increases the chance for successful hunting. Hildebrandt and McGuire (2002) argue that hunter-gatherer groups pursue large game for the perceived fitness benefits gained by successful hunters. Thus, it makes sense that a higher hunting success rate in a given area due to concentration of large game animals into climate-shrunk habitats would encourage increased hunting. Combined with competition between species, this increase in hunting activities could very well push a species to extinction. This blending of arguments, originally supporting opposing hypotheses, works very well to support combining the effects of *Homo sapiens* with global climate change to explain megafaunal extinction in a multifaceted model.

Whereas a resolution to the debate between the climate and overkill hypotheses may not be possible to reach, a compromise might serve as a better conclusion. The opposing arguments agree on the fact that the late-Pleistocene climate changes occurred at approximately the same time as the arrival of *Homo sapiens* in the New World. There is simply no separating the two events. Since there is no differentiating between the time frames of these two events, it follows that the effects of climate and *Homo sapiens* on the megafaunal populations should not be differentiated either. Both hypotheses seem to have holes that are most logically filled with data from the opposing argument. Therefore, a multifaceted model, incorporating arguments and data from both sides of the debate, makes the most sense in explaining the late-Pleistocene megafaunal extinctions.

References

Alroy, John. (2001). "A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction." Science 292:1893-1896.

Edwards, William Ellis. (1967). "The Late-Pleistocene Extinction and Diminution in Size of Many Mammalian Species." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 141-154. New Haven: Yale University Press.

Gingerich, Philip G. (1984). "Pleistocene Extinctions in the Context of Origination-Extinction Equilibria in Cenozoic Mammals." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 211-222. Tucson: University of Arizona Press.

Grayson, Donald K. (1984). "Archaeological Associations with Extinct Pleistocene Mammals in North America." Journal of Archaeological Science 11(3):213-221.

Grayson, Donald K. (1991). "Late Pleistocene Mammalian Extinctions in North America: Taxonomy, Chronology, and Explanations." Journal of World Prehistory 5(3):193-231.

Guilday, John E. (1967). "Differential Extinction During Late-Pleistocene and Recent Times." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 121-140. New Haven: Yale University Press.

Haynes, C. Vance. (1984). "Stratigraphy and Late Pleistocene Extinction in the United States." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 345-353. Tucson: University of Arizona Press.

Hester, James J. (1967). "The Agency of Man in Animal Extinctions." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 169-192. New Haven: Yale University Press.

Hildebrandt, William R. and Kelly R. McGuire. (2002). "The Ascendance of Hunting During the California Middle Archaic: An Evolutionary Perspective." American Antiquity 67(2):231-257.

Horton, David R. (1984). "Red Kangaroos: Last of the Australian Megafauna." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 639-680. Tucson: University of Arizona Press.

Jelinek, Arthur J. (1967). "Man's Role in the Extinction of Pleistocene Faunas." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 193-200. New Haven: Yale University Press.

Kaplan, H. and K. Hill. (1985). "Food Sharing Among Ache Foragers: Tests of Explanatory Hypotheses." Current Anthropology 26: 223-245.

Kiltie, Richard A. (1984). "Seasonality, Gestation Time, and Large Mammal Extinctions." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 299-314. Tucson: University of Arizona Press.

Martin, Paul S. (1967). "Prehistoric Overkill." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 75-120. New Haven: Yale University Press.

Martin, Paul S. (1984). "Prehistoric Overkill: The Global Model." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 354-403. Tucson: University of Arizona Press.

McDonald, Jerry N. (1984). "The Reordered North American Selection Regime and Late Quaternary Megafaunal Extinctions." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 404-439. Tucson: University of Arizona Press.

Roberts, Richard G. et al. (2001). "New Ages for the Last Australian Megafauna: Continent-Wide Extinction About 46,000 Years Ago." Science 292:1888-1892.

Slaughter, Bob H. (1967). "Animal Ranges as a Clue to Late-Pleistocene Extinction." In *Pleistocene Extinctions: The Search for a Cause*. Paul S. Martin and H.E. Wright, Jr., eds. Pp. 155-168. New Haven: Yale University Press.

Whittington, Stephen L. and Bennett Dyke. (1984). "Simulating Overkill: Experiments with the Mosimann and Martin Model." In *Quaternary Extinctions: A Prehistoric Revolution*. Paul S. Martin and Richard G. Klein, eds. Pp. 451-465. Tucson: University of Arizona Press.

Robin Gibbons

Robin Gibbons is a senior majoring in Human Biology with an area of concentration in Evolutionary Studies. At Stanford, she has been a member of the Leland Stanford Junior University Marching Band and has served as a College Assistant at Freshman-Sophomore College. This is her first submission to a Stanford publication. After graduating in 2004, she hopes to attend medical school. She would like to thank Dr. Teresa Steele for all of her insightful comments and helpful suggestions in the writing of this paper.

Medical Anthropology

Presumed Consent for Organ Donation
Perspectives of Health Policy Specialists

In contrast to the growing organ shortage in the U.S., other countries' use of a presumed consent policy has increased donation rates. There is currently a lack of research addressing the political feasibility of a presumed consent policy. This study mapped the terrain of opinions among health policy specialists regarding presumed consent as an organ donation policy. Fifteen health policy specialists from Congressional and Senatorial offices on committees that deal with legislation related to healthcare were interviewed. Analyses were conducted using the qualitative methodology of grounded theory and constant comparison. Study participants' responses were clustered in six categories: semantics of 'presumed consent;' semantics in policy making; 'specified refusal' as a replacement term; political feasibility of presumed consent; factors influencing the outcome of presumed consent; and objections to presumed consent. This study found that responses to 'presumed consent' as a label were negative, such that renaming the policy, perhaps to 'specified refusal,' should be considered. The findings also suggest that methods of public education and communication (especially the media response) may influence the feasibility of passing presumed consent legislation. The results of this study additionally indicate that a presumed consent policy may have more political feasibility on a state rather than federal level.

Kenneth Gundle

Every day in the United States, 17 people die waiting for an organ transplant. The number of people on the waiting list for an organ has more than tripled over the last ten years and now exceeds 82,000 individuals; at the same time, the number of donors has remained relatively stagnant (OPTN, 2004). In 2001, 6,439 people died while waiting for a transplant, nearly double the 3,916 candidates who died while waiting just five years earlier in 1996 (OPTN, 2004; The Lewin Group, 2000; Wright, 1998). Neither the waiting list nor the number of patients who are dying each year is showing any sign of decrease – instead, there is only an increasing organ shortage crisis in the United States.

There has been much debate on the influence of policy on organ donation rates. While the United States operates under a system of 'expressed volunteerism,' with consent received from donors and their families, other countries such as Spain, Belgium, and Austria use an 'opt-out' policy frequently referred to as 'presumed consent.' Although the exact implementation varies, the general concept is that individuals opposed to donating their organs may list their objection on a national registry, rendering them ineligible to donate. People who do not register are considered eligible to be donors unless their family specifically objects.

Organ donation rates in Spain, Belgium, and Austria suggest that presumed consent might have a positive effect on rates of organ donation. In 1999 Spain had an organ donation rate of 33.6 donors per million people (a 142% increase in ten years) while in the same year, the United States had 21.8 donors per million (Matesanz et al., 1996; Matesanz and Miranda, 2000). Spain's success was not necessarily the result of presumed consent, as its infrastructure and education systems greatly improved during that same time period, but the effect of Spain's policy change to presumed consent may have been an important factor. After Belgium passed presumed consent legislation in 1986, its donation rates also rose dramatically (Michielsen, 1996). A frequently cited example is that of two similar transplant centers in Belgium – one in Leuven and one in Antwerp. Leuven switched to presumed consent with the passage of the law and in three years, its donation rate climbed from 15 to 40 donors per million, while Antwerp did not change its policy and only maintained previous levels (Kennedy et al., 1998). In Austria, presumed consent became law in 1982, and by 1990, the rates of donation had quadrupled, to the point where the number of patients awaiting kidneys nearly equaled the number of kidney transplants performed (Gnant et al., 1991). Implementation of presumed consent was followed by increased organ donation rates in these three countries.

Higher rates of organ donation not only result in saved lives, but frequently in saved financial resources. Spain estimates that its 10,000 renal transplants save approximately $207 million every year (López-Navidad et al., 2002). Compared to dialysis, transplanting a kidney is beneficial both in quality of life for the patient and in money spent. In the United States, there are currently over 50,000 people on the waiting list

for kidney transplants, which potentially represents a large savings in healthcare expenditures.

Although studies have discussed both the ethics and the merits of presumed consent in the United States, there is a lack of research addressing the political feasibility of the policy (political feasibility being the possibility of a policy proposal actually becoming law). Efficacy alone will not determine if presumed consent can be adopted in the United States. Understanding the cultural, societal, and political climate is also necessary to determine a solution to the rising shortage of organs.

This study mapped the terrain of opinions among health policy specialists in Washington, D.C. regarding presumed consent as an organ donation policy. Study of the perspectives of members of this community can lead to a better understanding of the political feasibility of an 'opt-out' organ donation policy for the U.S.

Methods

Participants and Procedures

A staff member involved with health legislation from the office of each Member of Congress from the Subcommittee on Health and the Environment (part of the Energy and Commerce Committee, 31 offices total) and from the office of each Senator from the Health, Education, Labor and Pensions Committee (21 offices total) was contacted requesting participation in the study. Of the 52 offices contacted, 14 did not respond and 18 declined participation. The remaining 20 offices agreed to be interviewed, though at the time of interview, five did not consent to the informed consent sheet and were not interviewed. Two reasons were given for declining participation: an office policy not to participate in any form of research (22 offices); and a lack of a staff member focused on health issues (one office).

These offices were selected because they work with the committees where legislation relating to organ donation would likely be submitted and reviewed. These offices were also likely to have staff members who focus on health policy. Participating staff members had the title of Legislative Director (1), Legislative Aide (13), or Legislative Correspondent (1). All focused on health issues alone or healthcare along with one or two other topics.

Semi-structured interviews were conducted with each participant following a field guide, and participants were encouraged to comment freely about their opinions. Interviews lasted between 30 and 90 minutes, and field notes were written during the interview and immediately afterwards. Confidentiality was assured. This study was approved by Stanford University's Institutional Review Board.

Analyses

Transcripts of field notes from the fifteen interviews were read and reread. Statements pertaining to presumed consent were excerpted. These statements were analyzed together using hard-copy analysis[1] and NVivo qualitative analysis software and hard-copy analysis. An experienced team of qualitative researchers used a constant comparative methodology. This grounded approach sought to 'map the terrain' of study participants' perspectives on presumed consent as an organ donation policy. Study participants' statements about presumed consent clustered in six categories.

Results

The six categories that the excerpted statements reflect specific issues repeatedly mentioned in the interviews. For each category, citations are listed to give a sense of the nature and range of participants' stated perspectives.

Terminology of 'Presumed Consent'

The most striking theme that emerged was how these staff members who work with health policy responded negatively to 'presumed consent' as a term. Some informants considered the term to be a major barrier to the policy's acceptance by the public. The term was associated with assuming an individual's choice and the taking away of rights.

Citations:

'Presumed consent seems coercive.' 'It sounds like it takes away a patient's rights, period.' 'Presumed consent has negative connotations. It sounds like something is being taken away. It sounds like the opposite of informed consent.' 'Presumed consent sounds like you're thinking something about me without asking me.' 'I wouldn't use it in my media.' 'Presumed consent is a dicey term on the continuum of stem cells.'

Importance of Semantics in Policy Making

Related to the previous category, informants indicated that semantics play a large role in the passing of legislation. All responding participants agreed that how a policy is perceived is critical. Other examples of recent issues where naming has mattered were mentioned, such as the stem cell debate and medical liability reform.

Citations:

'If you are going to try to get a [bill] passed don't call it presumed consent. Words definitely matter.' 'Names [do] matter on the Hill. Washington, D.C. is all about semantics.' 'Terms matter in a debate.' 'The worst outcome would be if [presumed consent] were villainized; it could go the way of cloning in a debate.'

Specified Refusal as a Replacement Term

Informants were asked if the term 'specified refusal' would be better than 'presumed consent.' While the policy specialists thought the term accurately described the organ donation policy, responses varied significantly among interviewees as to whether the term "specified refusal" would be better.

Citations

'Specified refusal is too cold. It needs something more.' 'It sounds more clean and less hairy.' 'It doesn't draw the same immediate suspicions, but it is splitting hairs.' 'Presumed consent is lost on the public, but it is probably as good as it can get."

Political Feasibility of Presumed Consent

Interviewees' comments cast doubt on the chances of presumed consent legislation being passed at the federal level. Most informants indicated that federal legislation is not currently a possibility, and suggested that this issue may be better addressed as a state ballot measure. Some thought that organ donation policy is a state issue and should not be legislated by the federal government.

Citations:

'It is theoretically a good idea… but half the people in Congress wouldn't go for it.' 'It has no political feasibility federally. It should be done state-by-state, but organs fly all over the U.S., so it would be complicated.' 'It should be done as a ballot measure.' '[Legislation would] bring donation to the national forefront and help to educate the population, even if the legislation isn't passed.' 'It is worth doing in a state ballot. It can't go zero to 100 percent on the federal level…. For it to work on the federal level you would need a crisis beforehand.' 'No, it isn't politically feasible on the federal level. Organ donation is more of a state issue. It might be feasible on the state level.' 'I can't imagine it going through federally; it is too big a step. If a state were to pass a bill first, then federally we could look at it again.' 'I also see this as a States' Rights issue.'

Factors Influencing the Outcome of Presumed Consent

Many interviewee comments dealt with factors that might alter presumed consent's political feasibility and its acceptability by the public. These statements fit into two subcategories: education and communication.

Education and Knowledge of Presumed Consent

There was disagreement between informants regarding the effects of education and public knowledge of presumed consent on the policy's acceptability and effectiveness at increasing donation rates. While some felt that presumed consent would only raise donation rates if the public were uninformed, others believed education is critical to avoid a backlash by the public against the policy. Also mentioned was the possibility that public outcry and media attention itself would actually lead to greater awareness and higher donation rates as the public learns about the organ shortage.

Citations:

'Without education there would be public outcry and a backlash.' 'Some people would object but the outcry would provoke discussion and raise awareness.' '[Presumed consent] would increase donation if people don't know about it, but the numbers would stay the same if people are knowledgeable.' 'Maybe we should go into schools to educate people about organ donation. Presumed consent could backfire without education.'

Media and Communication Affecting Presumed Consent Acceptance

Considering the extent of media coverage in politics and healthcare, it is not surprising that informants repeatedly suggested that the nature of presumed consent's media coverage would be important to how the public responds. Additionally, it was suggested that healthcare policy requires good publicity and marketing.

Citations:

'Where I'm from, people don't like big government. But whether or not presumed consent would work would depend on the publicity.' '[Political feasibility] would depend on the communication campaign.' 'Presumed consent would have to go through a lot and would need media and public support.' 'It would need good media coverage.'

Objections to Presumed Consent

Respondents frequently suggested problems they had with presumed consent, and the responses were quite varied. Objections ranged from concerns about exploiting minorities, to respecting religious and cultural beliefs, to issues of spending and individuals' inherent rights.

Citations:

'I don't want a system that leaves out the exception to the rule. Presumed consent is fine if opposition is allowed, but what if English is not the first language?' 'I don't think that's right; it's unethical. People have religious, cultural, and ideological beliefs.' 'There are rights you are entitled [to] regardless of registry. People in our district don't even pay their federal taxes and don't think the government should know their income.' 'Why not just spend the money to educate people about the current system?'

Discussion

Methodological issues

The informants were selected because they worked for committees that would be the first to deal with legislation related to organ donation policy. Many offices chose not to participate in this study, frequently due to office policy against participating in research. Because of this policy, the number of respondents was reduced, and it is unsure if the responding offices have differences that result in biased results. Republican and Democratic offices were represented from both the House and Senate committees.

The interviews were conducted by a single interviewer who followed a semi-structured interview field guide. The analyses of the interviews were conducted primarily by the author, who

conducted the interviews, but the categories were determined by a team of researchers using a constant comparative methodology. There is still a certain risk that some phenomena were exaggerated and others overlooked, though working within an analysis team should have minimized this possibility.

Terminological Issues in Organ Donation Policy

The results indicated that health policy specialists in Washington, D.C. had significant problems with 'presumed consent' as a term. Informants indicated the term had negative connotations, sounded coercive, and would be difficult for the public to understand. 'Specified refusal' received mixed responses as an alternative to 'presumed consent'. Some considered this terminology such a problem that continuing to name this 'opt-out' policy presumed consent might prevent its adoption as policy.

Without exception, participants believed that 'words matter' in the context of policymaking. Additionally, positive media coverage and the right public education were labeled as factors influencing whether or not presumed consent would be accepted and passed into law. These findings suggest a significant change in how addressing the organ shortage from a policy standpoint should be approached.

Having a policy that can potentially increase rates of organ donation is not enough in and of itself. These informants, who are involved in the creation of health policy, suggest that semantics and marketing are just as important as efficacy in influencing political feasibility.

Federal Political Feasibility of Specified Refusal

Many of the responses suggest that an 'opt-out' organ donation policy would not be passed as federal legislation. Some respondents felt it would be too great a step, while others were concerned that organ donation policy is not the responsibility of the federal government. However, many suggested that such a policy would be much more acceptable at the state level, and more specifically as a state ballot initiative that the public itself would vote on. Informants felt that regardless of such a bill's passage, the discussion that such a vote would generate could broadly increase awareness and donation rates.

Implications of this Study

This study has identified several issues for consideration by those seeking to increase organ donation rates through policy changes, particularly through any form of presumed consent. The terminology describing this organ donation policy may need reevaluation. This study's results indicate potential negative consequences of using the term 'presumed consent;' a different label and a change of title, perhaps to 'specified refusal,' needs to be considered. Also, although literature exists on the effectiveness of an 'opt-out' policy, research is needed regarding how best to educate the public about new organ donation policy and how to communicate effectively with the media. Informants in this study felt that both education and communication could strongly affect the outcome of presumed consent legislation. Beyond efficacy, several factors may contribute to a policy's success, and their importance should not be ignored. Another potential implication is that new types of organ donation policy should first be tried on the state level. Informants expressed resistance to any sudden change, and many suggested that first passing legislation in a state would increase the chance of federal legislation.

Author Notes

* The key features of the data analysis and coding process are (1) Explicitly stating what each excerpt conveys (and convincing other member/s of the coding team that your characterization is correct). (2) Explicitly justifying why a particular excerpt is being placed in a particular category/cluster (and convincing the other member/s of the coding team that your categorization of this item is correct).

References

Gnant MFX, Wamswer P, Geotzinger P, Sautner T, Steininger R, Muehlbacher F. The impact of the presumed consent law and a decentralized organ procurement system on organ donation: Quadruplication in the number of organ donors. *Transplantation Proceedings* 1991;23(5):2685-2686.

Kennedy I, Sells RA, Daar AS, *et al.* The case for "presumed consent" in organ donation. *Lancet* 1998;351:1650-1652.

The Lewin Group, Inc. (for the Office of the Assistant Secretary for Planning and Evaluation). Published January 4, 2000. Analysis of state actions regarding donor registries (25 July 2003). Available at http://www.organdonor.gov/aspehealth.html

López-Navidad A, Caballero F, Cortés U, Martinez J, Solá R. Training course on donation and transplantation for 16- to 18-year-old schoolchildren in the Hosptial de Sant Pau. *Transplantation Proceedings* 2002;34:29-34.
Matesanz R, Miranda B. Expanding the organ donor pool: The Spanish model. *Kidney International* 2000;59:1594.
Matesanz R, Miranda B, Felipe C, Naya T. Continuous improvement in organ donation. *Transplantation* 1996;61: 1119-1121.
Michielsen P. Presumed consent to organ donation: 10 years' experience in Belgium. *Journal of the Royal Society of Medicine* 1996; 89(12):663-666.
The Organ Procurement and Transplantation Network (OPTN). Main page (22 January 2004). Available at: http://www.optn.org
Wright DH. Survey report: Advance directives and donor card effectiveness. UNOS. 1998.

Kenneth Gundle

Kenneth Gundle is a junior majoring in Human Biology with an Area of Concentration in "Biological and Social Issues in Organ Transplantation." He first became interested in organ donation policy during a class entitled "The Nation's Health." The research presented here was supported by a Chappell-Lougee Scholarship and a Haas Summer Fellowship. Readers interested in learning more about presumed consent should contact the author or visit www.presumedconsent.org. Kenneth would like to thank his mentors for their insight and encouragement throughout his research, and his family for their constant support. He would like to thank his mentors Linda Hogle, Stan Wanat, Philip Lee, Geoffery Heller, and Barbara Koenig

Computer Science

Ultra-Fast Matrix Multiplication: An Empirical Analysis of Highly Optimized Vector Algorithms

The development of high-performance matrix multiplication algorithms is important in the areas of graph theory, three-dimensional graphics, and digital signal processing. We present a quantitative comparison of the theoretical and empirical performance of key matrix multiplication algorithms and use our analysis to develop a faster and more modern algorithm. Starting with the mathematical definitions of three classical algorithms (the naive algorithm, Winograd's algorithm and Strassen's algorithm), we derive their theoretical run-time complexity and then compare these theoretical values to each algorithm's real-world performance. We find that Strassen's algorithm, which is the fastest theoretically, suffers a severe performance loss due to its extensive use of recursion. Thus, we propose a Hybrid Winograd-Strassen algorithm that improves the performance of both advanced algorithms. Finally, we discuss the superior performance of algorithms when utilized on modern computational optimizations such as the AltiVec velocity engine, a highly optimized vector processing unit developed by Apple in conjunction with Motorola and IBM.

Boyko Kakaradov

Matrices play an important role in mathematics and computer science, but more importantly, they are ubiquitous in our daily lives as they are instrumental in the efficient manipulation and storage of digital data. Matrix multiplication is essential not only in graph theory but also in applied fields, such as computer graphics and digital signal processing (DSP). DSP chips are found in all cell phones and digital cameras, as matrix operations are the processes by which DSP chips are able to digitize sounds or images so that they can be stored or transmitted electronically. Fast matrix multiplication is still an open problem, but implementation of existing algorithms [5] is a more common area of development than the design of new algorithms [6]. Strassen's algorithm is an improvement over the naive algorithm in the case of multiplying two 2×2 matrices, because it uses only seven scalar multiplications as opposed to the usual eight. Even though it has been shown that Strassen's algorithm is optimal for two-by-two matrices [6], there have been asymptotic improvements to the algorithm for very large matrices. Thus, the search for improvements over Strassen's algorithm for smaller matrices is still being conducted. Even Strassen's algorithm is not considered an efficient reduction as it requires the size of the multiplicand matrices to be large powers of two.

The following two sections present the naive, Winograd's, and Strassen's algorithms along with discussions of the theoretical bounds for each algorithm. We then present a confirmation of the theoretical study on the running times of the algorithms, followed by the results of an empirical study. In the final two sections, we present an improved Hybrid algorithm, which incorporates Strassen's asymptotical advantage with Winograd's practical performance, and discuss the stunning performance of the AltiVec-optimized Strassen's algorithm.

Naive Algorithm

The naive algorithm is solely based on the familiar mathematical definition for the multiplication of two matrices, as shown in equation (1) on the following page. The lowercase letters on the left hand side of the equation indicate that the individual matrix components are combined in scalar multiplication. To compute each entry in the final $n \times n$ matrix, we need exactly n multiplications and $n - 1$ additions. And since each of the n^2 entries in the first matrix (A) is multiplied by exactly n entries from the second matrix (B), the total number of multiplications is $n \times n^2 = n^3$, and the total number of additions is $(n - 1) \cdot n^2 = n^3 - n^2$. Thus, we classify the naive algorithm as an $O(n^3)$ algorithm, because its running time increases as the cube of the given parameter n, so doubling the size of n will increase the number of multiplications eight-fold. While the algorithm performs n^3 scalar multiplications, the input itself is on the

order of n^2. So, the naive algorithm has an almost linear running time with the size of the input.

Advanced Algorithms

We present two famous algorithms for matrix multiplication which use a reduced number of multiplications and therefore run faster. Winograd's algorithm achieves a run-time complexity of $O(n^3)$; that is to say, as n increases, it approaches the run-time of the naive algorithm. On the other hand, Strassen's algorithm uses a peculiar property of sub-matrices to reduce the run-time complexity to approximately $O(n^{2.78})$, which means that for a large enough n, Strassen's algorithm should theoretically perform significantly faster than either of the two previous algorithms. Even though Winograd worked to improve Strassen's algorithm by reducing the number of sub-matrix additions from 18 to 15 [5], he also developed a matrix multiplication algorithm of his own, earlier than Strassen.

Winograd's Algorithm

As described in [3], Winograd's algorithm trades multiplications for additions, much like Strassen's method. However, it is asymptotically the same as the naive algorithm. Instead of multiplying individual numbers as in the naive algorithm, Winograd's algorithm uses pairwise multiplication of couples of entries and then subtracts the accumulated error. As demonstrated in [3], Winograd's algorithm is defined as shown in equations (2), (3), and (4) above.

Since A_i and B_j are pre-computed only once for each row and column, they require only n^2 scalar multiplications. The final summation does require $O(n^3)$ multiplications, but only half of those in the naive algorithm. Thus, the total number of scalar multiplications has been reduced to $\frac{1}{2}\, n^3 + n^2$. However, the number of additions has been increased by $\frac{1}{2}\, n^3$. Winograd's algorithm is theoretically faster than the

Naive Algorithm

$$C_{i,j} = \sum_{k=1}^{n} a_{i,k} \cdot b_{k,j} \tag{1}$$

Winograd's Algorithm

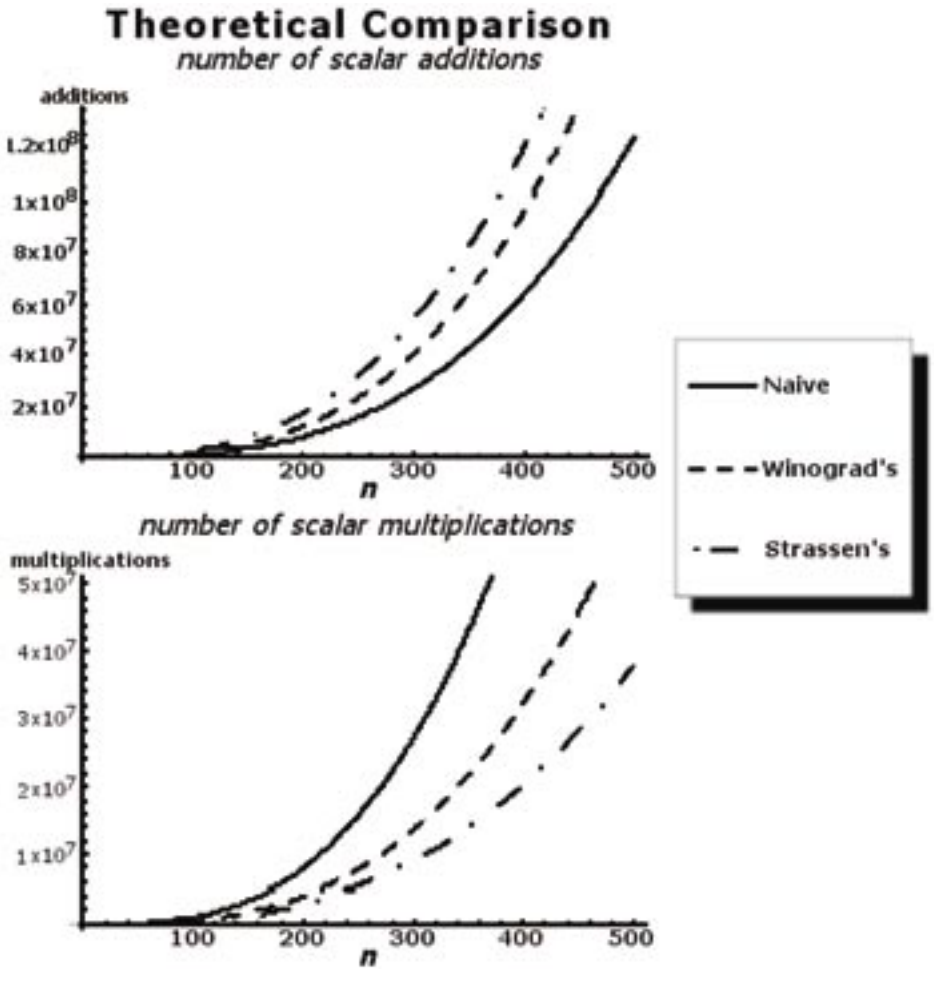
$$A_i = \sum_{k=1}^{n/2} a_{i,2k-1} \cdot a_{i,2k} \tag{2}$$

$$B_j = \sum_{k=1}^{n/2} b_{2k-1,j} \cdot b_{2k,j} \tag{3}$$

$$Ci,j = \sum^{n/2} (a_{i,2k-1} + b_{2k,j})(a_{i,2k} + b_{2k-1,j}) - A_i - B_j \tag{4}$$

Strassen's Algorithm

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \times \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix} \tag{5}$$



naive algorithm, because computers add faster than they multiply (just as humans do), while the total number of operations remains almost unchanged.

Strassen's Algorithm

Strassen devised a clever method of dividing the given matrices into four sub-matrices and then recursively multiplying them to obtain the resultant matrix. This method is known as "divide-and-conquer" and adds not only elegance but also improved theoretical performance. First, the two matrices are divided into four quadrants, or submatrices, of size *n/2* by *n/2*. Now, via a series of intermediate variables, Strassen's method uses only seven multiplications and eighteen additions for each recursive call in order to compute the final matrix product.

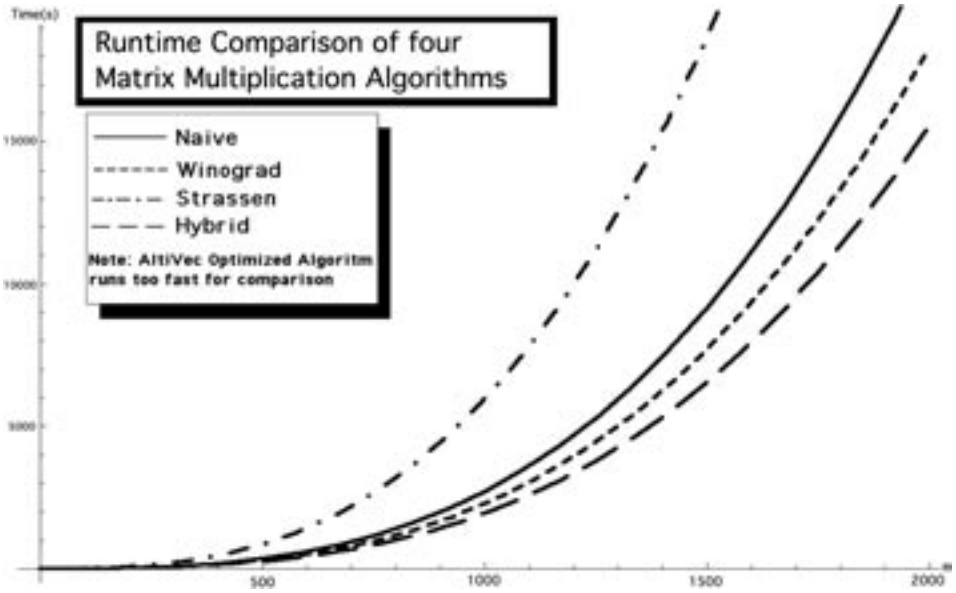
Since Strassen's algorithm is recursive, we can solve a set of recurrence relations for the number of scalar multiplications and additions. It is relatively straightforward to show that the number of multiplications for an *n*-by-*n* matrix is $n^{2.807}$, and the number of additions is $(4.5)\, n^2 \log_2 n$.

Unlike the Empirical Study (which is the next topic of this paper), the theoretical results were consistent with the asymptotic models of the algorithms. The naive algorithm always had the lowest number of scalar additions with the largest number of scalar multiplications. The two graphs on the previous page compare the three algorithms in terms of the number of additions and the number of multiplications as a function of the matrix parameter *n*.

Empirical Study

The naive and Winograd's algorithms were implemented very easily. The empirical tests agree with the theoretical derivations made earlier. Both algorithms appear to run as the cube of *n*, but Winograd's algorithm is faster by a constant because of the tradeoff between additions and multiplications. The graph below illustrates the comparative run-times of the $O(n^3)$ algorithms.

However, Strassen's algorithm presented problems because it is based on recursion. Initially, we manually copied the input matrices into the respective quadrants, a measure which was very time consuming. To circumvent this performance bottleneck, we needed to perform all operations on the input matrices themselves by keeping track of the partitioning. Despite optimizations of Strassen's algorithm, its running time decreased by only 25%, and despite its lower theoretical run-time complexity of $O(n^{2.807})$, it still took three times longer to execute than the $O(n^3)$ algorithms. This is a result of the large number of stack operations caused by the extensive recursion of Strassen's algorithm, which requires large addressing headers in order to contain all four sub-matrix pointers. As we see in a later section, using optimized code with 128-bit headers significantly improves the actual running time of Strassen's algorithm.



For the purpose of these performance comparisons, we conducted tests primarily on a personal laptop computer—an Apple iBook G3 500MHz with 384MB of RAM. The figure below presents the running times (in seconds) of the algorithms as a function of the matrix parameter *n*. This figure clearly shows the polynomial behavior of the algorithms. Note also that these specific run-time approximations depend on the processor speed and so are specific only for the iBook computer. While they serve as a good graphical comparison, they also allowed us to predict the points (values of *n*) of intersection where Strassen's algorithm becomes faster than the $O(n^3)$ algorithms. Looking at the combined graph for all three algorithms, you can see that the intersection points will be for very large *n*. The value of *n* for which Strassen's algorithm (in this implementation) becomes better than the naive algorithm is $n = 60{,}750$. In fact, it would take the iBook computer 19.5 years to confirm this result. The intersection point for the Winograd algorithm is even more terrifying, as you can see from the graph, from which it becomes evident that Winograd's algorithm is best for reasonably-sized matrices (i.e. *n* between 2 and 60,000).

Hybrid Algorithm

Strassen's algorithm performs more slowly than expected in the empirical tests mainly due to the large number of recursive calls, which theoretically should not take up too much time, but in fact account for most of the running time as seen on the previous graph. In the search for a more efficient algorithm for matrix multiplication, we constructed a Hybrid algorithm, which uses the

Strassen's method until a predetermined cutoff size of the seven sub-matrices, after which Winograd's algorithm takes over. This kind of optimization is common to other computationally intensive tasks such as sorting. It has been found that the theoretically faster quick-sort algorithm performs better in real-world applications when the recursion is stopped and the slower, but simpler, insertion sort is used to complete the sorting. In a similar way, we reduce the number of recursive calls exponentially while still taking advantage of Strassen's reduced number of multiplications.

Empirical tests (the long-dashed line in the comparison figure above) show that the resultant Winograd-Strassen hybrid algorithm performs significantly better than both of its predecessors. While, theoretically, the Hybrid algorithm performs almost $O(n^3)$ multiplications (reduced by the use of Strassen's algorithm), with a smaller scaling constant (due to the use of Winograd's algorithm), it is empirically faster than the original Winograd's algorithm by about 15%. The critical level of recursion before switching to Winograd is related to the Hybrid algorithm's performance, and both depend primarily on the starting value for n, with a larger n yielding a more significant speedup.

AltiVec Optimization

In 1998, Apple Computer introduced a specialized vector computation unit called AltiVec, which handles multiple vector operations with increased performance, as in the SSE instruction set on Intel processors. In 2000, Apple's Advanced Computation Group [4] arrived at the revolutionary conclusion that AltiVec instructions drastically improve the running time for Strassen's algorithm. According to their results, a 500 MHz G4 processor could multiply two 2500×2500 matrices in about forty seconds.

Our own AltiVec experiment confirms the stunning performance of AltiVec code with an even more drastic result: a multiplication of two random 2500×2500 matrices finishes in less than a second on a 1.25 GHz G4 processor, presenting a sustained performance of 3.45 TFLOPS (Trillion Floating Point Operations Per Second). Note that the same operation takes 130 seconds to complete by a non-AltiVec optimized implementation.

The efficiency of Apple's AltiVec velocity engine results from the fact that it is able to process four multiplications simultaneously by multiplying four-dimensional vectors of numbers instead of individual numbers. This not only speeds up the pace of actual number crunching, but also allows the four sub-matrix references (pointers) to be contained in a single 128-bit header (because each is exactly 32 bits), which practically eliminates the slowdown observed earlier in the non-optimized Strassen's algorithm.

Conclusion

The empirical results of our study lead us to conclude that Strassen's algorithm is an efficient solution for highly-optimized processors such as the G4 with the AltiVec velocity engine by Apple, which most recently implemented fast matrix multiplication in a technical package called AG-BLAST, which significantly accelerates protein and DNA searches used in biomedical research and drug discovery. However, in everyday applications, such as embedded DSP processors in cell phones and digital cameras, where high-performance vector processors are not feasible, the computational bound for matrix multiplication could be effectively lowered by using our Hybrid algorithm.

References

1. Aho, Hopcroft, and Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley: New York, 1974.
2. Weiss, Mark. *Data Structures & Algorithm Analysis in C++*. Addison-Wesley: New York, 1998.
3. Manber, Udi. *Introduction to Algorithms: A Creative Approach.* pp 301- 304. Pearson Education: New Jersey, 1989.
4. Apple Computer, Inc. "Fast Matrix Algebra on Apple G4." http://developer.apple.com/hardware/ve/pdf/g4matrix.pdf.
5. Jacobson, et al. "Implementation of Strassen's Algorithm for Matrix Multiplication." http://www.supercomp.org/sc96/proceedings/SC96PROC/JACOBSON/INDEX.HTM.
6. McLoughlin, Aileen. "Using Computers to Search for Efficient Numerical Algorithms." http://www.compapp.dcu.ie/alumni/newsletter/scientist1.htm.

Boyko Kakaradov

Boyko Kakaradov is a freshman originally from Bulgaria, but who currently lives in Camden, Maine. While he is still undeclared, he is interested in artificial intelligence and nanobiology. He became involved in matrix multiplication through the extension of a high school algorithms class. He would like to dedicate this publication to his former Computer Science teacher, Duncan Innes, for his inspiration and academic support; Mr. Innes lived his lifelong dream of contributing to his students' academic and cultural enlightenment.

Policy Analysis

The Hanford Nuclear Waste Site:

A Legacy of Risk, Cost and Inefficiency

Since the cessation of plutonium production in 1987, the Hanford Nuclear Waste facility has been the site of the largest and most expensive environmental cleanup project in history. Without prior knowledge of the dangers associated with radiological wastes, the Department of Energy disposed of millions of gallons of these wastes directly into the soil and the nearby Columbia River. Though the federal government bears the financial responsibility for the Hanford cleanup, it has often found itself locked in dispute with the local government over the degree of remediation required. While the federal government attempts to strike a balance between personal and environmental safety and economic feasibility, the local government has consistently pushed for more extensive measures. The result of these conflicting interests has been the most expensive and arguably most inefficient cleanup project in environmental history.

Noah Lichtenstein

In 2004, the United States Department of Energy (DOE) will spend over $2 billion on the cleanup of the Hanford nuclear reservation—the nation's largest existing federal nuclear waste storage site (DOE, 2002). Located in southeastern Washington state, the Hanford site serves as a storage site for radioactive waste produced during the Manhattan Project and throughout the Cold War. With limited knowledge of the dangers associated with radioactivity, DOE officials at Hanford disposed of millions of gallons of highly radioactive materials directly into the soil. Today, large quantities of these radioactive substances have been detected in the groundwater beneath the site, contaminating the water that feeds the Columbia River. Faced with this enormous threat to personal and environmental safety, the federal government is now struggling with financing and implementing the estimated 50 year, $60 billion remediation plan—the largest and most expensive environmental cleanup project in history (DOE, 2002)

Amidst a myriad of political, economic, and environmental challenges that accompany the cleanup effort, the federal government is struggling to appropriate funds in a manner that both provides a reasonable reduction of personal and environmental risk and remains economically feasible in a system operating with limited resources. This paper provides a history of the Hanford legacy and an analysis of the conflicting cleanup incentives between the local and federal government that contribute to the project's inefficiency.

Background and Cleanup Challenges

The Department of Energy's Hanford site is a 560 square mile nuclear reservation located in southeastern Washington, 35 miles north of the Oregon border. The Columbia River flows through the northern portion of the site and forms much of its eastern boundary. Approximately 175,000 people live directly downstream, in and around the cities of Kennewick, Pasco and Richland, and the site is located 215 miles upstream from Portland, Oregon. Between 1944 and 1987, Hanford continually expanded its operations and came to play a pivotal role in the nation's defense, accounting for approximately 74 tons of plutonium for the U.S. nuclear weapons arsenal, almost two-thirds of the total plutonium production for use by the federal government (DOE, 2002).

During the course of Hanford's nearly fifty years of production, separation, and purification of plutonium for the nation's nuclear weapons, the DOE estimates that the Hanford site produced approximately 450 billion gallons of liquid waste (GAO, 1998). The vast majority of this waste was released directly into the ground through about 300 cribs,[1] ponds, and unlined trenches, while liquid wastes of varying levels of contamination were often pumped directly into the Columbia River (DOE, 2002). Studies show that these waste releases, though having met the then-existing disposal standards, have led to

the contamination of about 270 billion gallons of groundwater spread over 80 square miles beneath the site (DOE, 2002).

In addition to these releases directly into the soil and groundwater, the DOE is currently storing approximately 54 million gallons of the most radioactive and hazardous wastes[2] in 177 underground tanks, many of which were built in the 1940s to 1960s and have far exceeded their design life of 10-40 years (Hanford Group Inc, 2003). According to the DOE's reports, 67 tanks have leaked over 1 million gallons of highly radioactive wastes, and the DOE acknowledges that these wastes have also contaminated the groundwater that feeds into the Columbia River (GAO, 2003).

The cleanup of the Hanford site is guided by the *Hanford Federal Facility Agreement and Consent Order*, commonly referred to as the Tri-Party Agreement (TPA), signed on May 15, 1989 by the DOE, the EPA, and the Washington State Department of Ecology. The TPA seeks to achieve compliance with applicable federal regulatory provisions, and furthermore, it establishes legally binding deadlines for the completion of specific actions, defines and ranks cleanup commitments, and provides the basis for achieving full regulatory compliance and remediation (WA State Dept, 2003).

Costs of Remediation

From the outset, the institution and operation of the Hanford site was a national project, aimed at providing collective benefits to the entire country through the establishment of a nuclear defense arsenal. Throughout its construction and its nearly 50 years in operation, the Hanford site received complete federal funding and was overseen by the federal DOE. Today, with the focus of the facility having switched from plutonium production to extensive cleanup and remediation, the federal government is still responsible for bearing the entirety of the monetary burden.

Although the figures continue to fluctuate from year to year, complete remediation costs are currently estimated at $50-60 billion (DOE, 2002). Additionally, over the past several years, the costs of remediation for the Hanford site have steadily increased; in 1998, the DOE received $1.07 billion for cleanup programs at Hanford compared with $1.78 billion in 2002 and $1.95 billion in 2003 (Congressional Research Service, 2003). With such soaring costs, the annual process of allocating funds for the site's cleanup is naturally a very complex and hotly contested issue. While the DOE submits a budget in accordance with what it deems as necessary for the fulfillment of its cleanup requirements under the TPA, the amount of money allocated to the DOE by Congress is often much less than requested. This disparity in funding has widespread implications, often making it difficult, if not impossible, for the DOE to meet its "established milestones." For example, while the $1.78 billion allotment for 2002 was an increase of $320 million over the previous year, it still fell $56 million short of what the DOE's calculations indicated was necessary to meet its legal cleanup obligations (Congressional Research Service, 2003). As a result, each of the DOE's divisions responsible for specific cleanup procedures was expected to achieve the same level of remediation with less funding than was reportedly needed.

Faced with growing public dissatisfaction with project inefficiency and spiraling costs, President Bush has pushed to reduce federal funding for the cleanup project in each of the past three fiscal years. However, due to the efforts of a bipartisan nuclear cleanup caucus led by Washington Senator Patty Murray, a member of the Senate Appropriations Committee, the President's funding reduction proposals have been repeatedly thwarted (Webster, 2002). Following this series of intense budget wrangling from 2000-2003, the funding arrangement for Hanford underwent significant changes in fiscal year 2004. The Bush Administration called for the creation of a new Defense Site Acceleration Completion Account which would provide Hanford funding (Congressional Research Service, 2003). As passed by the House and Senate, the conference agreement on the Energy and Water Development Appropriations Act for FY2004 would provide $5.65 billion for this new Account, which is nearly $164 million less than the requested $5.81 billion (Congressional Research Service, 2003).

Analysis of the Problems Associated with the Hanford Cleanup

One of the most frustrating realities for policy-makers involved in the Hanford cleanup is that it is the most expensive, and arguably the least efficient, project ever conducted by the DOE. At a current cost of over $2 billion a year, and likely to take at least another 20 years to complete, the Hanford remediation has been, and barring a miracle, will continue to be, an excessive drain on Congress's checkbook. While a number of factors contribute to this problem, the principal cause is a conflict of motivation between those who bear the costs and those who bear the risks associated with remediation.

The primary bearers of risk in the Hanford cleanup are the members of the local population in and around the cities of Kennewick, Pasco, and Richland. It is the workers from these communities that are employed at the Hanford site and who have the greatest potential exposure to radiation. However, though the local populace is clearly at the greatest risk, there remains in the community a distinct lack of incentive to expedite the cleanup process. The reasoning behind this apparent divergence from rational thought has

been posited by Robert Budnitz, the chairman of the National Academy of Sciences (NAS) former Committee on Buried and Tank Waste (1993-1998). According to Dr. Budnitz, because the dangers accompanying the site are not clear and present, the local citizens feel little pressure to push for immediate cleanup; since nobody can see the direct effects of contamination on the community, the citizens feel little pressure to combat the problem quickly (Budnitz, 2002). In other words, if not enough people are getting sick *today* as a result of contamination, the local community sees little reason why it matters if the site is cleaned up in 3 years or 30 years.

While undoubtedly the people living near Hanford want to see site remediation, the lack of an immediate tangible threat assigns the risk of contamination a lower priority than that of other issues concerning the community. Thus, the decision of whether to expedite or to prolong the cleanup effort flows from a simple evaluation of individual costs and benefits. Prior to the cessation of plutonium production in 1987, Hanford employed approximately 8,000 workers; in contrast, there are currently over 11,000 employees working on the cleanup effort (Budnitz, 2002). In an area that has grown alongside Hanford and has been economically dependent upon it since the beginning of World War II, the shutdown of the facility and completion of the cleanup will certainly bring economic downfall for the surrounding communities. While the remediation efforts are taking place, the local communities have not only sustained their existence but have also thrived, with the cleanup project's creation of over 3,000 new jobs. Without any clear and present danger posed by the site, the finite costs associated with the completion of Hanford's cleanup far outweigh the seemingly intangible benefits of remediation (Budnitz, 2002).

In addition, the local community would like to see the site completely uncontaminated, free of Hanford's footprint, an economically enormous, if not impossible task. In the absence of economic responsibility, the only choice the local government has to make is, "Would we prefer to have the site restored to prior greenfield status or to have large portions of the reservation fenced-off and unusable?" In contrast, Congress, working with limited resources, has to decide what degree of remediation provides the correct balance between reduction of health and environmental risk and economic feasibility.

Though the local communities' incentives for prolonging the cleanup efforts are understandable, they stand in direct conflict with the remediation goals of the federal government. Distinct from the objectives of the local citizenry, Congress's primary aim is to complete the remediation of the Hanford site in both an expedient and cost-effective manner. While safety is publicly stated to be the most important aspect of the Hanford cleanup, the goal of the federal government from a Congressional standpoint is to strike an appropriate balance between the reduction of risk at Hanford and the most economically reasonable remediation plan. However, because of the local benefits that accompany a prolonged remediation, coupled with the federal government's bearing of the entire monetary burden for the project, the local government has the incentive to manipulate the availability of federal funding. With proposed measures being paid for with federal funds, often the policies that the local government judges to be cost-effective for the community are inconsistent with the balance Congress is attempting to strike.

In addition to its attempts to restore Hanford to prior greenfield status, the local population has employed the issue of worker safety as a means for prolonging the cleanup efforts. While the local government pushes for a zero marker in worker exposure, Congress again turns to cost-benefit analysis to find the level of policy implementation that adequately balances worker protection with economic feasibility. However, because the issue at hand deals with personal health risk, as opposed to land use, the local community has achieved a greater level of success in receiving sympathy in the form of federal funds. Because the local community is worried about its citizens' exposure to risk, the reasoning behind conducting these studies is valid. But in practice, the studies have added to the overall inefficiency of the Hanford cleanup. In order to ensure the highest degree of personal safety, the local community has questioned all aspects the DOE's decisions and has insisted that the federal government conduct repeated studies of all cleanup procedures used. In doing so, the local government is ensuring that cleanup measures will come as close to the desired zero standard policy as possible. At the same time, local governments are adding a hitch to the remediation process, resulting in long delays and higher costs (Budnitz, 2002). Working with limited resources, and with annual cleanup costs for Hanford now in excess of $2 billion, it seems logical that Congress would be very concerned with the current level of project inefficiency. However, due to internal politicking and congressional logrolling, Congress has for the most part, failed to address the issue. Because Congress is in charge of appropriating funds, the process of allocating money for the Hanford cleanup is extremely complex, with every dollar specifically assigned. To add to the complexity of the issue, each of these specific allocation decisions is, in turn, influenced by the individual special interests of each Congressional member. Thus, in order to benefit from the proposal, a representative from one state might try to negotiate, promising to vote favorably on this issue on the condition that the waste is shipped to her state's treatment facility, an action that would benefit the economy and please her constituents

(Budnitz, 2002). Unfortunately, a great deal of economic inefficiency often accompanies this system of trade-offs because sending the waste to another state instead might save the government millions of dollars.[3]

Additionally, Congress is also under enormous public pressure to find the proper balance between risk-prevention and remediation cost—if Congress errs in either direction, the consequences can be severe. If a member of Congress is portrayed as being in support of reducing cleanup funds, the effects of such negative exposure could be politically devastating. Few politicians want to take on this risk. At the other end of the spectrum, Congress is expected to allocate funds efficiently, and eliminate wasteful spending. If the public becomes disenchanted with the Hanford cleanup effort and believes that there is a great imbalance between the risks involved and the remediation costs, then policy-makers risk alienating their constituents. As a result, when making appropriations decisions, members of Congress must find an acceptable balance to avoid being portrayed as tight-fisted or as excessive spenders.

Conclusion

After nearly a half century of waste management procedures that threatened both personal and environmental safety, the federal government is now faced with a cleanup task more complex than any other remediation project in environmental history. Not only are wastes that were dumped directly into the soil permeating the groundwater, but also over one-third of Hanford's storage tanks are leaking, adding to the flow of radioactive substances into the groundwater. The risks associated with exposure to these wastes are substantial; the effects of the Columbia River's contamination would not only be devastating to the local community, but would also impact all who live downstream and who depend on the Columbia for their livelihood.

In the face of all these risks, Congress and the local government have often been at odds regarding their goals for remediation. While the federal government wants to ensure the safety of the cleanup workers and the local citizenry, it also has to counter the challenges and criticism made by those who have to help bear Hanford's cleanup costs. At the same time, the local government is pushing to increase the extent of remediation beyond what is economically feasible for Congress. With so much of the local economy tied up in the remediation effort, the local government has a strong incentive to prolong the cleanup procedure.

The Hanford cleanup project has become a platform for Congressional logrolling. The local government and Congress are locked in dispute. Collectively, both parties must also counter the influence of special interest groups trying to benefit from the cleanup. In the upcoming year, with a new and increased budget from the federal government, the TPA stands ready to fulfill its end of the bargain. However, it remains to be seen whether this increase in funding will result in a safer and more efficient cleanup or in a more costly and prolonged remediation.

"Noah Lichtenstein, The Hanford Nuclear Waste Site: A Legacy of Risk, Cost, and Inefficiency, 44 Natural Resources Journal (forthcoming 2004)"

Author Notes

[1] A crib is an underground structure designed to allow liquid wastes to percolate to the soil.

[2] The waste in these tanks contributes about 215 million curies to Hanford's inventory of high-level waste.

[3] Coincidentally, instead of transporting all of its low-level wastes to existing treatment sites, DOE is constructing a new on-site facility specifically tailored for wastes at Hanford.

References

Budnitz, Robert. Chairman: National Academy of Sciences Committee on Buried and Tank Waste. (1993-1997). Phone Interview. 15 May 2002.

Congressional Research Service. Memorandum to Sen. Ron Wyden. "Environmental Cleanup at the Hanford Nuclear Facility." 25 November 2003.

CH2MHILL – Hanford Group, Inc. *Waste Tank Summary Report for Month Ending September 30, 2003*. HNF-EP-0182, Rev. 186. October 2003.

Department of Energy. *Performance Management Plan for the Accelerated Cleanup of the Hanford Site*. DOE/RL-2002-47. Rev. D. August 2002.
General Accounting Office. *Nuclear Waste: Challenges to Achieving Potential Savings in DOE's High-Level Waste Cleanup Program*. GAO-03-593. June 2003.
General Accounting Office. *Nuclear Waste: Understanding of Waste Migration at Hanford is Inadequate for Key Decisions*. GAO/RCED-98-80. March 1998.
Washington State Department of Ecology, U.S. Environmental Protection Agency, and U.S. Department of Energy. *Hanford Federal Facility Agreement and Consent Order*. 89-10 As amended through September 2003.
Webster, Todd. Press Secretary for Washington Senator Patty Murray. Phone Interview. 24 May 2002.

Noah Lichtenstein

Noah Lichtenstein is a senior majoring in American Studies, with a minor in Political Science. He was first introduced to the Hanford Nuclear Waste Site while taking Professor Rosnecranz's Environmental Policy and Law course. While serving as Legislative Intern for Oregon Senator Ron Wyden (as part of the Stanford in Washington Program), Noah continued to pursue his interest in the politics surrounding the cleanup of Hanford. While in Washington DC, he produced this manuscript—the complete version of which will appear in the Summer 2004 issue of the Natural Resources Journal. He would like to thank his mentor Armin Rosencranz.

Film & Cultural Studies

The Afterlife of Melodrama: Roger Michell's *Changing Lanes*

Written for the PWR course "The Rhetoric of Cultural Critique," this essay examines Changing Lanes *to argue that Hollywood films can embody social tensions in unexpected ways. The essay situates* Lanes *in the historical tradition of the film melodrama, with an emphasis on the genre's incarnation in Douglas Sirk's popular spectacles of the 1950's. By historicizing Lanes' visual and narrative conventions, the author tries to read the film "against the grain" to discover how it exposes a "powerful, pervasive ambivalence in our nation" about white treatment of black culture and "ideological insecurity" about whether people's lives are really the product of their choices. The film comes close to achieving a subversive "Sirkian" undercurrent thought to be extinct in today's Hollywood entertainment, but it eventually reaches a level of ideological confusion that surpasses the ambivalence observed in Sirk's features. In its unusual political incoherence, the film ends up in a category by itself, a unique Hollywood post-melodrama that addresses complex questions about racism in a fragmented 21st-century society, an America less defined by the clear racial divisions that Sirk considered in his own time.*

André de Alencar Lyon

In Roger Michell's frenetic road-rage melodrama *Changing Lanes* (2002), we find the hotshot Wall Street lawyer Gavin Banek (Ben Affleck) pitted against the "little guy" Doyle Gipson (Samuel L. Jackson) in all-out warfare: upper class versus working class, white versus black. After their fates suddenly collide in a highway fenderbender, ruining crucial court meetings for both of them, the two are soon vengefully sabotaging each other right and left. Each is prodded in this eye-for-an-eye rivalry by specific personal goals: Gavin needs a vital legal document he has accidentally exchanged with Doyle, and Doyle wants to somehow win back his kids and ex-wife in spite of his past behavioral and financial difficulties. Without any accepted moral standards to guide them within their corrupt, cutthroat, capitalist society, both characters find themselves pushed further and further to get what they want. Soon enough, not surprisingly, they end up teetering on the edge of utter moral dissolution.

The first time I saw the film, I was struck by this unusually bleak and critical vision of our twenty-first century American society, what a *Cinéaste* critic referred to as an allegory of modern American race and class relations (Landau 41). But this vision only holds up compellingly until the film's last few minutes, at which point it is abruptly shattered by a disastrously incongruous denouement. Once Gavin and Doyle finally arrive at their remorseful reconciliation, the filmmakers mop up their narrative mess by having the repentant Gavin usurp his corrupt father-in-law's firm (via "justified" blackmail) and thereby restore the disadvantaged Jackson to his financial stability and his family. We can assume they all live happily ever after. The End. At the time, I strongly concurred with the general critical consensus that the film had been "marred by a too-convenient, feel-good ending that [...] make[s] moviegoers feel cheated [rather than] challenged" (Graham). The director, Roger Michell, helped confirm my opinion when he revealed in his DVD audio commentary that this tidy, optimistic epilogue was in fact not originally part of the script: "I'm still agnostic about whether it's the right way to end the film or not" (*Changing Lanes*). So maybe the film's ending wasn't originally *intended* to be this pat and dissatisfying. "[P]robably the studio sales department sent it to a few focus groups whose members decided it was too much of a bummer" (Landau 41). What at first was essentially bothered me about this ending was simply that it felt like an artistic copout. I gave *Changing Lanes* a "thumbs-down" verdict solely because it failed to conform to my perceptions of what makes a successful film; in other words, it didn't "challenge" me. Of course, the film's failure to "challenge" me did not initially strike me as more than a disappointment for my Friday evening. But then I came across Ryan Gilbey's unique evaluation, and my eyes were opened to some significant and disconcerting social implications linked to this "merely disappointing" happy ending:

"[The film's] narrative actively upholds [the class and racial] divisions. At the end [...] power over black lives still resides in white hands [...] The film

says to its white viewers: [...] [P]lay Santa Claus to a black family in your neighbourhood today. [...] It's a choice that says more about white treatment of black culture [...] than the film-makers can possibly realize" (37) .

Gilbey's assessment leads us to realize that the film's ending is not just a pat, lazy "artistic copout." In fact, the film's resolution seems to actively communicate a racist social message, encouraging white viewers to serve as condescending benefactors to powerless African Americans. Furthermore, it seems to me that this message naturally entails encouraging African Americans to contently acquiesce to this patronage, to rely on help from white people and thereby remain fixed in their repressed, powerless position within the societal status quo.

But I believe Gilbey's assessment takes our understanding of this film's ideological significance only so far. If we turn back to the rest of the film and consider *Changing Lanes* as a whole in view of the resolution's disturbing implications, we can in fact find that the story and visual design operate together with the ending to produce meanings that seem to contradict the racist message Gilbey points out.

In order to sufficiently resolve this complicated concern, I want to more deeply explore *Changing Lanes*' relationship with a genre that has always held a very dynamic and problematic relationship with societal ideologies: the melodrama. According to the conventional understanding of the word "melodrama," *Changing Lanes* would certainly fit the category, with all its overheated emotions and manipulative plot contrivances. Historically, too, the film shares many rhetorical parallels with melodramas made in the 1950's and even with some extending as far back as the Silent Era. Comparing *Changing Lanes* to these specific melodramas can help us more fully understand the film's difficult relationship with our present status quo. By examining, from a historical perspective, *Changing Lanes*' melodramatic rhetorical methods of producing meaning for an audience through its narrative, style, and resolution, I believe we can discover in the film, despite and because of its happy ending, a much more complex variety of possible meanings than Gilbey gives it credit for. The film's unusual multiplicity of social meanings may mark it as a kind of post-melodrama: its ultimate incoherence mirrors the social complexities of a 21st century America, in which the racial divisions that an artist like Sirk had to address are not nearly so concrete and distinct.

To discover these hidden levels in Changing Lanes, it is important to first consider melodrama's historical tradition as a characteristically *conservative* genre, a genre normally *affirmative* of the societal status quo. We must firmly recognize the film's conservative roots before we can understand the way it doesn't necessarily conform to them.

Changing Lanes can best be grasped on this most basic, conservative level as following in the footsteps of one of the silver screen's first incarnations of melodramatic form: the Silent-Era social-problem film. Although they were often produced by progressive labor unions, most of these films, even while calling attention to social problems, actually "chose narrative structures that were conservative pleas for authority or passivity" (Sloan 34). By comparing story structures between *Changing Lanes* and one of these melodramas, we can recognize more precisely how *Changing Lanes*' narrative at least *seems* to bolster a racist social message.

A particularly fitting melodrama for comparison would be the National Association of Manufacturers' picture *The Crime of Carelessness* (1912), a melodrama set in a factory being poorly maintained by its owners for fire-safety regulations. The film of course presents us with a vision of imminent disaster, but this disaster only erupts when a worker carelessly tosses his cigarette into a pile of trash: the factory burns down because of a particular *individual's* mistake. The narrative concern then shifts from the grim reality of factory conditions to the irresponsible worker's sudden ostracism from his community; the film concentrates exclusively on this one worker eventually reconciling with his boss and helping to rebuild the factory. In this way, the original *social* conflict plays out on a uniquely *personal* level. This focus on the individual helps make "inequality or labor exploitation appear to be a universal product of the human condition, an eternal dilemma of life, rather than a problem artificially made by the specific economic or social conditions of industrial capitalism and the Progressive Era" (Sloan 36). Moreover, the film assures viewers that this "eternal dilemma" can in fact be resolved on a personal scale without altering the social system that technically produced the dilemma in the first place. Hence, by "ironically tapping the [audience's] desire for change," *The Crime of Carelessness* diffuses the need for social reform onto appeals for individual responsibility within an "open" society, in which tensions and conflicts can supposedly be resolved through people's noble devotion to moral principles (Sloan 34).

It is certainly reasonable to assume that this trend has survived within the genre up to the present day, and *Changing Lanes* seems to adhere to it, especially in the way it presents Doyle's suffering within the fast-paced, dehumanizing nature of his society. Doyle, too, enacts his own form of revolt, not just against Gavin, but also against the social system that seems to be walling him in. One high point of his revolt is vividly symbolized by his violent desecration of one of the bank's computer monitors: he destroys a metonymic representation of the control his technology-driven society seems to have over his life. Significantly, the motivation for this rejection seems to be deeply rooted in a sense of racial oppression,

as is indicated in a bar scene where Doyle fiercely insults two racist white men. However, *Doyle* is the one, rather than the seemingly racist outside world, who ultimately shoulders the blame for his problems when he's scornfully diagnosed as being "addicted to chaos." The film reminds us, as co-writer Michael Tolkin explains in his DVD interview, that *he* is the one with the alcohol addiction and the volatile temper (*Changing Lanes*). The film holds *him* responsible for maintaining his own dignity and composure, regardless of whatever nerve-wracking (and perhaps racially discriminatory) social circumstances he's up against. Rex Reed highlights this dimension when he asserts, "*Changing Lanes* opens your eyes both to the terrible things ordinary people can do to each other [...] and the ethics with which men reduced to monsters can still find redemption" (par. 4). Indeed, as its Christian imagery emphasizes, the film does seem to position its conflict on a plane of universal human morality. In doing so, it draws attention *away* from how much the social system itself may have contributed to Doyle's predicament. Finally, the film employs what could be aptly labeled a *deus ex machina*, or "narrative rupture" (Heung 319), when it *magically* resolves Doyle's problems through the work of "higher forces" Doyle is never made aware of in the story. In the end, although Doyle remains in his *passive* position within a fundamentally unchanged society, he finally does attain his redemption and contentment even within the confines of his "rightful (race- and class-defined) place in the world." So *Changing Lanes*, like any old-fashioned, affirmative melodrama, "reconcile[s] the suffering individual to his or her social position by affirming an 'open' society where everything [is] possible" (Elsaesser 515). Such a reading supports Gilbey's claim that the film prejudicially upholds the social divisions it depicts: the resolution encourages the individual to passively *endure* the present social reality, as Doyle eventually does, rather than actively *challenge* it.

This reading grows more complicated, however, when we consider *Changing Lanes* next to what Thomas Elsaesser has dubbed the "sophisticated family melodramas" of the 1950's (qtd. in Klinger15). These films, best exemplified by those of melodrama maestro Douglas Sirk, have been classified by Paul Willemen in a unique category originally defined by J. L. Comolli and J. Narboni: "films which seem at first sight to belong firmly within an ideology and to be completely under its sway, but which turn out to be so in an ambiguous manner" (qtd. in Willeman, "Distanciation" 272). Although often limited by the "narrative rupture" of a happy ending these films employed various methods of narrative and stylistic irony in order to subtly "work against" their superficial conservative messages. One could in fact discover scathing cultural criticisms buried beneath the surface simply by interpreting the films' storytelling techniques "against the grain." By examining *Changing Lanes* in conjunction with these specific melodramas, we can discover how the film does indeed operate subversively in spite of its affirmative surface meanings.

Sirk's last film, *Imitation of Life* (1959), would be the most appropriate to compare with *Changing Lanes* for this purpose, as much of its story focuses on an African-American girl named Sarah Jane trying to break out of her disadvantaged social position. In trying to "pass" for white and escape her racial roots, Sarah Jane conducts her own revolt against her conservative mother Annie, who consistently confines Sarah Jane to her racial identity by publicly acknowledging the light-skinned girl as her daughter. Wherever Sarah Jane flees in order to assimilate herself in white society, Annie protectively pursues her there, pleads for her to "come home," and unmasks her true self before the world. Consequently, Sarah Jane is continuously rejected and repressed by the white culture she so desperately wants to fit into.

Imitation of Life represents a bold departure for Sirk because it addresses more explicitly than his other works the horrific reality of social prejudice in 1950's America. In its "remarkable candor about [...] racial double-standards" (McKegney 72), the film's "representation of social contradictions [...] leads [it] far beyond the individual focus of the melodramatic mode" (Flitterman-Lewis 328). However, the picture was not considered subversive upon its initial release: the *Monthly Film Bulletin*, for example, panned the film with the claim that "its attitude toward its racial problem [was] debased and compromised" (qtd. in Klinger 78). The conservative implications of its final scene would certainly account for this reception: as in Doyle's case, Sarah Jane's rebellion is at last *contained* when, at Annie's funeral, she penitently confesses her rejection of her mother, accepts her racial identity, and implicitly resumes her servant position in the white family unit that had employed her mother. "[T]he last shots of the film [...] reinforce [Sarah Jane's] acceptance of the symbolic position of the black woman" (Flitterman-Lewis 329). The film resolves Sarah Jane's racial identity conflict through personal moral repentance, and it implies that her situation will somehow be alleviated when she enduringly resigns to her ascribed servant status. At first sight, then, *Imitation of Life* seems to fail as a forthright indictment of a prejudicial society.

But in spite of its affirmative resolution, the essential whole of the film still operates as an effective cultural critique, primarily because its *form* calls sharp attention to a social reality that's strongly influencing this intense "personal" drama. Through several narrative and visual details, *Imitation of Life* actively works against its conservative surface message to highlight racist social conditions and to lead the viewer to consciously perceive and

question those conditions, regardless of how the plot resolutions may superficially oversimplify them.

The most immediately obvious elements of this progressive dimension involve the film's narrative. Although the film seems to morally blame Sarah Jane for coldheartedly rejecting her mother (and the conservative values her mother represents), certain features of the story clearly present the social forces that push her *to* that rejection: "The fact remains that Sarah Jane [...] is beaten up by her boyfriend and fired from at least one job simply because of her race, and America is shown to make room for her black citizens most readily in kitchens and dressing rooms, her high life only glimpsed from back doors and back alleys" (McKegney 72).

The evidence of Sarah Jane's social confinement speaks quite powerfully for itself. Clearly, the racist structure of her community has irrevocably limited her options and made it essentially impossible for her to actualize herself in any meaningful way. When Sarah Jane's personal "liberation" only leads her to prostitute her image as a dancer in a white nightclub, the film itself does not really blame her for this "ignoble," "irresponsible" outcome of her revolt. In illustrating how she merely debases herself in her rebellion, the film is not necessarily repudiating her revolt as a moral error, but is instead revealing that Sarah Jane truly has "no exit," as Sirk puts it, from this cycle of social oppression no matter what she does (Sirk 119). Because "[t]he social pressures are such [...] that the range of 'strong' actions is limited," Sarah Jane is unable "to act in a way that could shape the events [of her scenario], let alone change [her] stifling social milieu" (Elsaesser 524). Consequently, this narrative indication of her impotence subtly "reproduc[es] the patterns of domination and exploitation existing in [American] society, [...] emphasizing [...] an emotional dynamic whose social correlative is a network of external forces directed oppressingly inward" (Elsaesser 532).

In *Changing Lanes,* Doyle seems to be in a very similar position. For one matter, the demanding, fast-paced dynamics of his urban community seem to relentlessly push him into a corner: frenzied traffic and strict, institutionally imposed time constraints are technically what deprive Jackson of the chance to win visitation rights over his kids. More important, however, is the film's potent suggestion that Doyle, like Sarah Jane, is individually at a unique social disadvantage because of his *race*. This implication arrives most forcefully in the aforementioned bar scene, in which Doyle's prolific and heated insult toward the two men ends up igniting a brief violent conflict. Although Doyle can be held responsible for initiating this conflict, the riled men are the ones who first approach Doyle and threaten him. Tolkin may describe this scene simply as another illustration of Doyle's explosive personality (*Changing Lanes*), but the fact remains that Doyle is unwillingly *placed* by these men in an antagonized position, giving him little choice but to react. The scene doesn't call attention to Doyle's unstable temper as much as it reveals Doyle's subjection to an outside racist threat. Moreover, the harsh, calculated intensity with which Doyle delivers his insult in the first place indicates that he must have very deep-seated frustrations over racial prejudice on a much broader level. It is reasonable to infer from the emphasis placed on this scene that this is far from his first encounter with racism in his social milieu.

Indeed, this scene is *dwelt* on so deliberately that it leads us to seriously consider how large a role race actually plays in Doyle's predicaments and other social interactions within the story. One of the first and most obvious examples of this racist undercurrent can be detected in the scene with the white divorce-court judge, who's blatantly unsympathetic to Doyle's pleas and explanations for his delay. After Doyle and his family leave the courtroom at the end of the scene, the camera tellingly lingers over a final shot as the next client, an elderly white woman, is politely welcomed by the same judge. Doyle might possibly be receiving a much lower order of treatment from his society simply because of his race. This possibility is corroborated by a significant pattern sustained throughout the story: the oppressive figures in power over Doyle are always white. Gavin, the hacker, the bank clerk, the elementary-school guards, and even Doyle's rather preachy, denunciatory Alcoholics-Anonymous counselor—all are white. As much as the narrative explicitly blames Doyle for his volatile personality, all these hints of racism indicate that the film is also implicitly justifying his revolt, revealing that there are indeed harsh and constricting social injustices he has to contend with. Like *Imitation of Life*, *Changing Lanes*' narrative reveals through the *failure* of the protagonist's revolt the limitations his society imposes on his ability to take "strong action," to confidently take control over his life.

But how can we be further convinced that the film is actively encouraging this progressive reading? To more solidly confirm the progressive dimension as a strong element in the text, it is useful also to examine how the film's melodramatic visual design emphasizes these subversive narrative implications.

A consideration of *Imitation of Life*'s style can lead us to recognize this progressive aspect of *Changing Lanes*' style, as *Changing Lanes* can be interpreted to draw at least indirectly from Sirk's uniquely expressive use of cinematography and melodramatic mise-en-scène. Because Sirk's protagonists often cannot vent their pent up emotions through "strong action," his melodramas employ "a sublimation of dramatic conflict into décor, color, gesture, and composition of frame, thematized in terms of the characters' emotional and psychological predica-

ments" (Elsaesser 521). By dramatizing the characters' repressed subjectivities this way, Sirk's films vividly articulate how cultural pressures and expectations take a profound toll on the individual psyche. *Imitation of Life* exemplifies this technique perfectly, as it uses "off-kilter angles, unnaturalistic lighting, character (and camera) movement and dynamic set design [as] visual equivalents of the repressed cultural tensions of the 1950's" (Butler 299). For its specific focus on the race issue, the film also uses style to illustrate the firm social divisions contributing to Sarah Jane's predicament:

"Sirk designs a color system in the film which articulates [the] social [reality] in a subtle yet pervasive way. For, while the film enjoys a wide and vibrant spectrum throughout [...] this is all utterly reduced by the [...] end to the irreconcilable blacks and whites of Annie's funeral, spectacular in their silent testimony to the real tragedy of Annie" (Flitterman-Lewis 329) .

As is the case with many of Sirk's other films, the spectator can consciously *read* the social statement contained within the film's visual construction. *Changing Lanes* employs a different, but certainly comparable form of subjective stylistic expression in which viewers can detect a very similar implicit social indictment. Most obviously, the cinematography reveals the emotional pressures being imposed on Doyle by mimicking his subjectivity in the jittery handheld camerawork, often intensified by swing pans and fast, jarring cuts. Also, the constant facial close-ups constrict Doyle in the frame to suggest the psychological constriction he's often struggling with, and this sense of claustrophobia is further heightened by the deliberately crowded mise-en-scène in settings like the courtroom lobby. Within this cramped, claustrophobic environment, however, there still remain many physical *divisions* between people, as is seen in the freeway scenes, where people are isolated in their own separate "prisons," their cars, absorbed in their own thoughts and activities. As a whole, the film's mise-en-scène limits human contact even while it often presses people so close together. The modern world Doyles experiences is depicted as an ant farm drained of human intimacy, an emotional wasteland somberly characterized by the film's generally cold, metallic color palette of washed-out blues and grays.

All these techniques demonstrate that the filmmakers clearly want to position the spectators in Doyle's confined and alienated frame of mind within his social surroundings. In other words, the film persuades us to *visually identify* with Doyle, as is signified even more clearly when we receive a literal point-of-view shot from Doyle's visual field while he's being arrested. Michell himself remarks in his commentary at this point, "Imagine what it would be like if it happened to you" (*Changing Lanes*). The film encourages viewers to intensely *feel* Doyle's sense of social imprisonment. His confinement is even more overtly symbolized when he's framed in two successive shots behind rail posts metaphorically resembling prison bars just as he's dashing into his sons' school and headlong into *literal* imprisonment. The shots are reminiscent of a poignant image in *Imitation of Life*, in which Annie collapses weeping and helpless behind an oppressive iron railing.

Indeed, both *Changing Lanes*' and *Imitation of Life's* myriad visual techniques, especially when examined together like this, help secure the subversive implications of the narrative, providing a remarkably strong case *against* both films' facile endings and final conservative appeals for "endurance" and "individual responsibility." And now that we can sharply perceive these films' subversive undercurrents in their style and narrative, we can also recognize how even the happy ending in each film's case can contribute to this subversion and "take on a socially critical edge" in and of itself (Bordwell 7).

According to the predominant academic reading, most of Sirk's films' happy endings employ a form of Euripidean irony, as Sirk himself termed it in an interview (Sirk 132). The narrative rupture of the happy ending can be seen as a conscious, "creative use of discontinuity" meant to dissatisfy the viewers enough to make them question the condescending simplicity of the film's proposed solution ("Sirkian System" 131). As Laura Mulvey writes, "the strength of the melodramatic form lies in the amount of dust the story raises along the road, a cloud of over-determined irreconcilables which put up a resistance to being neatly settled in the last five minutes" (qtd. in Heung 319). Hence, the fake resolution can be *meaningfully* insufficient. Elsaesser notes that the *deus ex machina* creates a "conflict between the intense inscription of neuroses at the stylistic level and any ending which would facilely suggest that the problems of the film had been resolved" (paraphrased in Klinger 17). *Imitation of Life*'s conclusion follows this trend, using Sarah Jane's return to her accepted position not to affirmatively support her social imprisonment, but rather to *underscore* the irrevocability and "unavoidability of racial difference" in American society (Flitterman-Lewis 329). Sirk explains in reference to this film, "[Y]ou're not really supposed to [believe the happy end] [...]. You sense it's hopeless, even though in a very bare and brief little scene afterwards the happy turn is being indicated. Everything seems to be OK, but you well know it isn't" (Sirk 132).

Changing Lanes, too, can be seen to "flaunt" the disparity between what we ask of art and what we know of social life (Bordwell 7). As brief, sudden, and perfunctory as it seems, the film's happy ending, whether intentional or not, remains so plainly, frustratingly inharmonious that it prompts astute viewers (like Gilbey) to see beyond the optimistic message, to see that, despite

what the ending overtly communicates, Doyle truly has "no exit" from passivity and powerlessness within his racist social structure. He has no choice but to remain "acquiescent to the ways of world" (Elsaesser 524).

In light of what this ending can suggest, Gipson's Alchoholics Anonymous slogan takes on a subversive new resonance: "God, grant me the strength to accept what I cannot change!" His words vividly recall Rainer Werner Fassbinder's reaction to *Imitation of Life*: "[N]o one will be able to help [Sarah Jane] [u]nless we change the world [...]. [A]ll of us in the cinema wept because changing the world is so difficult" (qtd. in Taubin 26).

But if the world is indeed so difficult to change, should we still regard *Changing Lanes* strictly as an adamant (and therefore blatantly unrealistic) appeal for social transformation and purification? Now that the film's subversive dimension has been firmly established, perhaps it would be wise to ask if this subversive dimension should remain the "best" or "most legitimate" side of the text, the dimension we should regard exclusively as the text's final "true meaning." Simply put, should the progressive reading remain the last word on the issue?

To answer this question for *Changing Lanes*, we must first take into account why the progressive reading didn't turn out historically to be the last word for Sirk. Of course, Marxist and other academic critics have been imposing these "against-the-grain" readings on Sirk's films for decades, but these attempts to establish his films as "unassailably progressive" have only "remove[d] [them] from the variations in meaning [they were] subject to throughout [their] history" (Klinger 33-34). For example, before Sirk was finally taken seriously as an *auteur*, mass audiences and critics alike actually regarded his films as culturally affirmative pop trash, as the films were often marketed and received as sensationalistic debauches centered around the sexual objectification of their female stars (Klinger 3). After critics in the '60's rehabilitated Sirk's reputation by hailing him as a progressive social critic (based on the narrative and stylistic elements I've described), they soon began to dismiss any conservative or even non-political readings of his films as naïve (Klinger 3). Today, however, academic criticism more often acknowledges the *equal* validity of these texts' original conservative dimensions (Bukatman). A Sirk melodrama is no longer considered to contain one fixed meaning determined by a single (perhaps subversive) individual, but instead is seen to operate to a large degree as a "cultural collaboration," in which meaning is produced by *multiple* artists as well as by the audiences who interpret the texts with certain ingrained, contextually determined cultural assumptions (Saxton 20). The fact remains that "the end product [has] absorbed as many interpretations as there [were] contributors, and the production [has been] fashioned with audience comprehension and taste in mind" (Saxton 22). So a progressive reading of these films should not be considered absolute; conservative readings can often bear an equal credibility in the critical understanding of a given melodrama.

For example, regardless of "whatever ironic subtext Sirk might have intended" with *Imitation of Life*, Marina Heung, a critic who certainly isn't oblivious to Sirk's subversive reputation, counterargues in one article that "the sheer emotional power of its final scene [. . .] finally operates to lay to rest the subversive energy of Sarah Jane and to reinstate Annie, in her death, as the emotional and ideological center of the film" (319-320). Indeed, with *any* film like this, as prominent as the director's possible subversive intentions may be, these intentions often cannot simply *negate* the conservative meanings built into the production, even though the progressive dimensions are not necessarily overshadowed by the conservative messages, either. Because the films often are not dominated by one or the other ideological extreme, these melodramatic texts can't be considered either "fully affirmative or fully subversive" (Rodowick 246).

Considering *Changing Lanes*, too, we cannot deny that "contradictions arise in all texts, including 'radical' ones, [and] that all texts are limited by ideological [...] constraints" (Selig 14). We cannot make the mistake many critics did with Sirk by considering *Changing Lanes* as ultimately "transcending" its conservative dimensions. To an extent, the film *still* operates somewhat like those affirmative social-problem melodramas from the Silent Era. One could even employ Heung's argument on *Imitation of Life* for our examination of *Changing Lanes*: even considering *Changing Lanes*' subversive qualities, the scene in which Gavin blackmails and humiliates his boss is presented with so much crowd-pleasing vigor that it lends significant weight to the genuineness of the happy ending. As oversimplified and false as this last-second resolution may appear to *us*, the filmmakers do not present it completely perfunctorily; the film even actively encourages us to cheer for Gavin, although he is acting as the purveyor of Doyle's affirmative *deus ex machina*. As with *Imitation of Life*, we can detect enough emotional sincerity in *Changing Lanes*' conservative resolution to treat it as a legitimate, meaningful portion of the picture.

So no matter how much we try to emphasize *Changing Lanes*' subversive dimensions, we can still read the film just as validly as a conservative, moralistic criticism of Doyle as we can a progressive criticism of Doyle's society. The question of where the film is ultimately leading us to place the blame for Doyle's problems—on the individual or the culture—remains unresolved either way we read it. We must simply conclude that the film's two ideological currents coexist in a state of inextricable opposition.

In its final illegibility, *Changing Lanes* may remain a something of a disappointment as a film, but that does not mean we cannot learn from observing its confusion. In fact, the true "importance of [this kind of] melodrama lies precisely in its ideological failure. Because it cannot accommodate its problems [...] but lays them open in their shameless contradictoriness, it opens a space which most Hollywood forms have studiously closed off" (Nowell-Smith 273). Though we can easily dismiss the film as an unresolved mess, we can still find it interesting for the richness of its contradictions and for what these contradictions can reveal about ourselves and our society. In much the way *Imitation of Life*'s political ambiguousness can certainly reflect the climate of ideological ambivalence regarding race relations in the late 1950's (Bukatman), *Changing Lanes'* representational contradictions can illuminate our own contradictory perceptions of our present-day American society. We might even say that Lanes reaches a level of political uncertainty so far beyond that of Imitation of Life that it verges into a new category of melodrama. It is in fact more of a post-melodrama: its extreme illegibility reflects the state of a society in which racism and racially defined social hierarchies remain present but become more difficult to identify concretely. The film operates in an environment that has progressed beyond the clearly prejudicial structures of '50s America but has not left its problems behind. The same social problems exist but in less obvious forms. The specter of prejudice lingers, but seems to emanate from everywhere and no where at once.

Returning to Gilbey's point now, we can still consider his observation valid, yet his insight has led us to discover not so much about "white treatment of black culture," but rather more about this powerful, pervasive ambivalence in our nation, ambivalence perhaps concerning, among many other things, what white treatment of black culture, even in a context far beyond the 1950's, can imply about the nature and structure of our society as a whole. Because of the ambiguity regarding what the film ultimately communicates to its audience, *Changing Lanes* successfully taps into our contemporary collective psychology, and it thereby manages to say more about the ideological perspective of our own culture than the filmmakers (and Gilbey) might necessarily realize.

References

Bordwell, David. "Happily Ever After, Part Two." *The Velvet Light Trap*. 19 (1982): 2–7.

Bukatman, Scott. Personal Interview. 26 Nov. 2002.

Butler, Jeremy G. "*Imitation of Life* (1934 and 1959): Style and the Domestic Melodrama." Fischer 298–301.

Changing Lanes. Dir. Roger Michell. DVD. Paramount, 2002.

Elsaesser, Thomas. "Tales of Sound and Fury: Observations on the Family Melodrama." *Film Theory and Criticism*. Oxford: Mast, Cohen & Braudy, 1992. 512–535.

Fischer, Lucy, ed. *Imitation of Life*. New Brunswick, NJ: Rutgers University Press, 1991.

Flitterman-Lewis, Sandy. "*Imitation(s) of Life*: The Black Woman's Double Determination As Troubling 'Other.'" Fischer 325–335.

Gilbey, Ryan. "Street Legal." *Sight & Sound*. 12.11 (2002): 36–37.

Graham, Renee. "*Changing Lanes* merges two lives but breaks down in its search for meaning." \ *Boston Globe*. 12 Apr. 2002. <http://ae.boston.com/movies/display?display=movie&id=1285>.

Heung, Marina. "'What's the Matter with Sarah Jane?': Daughters and Mothers in Douglas Sirk's *Imitation of Life*." Fischer 302–324.

Klinger, Barbara. *Melodrama and Meaning: History, Culture, and the Films of Douglas Sirk*. Bloomington: Indiana University Press, 1994.

Landau, Saul. "*Changing Lanes*." *Cineaste*. 27.4 (2002): 40–41.

Landy, Marcia, ed. *Imitations of Life: A Reader on Film & Television Melodrama*. Detroit: Wayne State University Press, 1991.

McKegney, Michael. "Film Favorites: *Imitation of Life*." *Film Comment*. 8.2 (1972): 71–73.

Nowell-Smith, Geoffrey. "Minnelli and Melodrama." *Landy* 268–274. Reed, Rex. "A Game of Hit-and-Run." *New York Observer*. 15 Apr. 2002. <http://www.nyobserver.com/pages/story.asp?ID=5727>.

Rodowick, David N. "Madness, Authority and Ideology in the Domestic Melodrama of the 1950's." *Landy* 237–247.

Saxton, Christine. "The Collective Voice as Cultural Voice." *Cinema Journal*. 26.1 (1986): 19–30.

Selig, Michael E. "Contradiction and Reading: Social Class and Sex Class in *Imitation of Life*." *Wide Angle.* 10.4 (1988): 13–23.

Sirk, Douglas. Interview. *Sirk on Sirk: Interviews with Jon Halliday*. New York: Viking Press, 1972.

Sloan, Kay. "A Cinema in Search of Itself: Ideology of the Social Problem Film During the Silent Era." *Cineaste*. 14.2 (1985): 34–37, 56.

Taubin, Amy. "In every dream home." *Film Comment.* 38.5 (2002): 22–26.

"Towards an Analysis of the Sirkian System" (Sirkian System). *Screen*. 13.4 (1972-73): 128–134.

Willemen, Paul. "Distanciation and Douglas Sirk." Fischer 268–272.

Andre Lyon

Andre Lyon is currently a sophomore at Stanford double-majoring in computer science and art history (with a film-studies emphasis and a studio-art minor). He was born in Oakland, CA and attended high school in the New Orleans area. Two of his essays have appeared in the *Boothe Prize Essays 2003* publication. He has been producing independent short features since high school and is currently exhibiting an animated short at several national film festivals.

Endocrinology

Hormonal and Physiological Profiles of Female *Haplochromis burtoni* as it relates to Affiliative Behavior

The African cichlid fish, Haplochromis burtoni*, has proven to be a useful model for understanding the social control of behavior. In females of this species, there is substantial remodeling of the brain-pituitary-gonadal axis controlling reproduction between gravid (egg bearing) and non-gravid states (White & Fernald, 1993). Moreover, female preference for male phenotype changes during the course of the reproductive cycle (Clement, Grens & Fernald, submitted). To understand what hormonal changes are associated with the behavioral and structural changes, we measured levels of several key hormones and peptides throughout the reproductive cycle of the female as a function of ovarian stage and development. We quantified plasma levels of testosterone and 17ß-estradiol using enzyme-linked immunoassays. The expression of gonadotropin-releasing hormone (GnRH), a hormone produced in the hypothalamus and involved in sexual maturation, was determined using quantitative real time RT PCR. In addition, the expression of two GnRH receptor types in the pituitary was also established using quantitative real time RT PCR. We relate the changes in hormonal profile to female affiliation preference to identify internal cues that might help orchestrate the shift to preference for dominant males.*

Jo Martin

Sexual selection, or mate choice, refers to any behavior in which one sex discriminates among several members of the opposite sex based on some salient criterion, such as size, status, or color. Researchers have documented sexual selection throughout the animal kingdom. Species of insects (Bonduriansky, 2001), spiders (Hebets, 2003), fish (Hughes, Du, Rodd, & Reznick, 1999; Kelley, Graves, Mugarran, 1999; Kodric-Brown, 1993; Clement, Grens, & Fernald, submitted), birds (Karubian, 2002), and mammals (Jurke, Price, & Debeli, 1995) have all been shown to engage in selective behavior. Choice behavior has been described in both females and males, though more frequently in females. Although the cues that may lead a female to distinguish between two possible mates are well described, rarely have the physiological factors determining such choice been examined. The objective of this research is to determine the underlying mechanisms responsible for the variations in female preference across the reproductive cycle.

Mate choice was first described in 1871 by Charles Darwin, who observed that many animals possess traits that appeared useless for survival, such as large plumage or bright colors that attract predators. Darwin postulated that these traits were prized by the opposite sex and thus enhanced the individual's probability of reproducing, which explained why they continued to be passed down from generation to generation. Darwin was also the first to refer specifically to female mate choice: "No doubt this implies powers of discrimination and taste on the part of the female which will at first appear extremely improbable; but I hope to shew (sic) that the females actually have these powers" (Darwin, 1871). Although Darwin's theory of natural selection was widely accepted, it was not until the late twentieth century that Darwin's theories of sexual selection began to gain prominence and recognition. Since then, the topic has been widely investigated in order to determine why sexual selection occurs and what cues animals use to decide among potential mates.

Numerous researchers have used fish to investigate mate choice in simple vertebrates. One common finding is that fish use visual features to distinguish between prospective mates. For example, female guppies prefer brightly colored males to males that are less colorful (Godin & Dugatkin, 1996), while female swordtail fish prefer males who have symmetrical vertical bars to those with asymmetrical ones (Morris & Casey, 1997). It has been suggested that these choice decisions are based on the genetic quality of the mate. In guppies, the male's ability to escape from predators, as well as his boldness towards predators, seem to be indicators of the male's genetic quality. Godin and Dugatkin (1996) established that the brightness of a male correlates positively with boldness towards and escape from predators, demonstrating that bright color may be a reasonable indicator of the quality of the male. Likewise, female swordtail fish may

prefer symmetrical males to asymmetrical males because genetic quality seems to be linked to the symmetry of sexual traits (Morris and Casey, 1997).

Recent research in sexual selection behavior has investigated whether female choice may vary according to the physical state of the female. Clement, Grens, and Fernald (submitted) demonstrated that females' willingness to affiliate with a dominant male changes across the reproductive cycle. They showed that in the African cichlid fish, *Haplochromis burtoni,* the reproductive state of the female determines whether she will show preference for territorial males. Their experiment tested female preference for the two types of *H. burtoni* males – territorial and non-territorial – who differ in size, coloration, status, and reproductive viability. Territorial males are large, aggressive, brightly colored, and control areas of territory and resources. They also have mature gonads capable of reproduction. In contrast, non-territorial males are smaller, less aggressive, camouflage colored, and do not control areas of territory. They only gain access to food when they are mistaken for females, and they are reproductively suppressed with immature gonads. As the female begins her reproductive cycle, she schools with females and non-territorial males. Only after she has become quite gravid (egg bearing and fertile) does she choose to affiliate with territorial males in order to spawn.

Clement, Grens, and Fernald investigated this behavior and found that when given a choice between a territorial male and a non-territorial male, gravid females prefer territorial males, while non-gravid females do not show a preference. This finding suggests that mate choice behavior may be influenced by the hormonal fluctuations during a female's reproductive cycle. However, in order to determine how hormones may affect mate choice either directly or indirectly, we first need to determine the time course of reproductive cycle and corresponding physiological variations. It has been documented that there are physiological changes between gravid and non-gravid females, and that these changes are likely signaled at a hormonal level. We investigated several hormones, hormone receptors, and physiological factors to determine which of these might mediate reproductive behaviors like mate choice.

White, Nguyen, and Fernald (2002) established that gonadotropin-releasing hormone I (GnRH I), a hormone released by the hypothalamus and involved in sexual maturation, varies across reproductive states in *H. burtoni*. We extended this research to look at the expression of GnRH I at several other time points over the entire reproductive cycle. We also examined GnRH I and GnRH III receptor expression in the pituitary. In addition, there are several key hormones known to regulate GnRH, including 17ß-estradiol and testosterone. These two hormones have also been shown to affect both behavior and physiology as they fluctuate during the reproductive cycle. As a result, we determined concentrations of 17ß-estradiol and testosterone in *H. burtoni* plasma. Lastly, we measured the gonadosomatic index (ovary weight/body weight x 100) in order to correlate these changes with the female's reproductive state.

Methods

Housing

We used stock populations of *Haplochromis (Astatilapia) burtoni* that were originally obtained from wild populations that live in Lake Tanganyika, Africa. Males and females were maintained together in aquaria that resemble the conditions of their natural environment: 28° C water temperature, pH 8 and a 12:12-hr light:dark cycle with full spectrum illumination (Fernald & Hirata, 1977).

To ensure that all the females began the reproductive cycle at the same time, we collected females after they had released mature fry (viable offspring). We checked every afternoon for females that were brooding eggs or immature fry and collected the females only when the fry were viable (*H. burtoni* females brood their offspring in their mouths for two weeks prior to release). We collected five to ten females at a time and housed them together in an aquarium with one to two territorial males per five females. The females were kept in these tanks until the day of sacrifice.

Sampling Procedure

Based on our observation that the female reproductive cycle is about 30 days long, we chose data collection points between the first and thirtieth days following the release of the fry. We collected data on days 1, 3, 6, 9, 12, 15, 18, 21, 23, 25, and 29 (Day 1 corresponding to the day that the fry are taken) from August to November of 2003. Sample size ranged from 3-5 subjects per data point. Prior to sacrifice and blood drawing, the fish were weighed and their length was measured. We drew blood from the caudal vein; the blood was centrifuged immediately and the plasma was drawn off and stored at –80˚C. We extracted the brain, pituitary, and ovaries from all of our samples and immediately froze them at –80˚C until analysis. Ovaries were weighed in order to calculate the gonadosomatic index (GSI = 100 x gonad weight/body weight) prior to freezing. Any eggs or immature fry were also immediately frozen and stored at –80˚C.

Determination of Reproductive Hormones and Peptides

We extracted plasma steroids from our plasma samples according to Cayman Chemical's Enzyme-Linked Immunoassay protocol. These extracts were analyzed for 17ß-estradiol and testosterone using enzyme-linked immunoassays (Cayman Chemical, Ann Arbor, Michigan).

We measured levels of GnRH I in the brain and GnRH receptor I and III in the pituitary using real time quantitative Reverse Transcription Polymerase Chain Reaction (RT-Q-RT-PCR), a sensitive assay used to determine levels of gene expression. Levels of GnRH I and GnRH receptor I and III were quantified using a protocol described in White, Nguyen, and Fernald (2002). All of these analyses were carried out with both positive and negative controls. After real time quantitative RT-PCR, we ran out the samples on a 2% agarose-ethidium bromide gel in order to determine the purity of the sample. Results were normalized with G3PDH and are given as percent expressions.

$$\textit{Percent expression} = \frac{\text{rel. exp. of G3PDH}}{\text{rel. exp. of gene of interest}}$$

$$\textit{Relative expression} = \text{Efficiency}^{\text{Average Cycle Threshold}}$$

Results

Gross Anatomy and GSI

As a female becomes gravid (egg bearing), her abdomen increases noticeably in size. The increase in abdominal size is due to the tremendous growth of the ovaries as the female approaches spawning (when she deposits her eggs at the anal fin of the male where they will be fertilized). At the beginning of the cycle, when the female has just released fry, the gonadosomatic index (GSI) is at its lowest point. By twelve days after fry release, most females are gravid again, and the average GSI shows a corresponding increase. Once the females start mouthbrooding their fry around day twenty-five, the GSI drops down to the initial level again (see Figure 1).

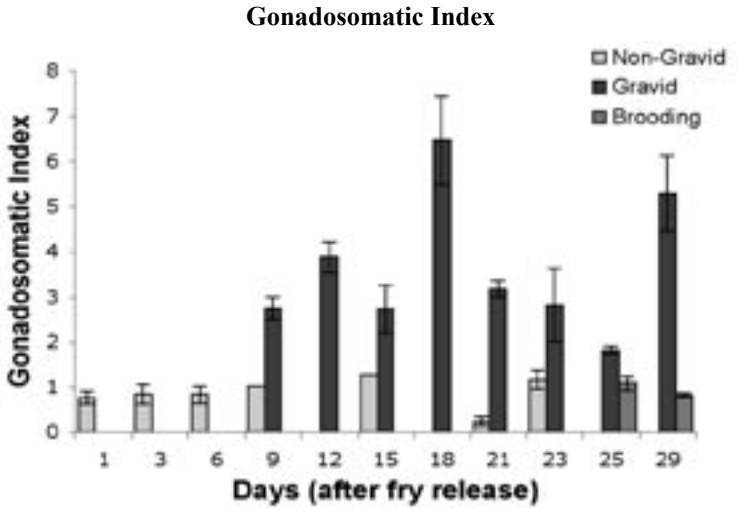


Figure 1. Gonadosomatic Index (GSI) across the female *Haplochromis burtoni* reproductive cycle. GSI = 100 x ovary weight/body weight.

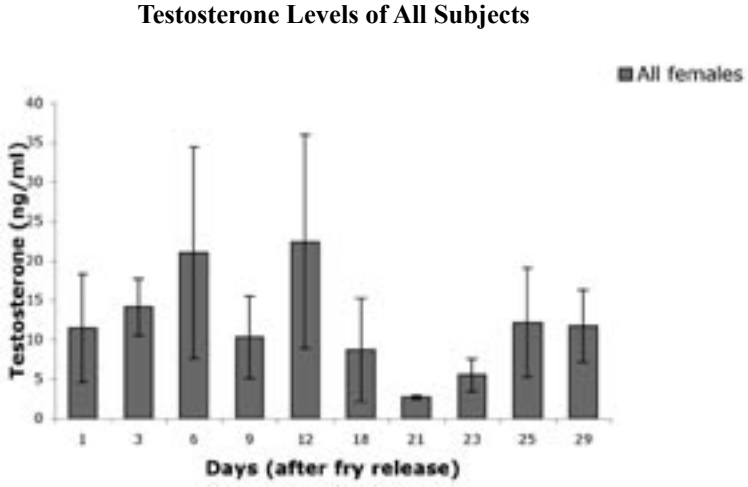


Figure 2. Plasma concentrations of testosterone in all subjects across the female *H. burtoni* reproductive cycle.

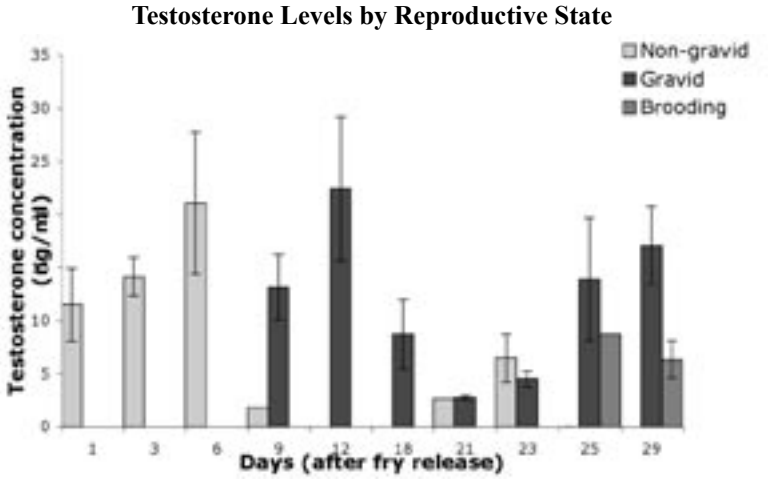


Figure 3. Plasma concentrations of testosterone across the female reproductive cycle, shown by reproductive state (Gravid, Non-Gravid, Brooding).

Hormone Profiles

Circulating plasma levels of testosterone and 17ß-estradiol were measured across the reproductive cycle. When females release their fry, testosterone and 17ß-estradiol concentrations are low. As females near the end of the non-gravid phase, testosterone concentrations increase first, reaching a peak around the sixth day after release (see Figures 2 and 3). They continue to rise as the female becomes gravid. Testosterone levels begin to decrease as the ovaries mature, and the females approach spawning. Testosterone concentrations are lowest during spawning, although they increase slightly during brooding (see Figure 3).

17ß-estradiol levels increase as the female becomes gravid and the ovaries start to mature. The levels peak just prior to ovary maturation, and begin to decrease as the female approaches spawning (see Figure 4). 17ß-estradiol remains low through both spawning and brooding (see Figure 5).

Real Time Quantitative RT-PCR results

Gonadotropin–releasing hormone I (GnRH I) is directly involved in the reproductive cycle of *H. burtoni* and is produced by neurons in the preoptic area of the hypothalamus. We performed real time quantitative RT-PCR in order to measure the amount of GnRH I mRNA transcript produced in the brain across the reproductive cycle. GnRH I transcript shows a peak at the eighteenth day after release, when the females are most gravid. It decreases slightly as the females spawn, and increases again during brooding (see Figure 6).

GnRH receptors I and III show similar patterns of expression across the reproductive cycle (see Figures 7 and 8). Both gradually increase as the female becomes gravid, peaking around the time that the female is most gravid, and then decreasing as she approaches

Estradiol Levels of All Subjects

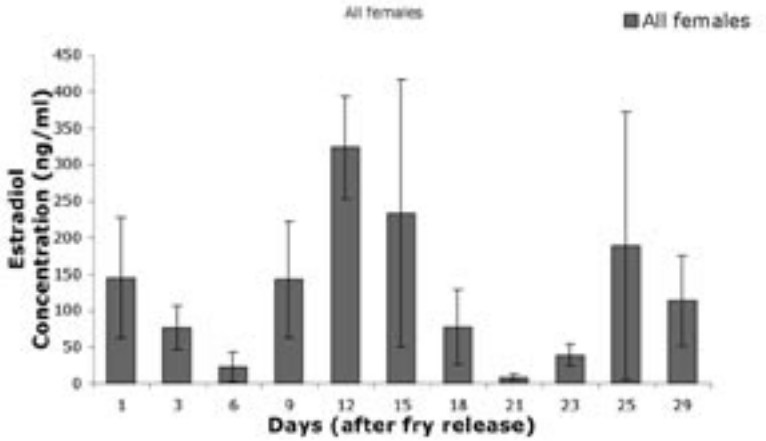


Figure 4. Plasma concentrations of 17ß-estradiol in all subjects across the female *H. burtoni* reproductive cycle.

Estradiol levels by Reproductive State

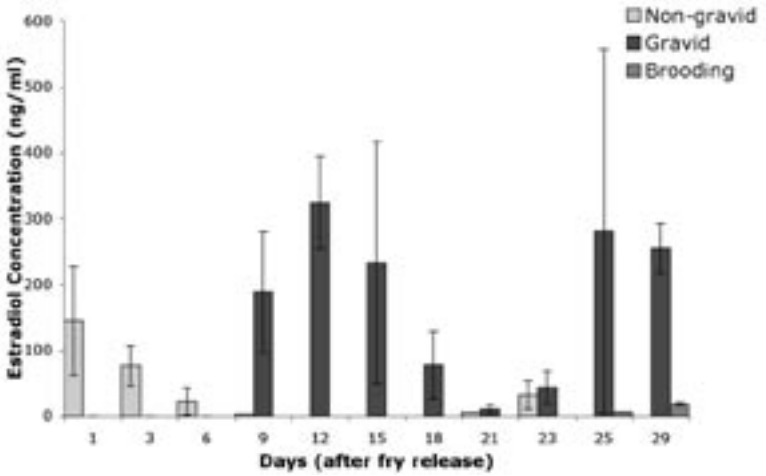


Figure 5. Plasma concentrations of 17ß-estradiol across female reproductive cycle, shown by reproductive state (Gravid, Non-Gravid, Brooding).

GnRH I Expression in the Brain by Reproductive State

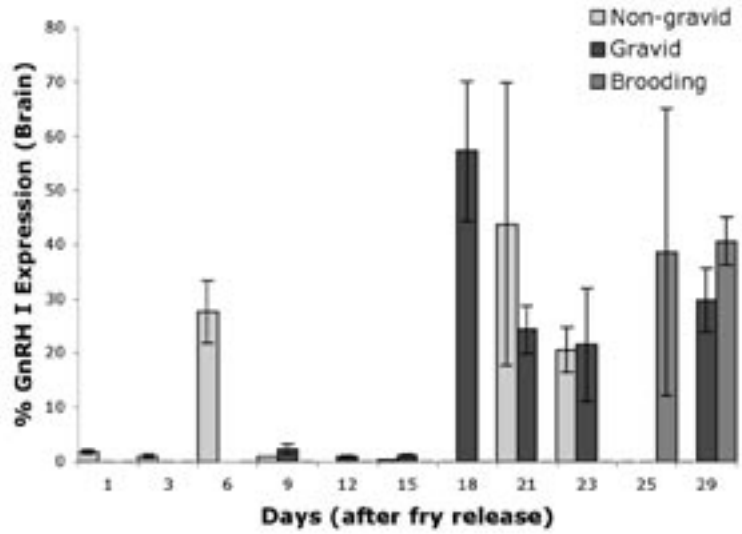


Figure 6. Gonadotropin-releasing hormone I (GnRH I) mRNA expression in the brain of female *H. burtoni*. Shown by reproductive state (Gravid, Non-Gravid, Brooding) across the reproductive cycle.

the end of the cycle. However, GnRH receptor I remains low during brooding while GnRH receptor III expression is high.

Discussion

The primary goal of this study was to identify the hormones involved in the reproductive cycle of female *H. burtoni* in order to correlate these fluctuations with changing behavior. To this end, we examined levels of 17ß-estradiol and testosterone in the plasma, GnRH I transcript in the brain, GnRH receptor I and III transcript in the pituitary, and gonadosomatic index. Our results indicate that when females release their fry, circulating levels of 17ß-estradiol and testosterone are low. The ovaries are regressed, and little GnRH transcript is produced in the brain. As females become gravid, testosterone concentrations increase, followed several days later by 17ß-estradiol. As 17ß-estradiol levels rise, the ovaries mature and increase in size. At the same time, the pituitary upregulates production of GnRH receptor I and III. When the female is very gravid, it appears that GnRH I transcript in the brain increases dramatically. However, additional data points are needed to determine when this hormone is most abundant.

This research extends previous research that indicated that gonadotropin-releasing hormone (GnRH) levels are correlated with the size of GnRH neurons in the preoptic area of the hypothalamus. White and Fernald (1993) demonstrated that GnRH neurons display remarkable plasticity throughout the reproductive cycle, as they are twice as large in females that have never spawned or are in the act of spawning than they are in females who are brooding. They also determined that preoptic GnRH transcript is significantly higher in spawning females than in brooding females (White, Nguyen, & Fernald, 2002), a result replicated here.

Although these physiological factors clearly fluctuate throughout the reproductive cycle, how might they direct reproductive behaviors such as mate choice? One possibility is that gonadal hormones may affect sexual motivation, as they do in primates (Wallen, 2001). Although high levels of gonadal hormones are necessary for reproductive behavior in most mammals, they are not necessary for sexual behavior in humans and primates (Wallen, 2001). For instance, many mammalian males require high levels of gonadal hormones (like testosterone) in order to maintain an erection. In contrast, both rhesus monkeys and human males can still engage in sex if testosterone is eliminated from their bodies, although they will lose sexual motivation. Likewise, levels of ovarian hormones may influence human female sexual motivation, but reproductive behavior is possible during all stages of the female menstrual cycle, and reproductive behavior is often influenced by social and cognitive factors (Wallen, 2001). It is possible that *H. burtoni* females are governed by a similar mechanism. Without the appropriate concentrations of ovarian hormones, female fish simply may not have the motivation to associate prefer-

GnRH receptor III Expression in the Pituitary by Reproductive State

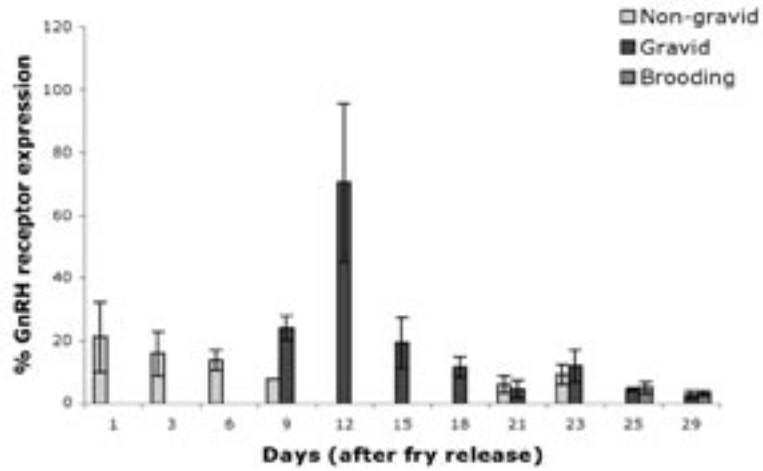


Figure 7. Gonadotropin-releasing hormone receptor I (GnRH receptor I) mRNA expression in the pituitary in female *H. burtoni*. Shown by reproductive state (Gravid, Non-Gravid, Brooding) across the reproductive cycle.

GnRH receptor III Expression in the Pituitary by Reproductive State

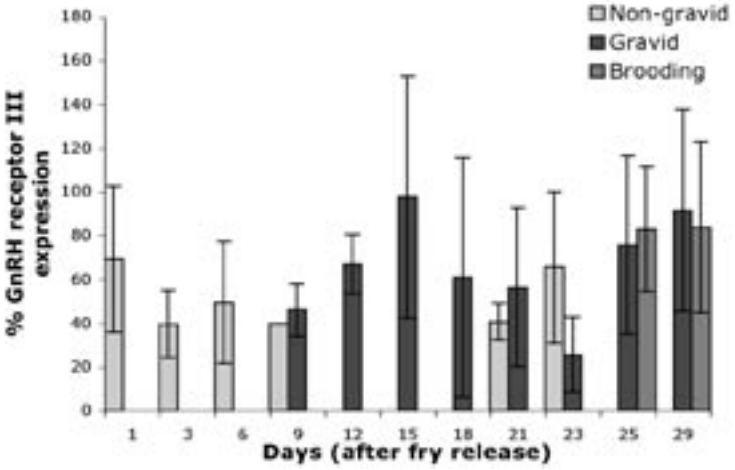


Figure 8. Gonadotropin-releasing hormone receptor III (GnRH receptor III) mRNA expression in the pituitary. Shown by reproductive state (Gravid, Non-Gravid, Brooding) across the reproductive cycle.

entially with territorial males. In particular, estradiol may play a large role in initiating reproductive maturation.

Although we hypothesize that a sex hormone (or perhaps several) mediates the shift in preference to territorial males, it is also possible that hormones may act by controlling behavior through changes in perception, and particularly through vision. The mate choice experimental design prevents any tactile contact between the female and the male, and it also limits the olfactory cues (the fish are separated by a clear plexiglass barrier). Hirata and Fernald (1975) have also determined that *H. burtoni* do not use sound for communication purposes. However, Hoke and Fernald (2002) established that estrogen receptor alpha is found in the retina of *H. burtoni*, so it is possible that either fluctuations in estrogen or estrogen receptors could subtly affect visual acuity or sensitivity. Intriguingly, estrogen, androgen, and progesterone receptors have also been found in several other species, including rats, rabbits, and humans (Wickham, Gao, Toda, Rocha, Ono, & Sullivan, 2000). If results indicate that visual perception changes across the reproductive cycle in *H. burtoni* (leading to changes in mate choice behavior), it is possible that the changing preference for male faces throughout the menstrual cycle in human females (Frost, 1994) may be mediated by a similar mechanism.

Human females also show variations in perception in several other sensory modalities: olfaction (Watanbe, Umezu, & Kurahashi, 2002), hearing (Davis & Ahroon, 1982), nocireception or pain sensitivity (Gaumon, Arsenault, & Marchand, 2002; Kerem, et al., 2002), vision sensitivity (Parlee, 1982), and even taste (Aaron, 1975) have been shown to change throughout the menstrual cycle. It is still unclear how hormones in the female menstrual cycle may influence these changes in perception, and it is likely that this connection will be understood in animal models (such as *H. burtoni*) before it is fully understood in humans.

Even if the *H. burtoni* mate preference shifts are not due to changes in perception, this research does have interesting implications for humans. Somewhat surprisingly, many of the same hormones that are prevalent in human females are also prevalent in *H. burtoni* females. In humans and fish, estradiol and testosterone are important gonadal hormones that govern reproduction and reproductive behaviors. Just as in fish, humans release gonadotropin-releasing hormone from the hypothalamus, which stimulates the release of gonadotropins from the pituitary. Due to this conservation of form and function across species, we can study these hormones and their effects in simpler systems and generate intelligent guesses about how these hormones may function in humans.

References

Aaron, M. (1975). Effect of the menstrual cycle of subjective ratings of sweetness. *Perceptual and Motor Skills*, 40, 974.

Bonduriansky, R. (2001). The evolution of male mate choice in insects: a synthesis of ideas and evidence. *Biological Review*, 76, 305-339.

Clement, T.S., Grens, K.E., & Fernald, R.D. (submitted) Female association preference depends on reproductive state in the African cichild fish, *Haplochromis burtoni*.

Darwin, C. (1871). *The Decent of Man and Selection in Relation to Sex.* New York: D. Appleton and Co.

Davis, M.J. & Ahroon, W.A. (1982). Fluctuations in susceptibility to noise-induced temporary threshold shift as influenced by the menstrual cycle. *The Journal of Auditory Research*, 22, 173-187.

Fernald, R. D. & Hirata, N. R. (1977). Field study of *Haplochromis burtoni*: habitats and co-habitat. *Environmental Biology of Fishes* 2(3), 299-308.

Frost, P. (1994). Preference for darker faces in photographs at different phases of the menstrual cycle: preliminary assessment of evidence for a hormonal relationship. *Perceptual Motor Skills*, 79, 507-514.

Gaumond, I., Arsenault, P., and Marchand, S. (2002). The role of sex hormones on formalin-induced nocireceptive responses. *Brain Research*, 958, 139-145.

Godin, J. J. and Dugatkin, L. A. (1996). Female mating preference for bold males in the guppy, *Poecilia reticulata. Procedings of the National Academy of Science*, 93, 10262-10267.

Hebets, E.A. (2003) Subadult experience influences adult mate choice in an arthropod: Exposed female wolf spiders prefer males of a familiar phenotype. *Proceedings of the National Academy of Sciences*, 100, 13390-13395.

Hirata, N. R. & Fernald, R. D. (1975). Non-intentional sound production in a cichlid fish (*Haplochromis burtoni, Gunther*). *Separatum Experientia*, 31, 299-300.

Hoke, K. L. & Fernald, R. D. (2002). Teleost fish retina contains estrogen signaling pathway. *Society for Neuroscience Abstracts*.

Hughes, K.A., Du, L., Rodd, F. H., and Reznick, D. N. (1999) Familiarity leads to female mate preference for novel males in the guppy, *Poecilia reticulata*, *Animal Behavior*, 58, 907-916.

Jurke, M.H., Pryce, C.R., & Dobeli, M. (1995). An investigation into sexual motivation and behavior in female Goeldi's monkey (*Callimico goeldii*): effect of ovarian state, male familiarity, and mate choice. *Hormones and Behavior*, 29(4), 531-553.

Karubian, J. (2002). Costs and benefits of variable breeding plumage in the red-backed fairy-wren. *Evolution; International Journal of Organic evolution*, 56(8), 1673-1682.
Kelley, J.L., Graves, J.A., and Mugarran, A.E. (1999). Familiarity breeds contempt in guppies, *Nature*, 401, 661.
Kerem, M., Akbayrak, T., Bumin, G., Yigiter, K., Armutlu, K., & Kerimoglu, D. (2002). A correlation between sex hormone levels and pressure pain threshold and tolerance in healthy women. *The Pain Clinic*, 14(1), 43-47.
Kodric-Brown, A. (1993). Female choice of multiple male critieria in guppies: interacting effects of dominance, coloration and courtship. *Behavioral Ecology and Sociobiology*, 32, 415-420.Kurubian
Morris, M. and Casey, K. (1998). Female swordtail fish prefer symmetrical sexual signal. *Animal Behavior*, 55, 33-39.
Parlee, M.B. (1983). Menstrual rhythms in sensory processes: a review of fluctuations in vision, olfaction, audition, taste, and touch. *Psychological Bulletin*, 93(3), 539-548.
Wallen, K. (2001). Sex and context: hormones and primate sexual motivation. *Hormones and Behavior*, 40, 339-357.
Watanabe, K., Umezu, K., and Kurahashi, T. (2002) Human olfactory contrast changes during the menstrual cycle. *Japanese Journal of Physiology*, 52, 353-359.
White, S.A. & Fernald, R.D. (1993). Gonadotropin-releasing hormone-containing neurons change size with reproductive state in female *Haplochromis burtoni*. *The Journal of Neuroscience*, 13(2), 434-441.
White, S.A., Nguyen, T., & Fernald, R.D. (2002). *Social regulation of gonadotropin-releasing hormone. The Journal of Experiemtnal Biology*, 205, 2567-2581.
Wickham, L. A., Gao, J., Toda, I., Rocha, E. M., Ono, M., & Sullivan, D. A. (2000). Identification of androgen, estrogen, and progesterone receptor mRNAs in the eye. *Acta Ophthalmologica Scandinavica*, 78, 146-153.

Jo Martin

Jo Martin is a senior from Lake Oswego, Oregon majoring in Psychology with honors. Her primary interests are neuroscience and neuroendocrinology, and following graduation she plans to attend medical school. She has been working in the lab of Professor Russ Fernald under the guidance of Tricia Clement since her junior year. In the winter of 2004, she presented her research at the Society for Integrative and Comparative Biology conference in New Orleans, LA.

Mentors: Tricia Clement and Russ Fernald

Urban History

Saving America's "Last Lovely City"

The San Francisco Freeway Revolt

One of the most striking aspects of San Francisco is its relative lack of freeways, unusual for a major American city. In the 1950s, however, local and state transportation authorities presented ambitious plans to build a total of nine freeways within the city's 49-square mile area. Residents of San Francisco angrily rose up to protest the freeways and -- through two influential Board of Supervisors decisions in 1959 and 1966 -- succeeded in blocking most of the proposed routes. In an era when many other American cities were enthusiastically constructing new infrastructure for the automobile, San Francisco's "freeway revolt" was quite unprecedented. In this paper, I trace the development of this "revolt" and argue that the economic and transportation benefits of freeways became subordinated to aesthetic considerations. While concerns over property values and tourism were important factors, the perceived damage that freeways would inflict on San Francisco's unique cityscape proved to be the greatest rallying point in the revolt. San Franciscans were unwilling to let "concrete monsters" ruin vistas, impact quality of life, and ultimately spoil what they termed as America's "last lovely city."

Dinyar Patel

"There will be a freeway on the moon before we get one in San Francisco," Mayor John Shelley declared in 1966 (Issel, 1999, 612). Shelley's statement was no act of hyperbole. In 1966, during an era when American cities were enthusiastically planning and erecting networks of highways, the San Francisco Board of Supervisors virtually arrested all further freeway construction within the city. Widely cheered, the Board's decision was the culmination of a decade-long "freeway revolt" aimed at preventing the implementation of transportation network plans first suggested in the 1940s. Despite an atrophying 19th century infrastructure and a worsening traffic situation within the city, business associations, chambers of commerce, and major newspapers joined in the celebrations over the defeat of the "concrete monsters."

Living in an environment gifted with a dramatic natural setting of hills and water, San Franciscans – members of what Herb Caen called "the best Club anywhere" – had long been wary of major building projects, fearing that incongruous development might impact quality of life, alter the city's dynamics, or – worst of all – spoil "the view" (Wirt, 182). These considerations became the pivotal factors in the San Francisco freeway debate. Overlooking serious economic and transportations considerations, San Franciscans instead stressed the *aesthetic* damage freeways would cause in the city. By the 1960s, aesthetic considerations had become so important that California highway authorities were forced to present highly elaborate, costly freeway alternatives with the sole intention of minimizing visual impact. These, nevertheless, were rejected by the Board in 1966. The "freeway revolt" of the 1950s and 60s set a lasting precedent for the city, ensuring that most of San Francisco was untouched by major infrastructure projects and even propelling a movement to demolish *existing* freeways. This, however, has come at a serious cost: traffic has worsened considerably over the last few decades, and strong opposition to intrusive freeways has virtually eliminated the city and state's ability to improve San Francisco's aging automobile transportation network.

Designing the Network: The Trafficways Plan (1940s-1950s)

Having sensed the coming age of the automobile with the completion of the Golden Gate and Bay bridges, San Francisco city officials began to debate methods for modernizing city infrastructure during the early 1940s. In 1945, the newly formed City Planning Commission adopted its first Master Plan, containing the "Major Thoroughfares Plan" transportation component. More an outline of general purpose than a detailed report, the "Major Thoroughfares Plan" called for

a system of freeways, parkways, and rapid transit lines designed to facilitate local transportation and provide easy access for an increasing number of suburban commuters. It sketched out approximate routes for four proposed freeways: the Bayshore Freeway,[1] an Embarcadero Freeway following the bayfront from Army Street to the intersection of Van Ness Avenue and Lombard Street, a Panhandle Freeway linking Golden Gate Park with the Civic Center, and a route parallel to the Bayshore Freeway that would follow Mission Street ("Four Freeways Added," 1951, 16).

In the following years, the Board of Supervisors authorized a comprehensive survey to study freeway proposals and draft final plans. A new transportation system had become a widely discussed and highly important issue in San Francisco, as the Planning Commission's 1948 *Report on Activities* stated:

> Of all the problems confronting San Francisco in its current development as a changing city, transportation . . . is of most immediate importance. More people talk about it, more columns of news and editorial space are devoted to it, and more city departments are directly concerned with it, than perhaps any other single problem confronting San Francisco. The very survival of San Francisco as a healthfully functioning city is dependent upon resolving the not-so-slow strangulation which an atrophying transportation system is inflicting upon the city (3).

Nevertheless, in these early years of enthusiasm for expressways and highways, some in San Francisco stressed caution. L. Demming Tilton, the city's first planning director, warned that a freeway "is a device that can make or break the city . . . It can cut the city into unrelated parts, or bind it together" (Issel, 621). UC Berkeley city planning professor Theodore J. Kent suggested that an expanded public transit system would be more appropriate for San Francisco and surrounding regions (Rose, 1990, 58).

Draft plans for the San Francisco freeway network published in 1947. Several of the freeways shown here, such as those along the oceanfront, San Mateo County line, and fringes of the Presidio, were eliminated before the 1951 'Trafficways Plan' was released.

Despite these opinions, on July 17, 1951 the City Planning Commission endorsed construction of an extensive network of new roadways by adopting the landmark "Trafficways Plan" as an element of the Master Plan. "Trafficways" authorized construction of three types of roadways: freeways, primary thoroughfares (major surface arterials that would be turned into expressways),[2] and secondary thoroughfares (city streets that would be widened and upgraded). In addition to the freeways recommended in the 1945 "Major Thoroughfares Plan," "Trafficways" added more routes: by the time it was updated in 1955, the plan included freeways bisecting the Pacific Heights, Mission, Sunset, Richmond, and Haight-Ashbury neighborhoods. Also included was a new transbay bridge to Oakland – the Southern Crossing – which would help "link all portions of this great metropolitan area so as to stimulate and soundly influence efficient and convenient transportation and traffic flow" (*The Case for the Southern Crossing*, 15; Nolte, "Southern Crossing,"A-1).[3] San Francisco now had a total of nine proposed freeways packed into its 49-square mile area.

The massive scale of freeway construction proposed did not immediately create protest; in fact, "Trafficways" enjoyed tremendous support in the early 1950s. Before it was approved by the City Planning Commission in 1951, "Trafficways" had been endorsed by the Downtown Association, the Chamber of Commerce, the Junior Chamber of Commerce, and many other organizations. Business groups hoped that freeways would inject new energy into a downtown that was beginning to suffer due to decentralization and suburban growth. Indeed, the main objectives of "Trafficways" seemed to cater to business interests: "Development of an efficient, economical, and balanced system of major trafficways . . . employed where it is most suitable and effective from the standpoint of present and prospective traffic movement and from the standpoint of the present and desirable future use of adjoining land areas" (*The Master Plan*, 1963, no page).

Aside from some protest in the middle-class Cow Hollow neighborhood, most city residents supported freeway construction and did not express concern over the highways' impact on their neighborhoods (*Annual Report: 1949-1950*, 5). This was largely due to the long-range timeframe adopted by "Trafficways": the plan was not supposed to be fully complete for another twenty years ("Four Freeways Added," 1951, 16), and thus "construction of its

elements seemed remote" (Lathrop, 1969, 2).

Beginnings of Dissent: The Original "Freeway Revolt" (1950s)

In his book, *Power in the City*, Frederick M. Wirt comments that in "perhaps no other city in America do citizens so seriously discuss" aesthetic implications of new development. Wirt claims that this "sensitivity to environmental beauty" stems from an overwhelming pride San Franciscans have in their city, and creates "an aesthetic of place" (182). Maintenance of San Francisco's beauty has evolved into a key quality of life factor. Freeways, therefore, were bound to become a significant topic of discussion not for their effect on local traffic patterns but because of their aesthetic impact.

The California State Division of Highways (DOH) was quick to realize the importance of aesthetic considerations in San Francisco. DOH officials claimed that the Bayshore Freeway – the first element in "Trafficways" to be constructed – created a "system of Skyways" where "the beauty which has long been San Francisco's fame . . . unfold[s] to the public entering from all directions" (Issel, 1999, 626). In its 1955-56 *Capital Improvement Plan*, the San Francisco Department of City Planning noted that "a national art award was recently granted to the Department of Highways for the design of its . . . junction of the Bayshore, Central, and Embarcadero Freeways . . . The state engineers have opened up new vistas that can be seen by motorists from these new viaducts" (4).

In contrast, San Francisco residents were horrified by the completed Bayshore Freeway. This artery did do a remarkable job in improving crosstown transportation and relieving surface streets – traffic on busy Potrero Avenue, for example, fell from 30,698 vehicles per day to only 9,068 vehicles per day after the freeway was opened (*Capital Improvement Plan: 1956-1957*, 6). Nevertheless, citizens condemned the creation of more "expensive, *unsightly*, and very durable civic blunder[s] for future generations to mourn over."[4] The *San Francisco Chronicle* slowly backed away from the endorsement of freeway construction it gave in the early 1950s (Issel, 1999, 627). Upon seeing how the Bayshore Freeway divided communities along its route and lumbered over blocks in South of Market, many San Franciscans restudied freeway proposals to see how future construction would affect their neighborhoods and other sections of the city.

The bayfront Embarcadero Freeway, in particular, propelled a nascent "freeway revolt." When construction began on the freeway in 1957, much of the central waterfront was still industrial and therefore thought to be of little scenic value. However, at the foot of Market Street stood the city's historic gateway – the Ferry Building – and any elevated structure threatened to block the sightlines on Market. San Franciscans became increasingly concerned when the entire Embarcadero route was changed from a single-level elevated structure to a double-decked structure, effectively forming a three-story wall along the waterfront. Tremendous public pressure to preserve the waterfront failed in the late 1950s: the DOH categorically rejected an $11 million proposal to tunnel the freeway in front of the Ferry Building. Angered citizens, in "one of the wildest San Francisco Board meetings on record," convinced city supervisors to pass a resolution stopping freeway building at Market, but the DOH continued construction. Once completed, the structure was reviled as "the concrete monster" (Leavitt, 1970, 201).

Property values were understandably a large reason for the "freeway revolt," but like the Embarcadero Freeway some routes were opposed because of the visual scars they would leave on cherished city landmarks. It is significant to note that some of the most vocal protests targeted the Panhandle Freeway – the crosstown route that would have had the least impact on established neighborhoods since it would mostly run through Golden Gate Park. San Franciscans regarded the freeway as an assault on sacred ground and stated that even DOH-proposed tunneled routes through the park would be unwelcome intruders. In one of the first cases of freeway opposition from within the government, the Recreation and Park Commission of San Francisco passed a resolution opposing surface construction anywhere in Golden Gate Park (Lathrop, 1969, 3). Amid growing concerns over the freeway's visual impact, the Panhandle Freeway was deleted from the revised "Trafficways Plan" of 1955 "pending further study" (*The Master Plan*, 1963, Transportation Plate 1).

As public opposition to the remaining "Trafficways" proposals and ongoing DOH projects swelled in the late 1950s, anti-freeway voices began to appear on the Board of Supervisors. Supervisor William Blake, who had campaigned against the Junipero Serra Freeway in Sunset, remarked that "San Francisco's most precious asset is its breathtaking geographical beauty," an asset that should not be jeopardized by "concrete monstrosities" (Nolte, "Obituary," 1996, E-2). By 1959 – the year that the Central Freeway was completed – 97 different San Francisco organizations and neighborhood coalitions had come out in opposition of further freeway construction. This unprecedented "freeway revolt" reached its climax on January 26, 1959, when the Board – in a unanimous decision – dealt a deathblow to "Trafficways" by withdrawing its support of six of the nine proposed freeways.[5] While cities around the nation were busily constructing more highways to facilitate growth, San Francisco had placed visual considerations on higher ground than economic ones and had halted

construction. JP Sinclair, a traffic engineer for the heretofore unchallenged DOH, remarked that he had "never seen anything like it" ("Board Kills Plans," 1959, 1-2).

Second Round: Crosstown Routes Revisited (1959-1966)

Stunned by its first major highway-building setback, the DOH struggled to understand the reasons behind the freeway rejections and its implications for future plans in San Francisco. Public relations campaigns were stepped up, such as the statewide distribution of DOH booklets entitled *Freeway Facts*. Many parts of *Freeway Facts*, like the frequently asked questions section, seem to be especially written for San Franciscans:

> Will the freeway spoil the appearance of the town?
> On the contrary, design of structures . . . will often enhance the appearance of the community.
> What effect will the freeway have on nearby homes?
> Interviews with people living alongside freeways showed that to 5 out of the 10 freeways made no difference, 3 out of 10 ***preferred*** their location and 2 out of 10 did not like it (11).[6]

Immediately after the Board's 1959 decision, state and city agencies began revising rejected freeway plans. In November 1960 they published "Trafficways in San Francisco – A Reappraisal." The report called for an extra 9.2 miles of freeways within the city limits – considerably less than what was proposed in 1951 – to be built in the form of resurrected Golden Gate and Panhandle freeways (*Annual Report: 1960 –1961*, 11).

Transportation planners were conscious of aesthetic objections to the original routes and presented costly, elaborate freeway proposals designed to have minimal visual impact on their surroundings. The Golden Gate Freeway is a particularly good example.

Various proposals put forth in the 1960s for constructing the Golden Gate Freeway through Fisherman's Wharf and the Marina District. Options included a subaqueous tube in the bay (FW) and tunnels under major Wharf thoroughfares (FT). In Freeway Studies: Panhandle Parkway and Golden Gate Freeway, 1966.

Original "Trafficways" designs of the 1950s called for an above-ground route: with the exception of Russian Hill (where a tunnel would be bored), many of the city's most famous neighborhoods would have been bisected and vistas along the waterfront would have been destroyed. Revised plans in the 1960s proposed extensive use of tunnels. In one design option, the Golden Gate Freeway would pass through Fisherman's Wharf in three tubes – one each under Bay, North Point, and Beach streets – and continue under Marina Boulevard in a cut-and-cover tunnel. The estimated cost for this underground route was a staggering $311.9 million. In another, more elaborate proposal, a "subaqueous" double-decked tube would be constructed in the bay paralleling the northern waterfront. The top of the tube would be partially visible, but the route would not interfere with the landscape of the Marina and Fisherman's Wharf areas. In this design scheme, transportation considerations were entirely subordinated to aesthetic concerns: subaqueous tubes would have been ineffective in solving northern San Francisco's transportation problems as they would have been inaccessible to local traffic. With estimates totaling over $250 million, the tubes would also have been extremely cost-inefficient (*Freeway Studies*, 1966, no page).

The state's gestures of goodwill did not last long, however: as the 1960s progressed, the DOH rejected the costly tunneled alternatives and renewed its push for mostly above-ground Panhandle and Golden Gate freeways. Along the northern waterfront, San Franciscans once again rallied around the issue of aesthetics to denounce new infrastructure projects. By this decade, the San Francisco preservation cause had gained a powerful new ally: the tourist industry. The post-war years had seen a dramatic increase in visitors drawn to San Francisco's unique cityscape and beautiful neighborhoods. Businesses dependent on the tourist trade feared that freeways would destroy the city's appeal. George Burger, president of the Fisherman's Wharf Merchant's Association, explained, "The northern waterfront is the heart of San Francisco's lucrative tourist industry. The mere notion of putting an eight-lane freeway through here would be laughable if it weren't such

a deadly serious proposal . . . The tourist does not come to San Francisco to look at the James Lick Freeway" (Wax, 16). Recognizing these concerns, the San Francisco Chamber of Commerce began retracting its support for aboveground routes.

As in the years leading up to 1959, San Francisco neighborhood coalitions banded together to create a citywide alliance opposing all future construction. Initial DOH attempts to design freeways with minimal visual impact failed due to lingering public suspicion from the 1950s and the department's unwillingness to fund exorbitant projects. Citizens concerned about the city's aesthetics found a sympathetic ear in Mayor John Shelley, who spoke out against the destruction of cherished neighborhoods by telling a transportation planning conference, "In our central cities we must, at all cost, preserve the value of 'there'" (Leavitt, 1970, 202). Many on the Board of Supervisors had similar views: on March 21, 1966, the Board voted to reject plans for a Golden Gate and Panhandle freeway. As Department of City Planning members noted, the Board decision "removed virtually all possibility" of future freeway construction in San Francisco; subsequently, the federal Bureau of Public Roads withdrew all funding for future interstate construction within the city (*Annual Report: 1965–1966*, 6; Issel, 1999, 612). The second and most decisive "freeway revolt" had succeeded.

Conclusion: An evolving yet continuing revolt

San Francisco's battle against freeways by no means ended with the Board decision of 1966; indeed, it continues to this day. Nevertheless, the nature of the battle changed significantly: San Franciscans began to clamor for the demolition of *existing* infrastructure after their victory against the DOH. Upon taking office in 1967, Mayor Joseph Alioto focused the city's attention on the much-hated Embarcadero Freeway and called for its demolition. In the following years, the city authored studies on the Embarcadero's demolition and the Board passed multiple resolutions endorsing such plans. The report *A Transportation System for the Embarcadero Area* was released in 1974, recommending pulling down the freeway from Beale Street to Broadway. Instead of justifying the report's pro-demolition conclusions with traffic analyses, the Department of City Planning explained that the freeway was jarringly out of place on the waterfront *landscape* and therefore had to be torn down: "The Embarcadero Freeway was designed as a route through a then primarily industrial area . . . Over time the presence and design of the Embarcadero Freeway became increasingly incongruous in light of the Embarcadero Center and massive private redevelopment . . . The elevated freeway is presently an impediment to the proper relationship between the City and the Bay, an aesthetic barrier, an environmental disruption . . ." (*A Transportation System*, 1974, 20-21).

In 1980, an amendment was added to the San Francisco Master Plan which recommended the freeway's demolition and stressed "the reconfirmed value of the waterfront as a unique resource which needs to be reintegrated with the city." Totally bypassing the Embarcadero Freeway's importance as an arterial for downtown and northern city traffic, reports from the 1970s and 1980s evaluated the route for its aesthetic worth. When the Loma Prieta earthquake of 1989 occurred, there was little debate about implementing two decades'-worth of pro-demolition policy. The first sections of the freeway were pulled down in 1991.

The Embarcadero Freeway was only one of the San Francisco arterials severely damaged in 1989 – on the other side of downtown the double-decked Central Freeway buckled and collapsed. Its northern-most section, from Fell Street to Golden Gate Avenue, was unsalvageable and torn down in 1991 after Mayor Art Agnos threatened to sue Caltrans[7] if it constructed a replacement (Epstein, "Years," 1996, A-13). The fate of the remaining Central Freeway was to become the biggest transportation issue in San Francisco during the 1990s. The arterial was an important conduit for crosstown and Marin-bound traffic: on an average weekday, 80,000 cars traveled on the elevated structure (Nolte, "Traffic," 1996, A-1). However, after the freeway's upper deck was removed due to sustained earthquake damage, San Franciscans decided they liked less freeway in Hayes Valley and in 1998 passed city Proposition E to demolish the entire structure north of Market. City reports cited "a number of environmental, quality of life, . . . and urban design issues" including "the visual quality of the overpass as it crosses Market Street and enters Hayes Valley" as reasons behind the success of Proposition E (*Central Freeway Replacement Project*, 1999, 5-6).

Instead of a freeway, Proposition E called for construction of a surface level boulevard where the elevated structure stood, along Octavia Street from Market to Fell. Freeway ramps would touch down at the intersection of Market and Octavia, and northbound crosstown traffic would flow onto the boulevard to connect with the Oak and Fell corridors. In official reports, heavy emphasis was placed on the aesthetic qualities of this new roadway: San Francisco's Central Freeway Project Office stated that "drivers, both local and those passing through, should feel they are in a special environment and look forward to the experience." A 1999 publication, *Central Freeway Replacement Project: Octavia Boulevard Alternative*, suggested that "Octavia Boulevard could be celebrated with one or two gateway features" – such as portals – at its intersection with Market. With its vivid descriptions of landscaping and design features, the

report nevertheless seemed to neglect how the elegant thoroughfare would handle displaced Central Freeway traffic. San Franciscans, however, saw boulevard construction as a great aesthetic improvement, and in 2001 a city task force even considered proposals to tear down the Central Freeway at Bryant, six blocks south of the Market and Octavia intersection (King, 2001, A-14).

Despite the preserved beauty of San Francisco, nearly half a century of virulent freeway opposition has begun to take its toll. The city's limited and aging freeway system – composed only of the Bayshore, Southern, Southern Extension (I-280 east of Bayshore), and the truncated Central freeways – is becoming less capable of handling increased traffic flow. Surface streets have become hopelessly congested. In Sunset, where no further attempts to construct a freeway were made after the Junipero Serra route was rejected, 19th Avenue has doubled as a major local arterial and a crosstown route for through traffic. As a result, the thoroughfare has become "beyond clogged": it carried around 120,000 cars per day in 1997, approximately the same amount of traffic as the eight-lane MacArthur Freeway in the East Bay (Nolte, "Grand Central," 1997, A-13). In 2000, a report published by the city's Transportation Authority painted a gloomy picture of future traffic conditions. The authority predicted that by 2020, travel times within the city would increase by 20 percent, vehicle miles traveled would increase by 18 percent, and the number of hours spent in congested traffic would increase by 200 percent. To counter this projected gridlock, traffic planners recommended $6.6 billion in transportation improvements (Epstein, "Even $5.6 Billion," 2000, A-16).

Unfortunately, San Francisco's aversion to further freeway construction has made comprehensive transportation improvement for the automobile nearly impossible. In a city where 1960s-era partially tunneled freeway plans were enough to stir the ire of residents, contemporary solutions to traffic problems have become *prohibitively* elaborate and expensive so as to avoid controversy over possible aesthetic impact. In April 2000, the Transportation Authority urged the Board of Supervisors to consider a freeway plan very similar to the 1951 "Trafficways" report: its components included routes along Van Ness Avenue, 19th and Park-Presidio avenues, and the Oak and Fell corridor.[8] However, these freeways would be *entirely* tunneled. Once completed, the city would have a 7.5 mile-long network of underground "supercorridors," including a continuous 5 mile-long tunnel in Sunset from the San Mateo County line to the Presidio (Epstein, "Underground," 2000, A-1). Construction would be a herculean task, as it would disrupt San Francisco's busiest thoroughfares possibly for years. The tunnels would also be hazardous: they would be bored through unstable ground consisting mostly of dune sand. The extravagance and costliness (absolute cost has not yet been determined) of the plans will likely sink it, and the proposal will run into legal problems: the Master Plan now contains language prohibiting further highway capacity within city limits. Yet, despite being dismissed as "exotic ideas" by the *San Francisco Chronicle*, these tunneled freeway plans are the only possible automobile transportation solutions in a city that absolutely refuses to construct more surface routes (Epstein, "Even $5.6 Billion," 2000, A-16). In an era where increasingly serious traffic issues are wholly subordinated to aesthetic considerations, San Francisco has an impossibility as its only freeway alternative. Such is the cost of preserving America's "last lovely city."

Author Notes

1. The Bayshore Freeway encompasses present-day US 101 and Interstsate 80, which begins at the San Francisco city line by Candlestick (3Com)Park and feeds into the Bay Bridge in downtown.
2. The best example of a built primary thoroughfare is Geary Boulevard, which was widened into Geary Expressway between Gough and Masonic during Western Addition redevelopment projects.
3. Aside from the Southern Crossing, another proposal, drafted in 1949, called for a transbay span from Telegraph Hill to San Leandro. The new bridge would intersect with the Bay Bridge on Yerba Buena Island, "which would be leveled for a giant freeway interchange" (*The Case for the Southern Crossing*, 15).
4. Italics added.
5. The rejected freeways included the Western Freeway (a 1955 modification of the Panhandle route), the Junipero Serra Freeway in Sunset, the Park-Presidio Freeway in Richmond, the Central Freeway north of Turk, the Golden Gate Freeway, the Mission Freeway north of 30th, and the Crosstown Freeway.
6. Italics and bold appear as in actual text.
7. In 1972 the DOH was absorbed into the new California Department of Transportation, or Caltrans.
8. In another throwback to the 1950s, Senator Dianne Feinstein in 2000 recommended a new study for a Southern Crossing. Between the 1950s and 2000, the Southern Crossing had been studied and rejected five times. Feinstein and bridge supporters, however, proposed a transbay span far to the south, near the San Francisco International Airport (Nolte, "$600,000 Sought," A-16).

References

1. *Annual Report: 1949-1950*. San Francisco Department of City Planning, 1950.
2. *Annual Report: July 1, 1960 – June 30, 1961*. San Francisco Department of City Planning, 1961.
3. *Annual Report: July 1, 1965 – June 30, 1966*. San Francisco Department of City Planning, 1966.
4. "Board Kills Plans for 6 SF Freeways." *San Francisco Chronicle*, January 27, 1959, pp. 1-2.
5. *Capital Improvement Plan: 1956-1957*. San Francisco Department of City Planning, June 1956.
6. *The Case for the Southern Crossing*. Published by the City and County of San Francisco, presented at the Assembly Committee Hearing, Sacramento, May 7, 1949.
7. *Central Freeway Replacement Project: Octavia Boulevard Alternative*. San Francisco Department of Public Works, 1999.
8. Epstein, Edward. "Even $5.6 Billion Plan Means Gridlock in SF Future." *San Francisco Chronicle*, April 19, 2000, p. A-16.
9. Epstein, Edward. "Underground Toll Roads Urged for SF." *San Francisco Chronicle*, April 18, 2000, p. A-1.
10. Epstein, Edward. "Years of Traffic Mess Likely." *San Francisco Chronicle*, August 21, 1996, p. A-13.
11. "Four Freeways Added to 1945 Master Plan." *San Francisco Chronicle*, July 18, 1951, p. 16.
12. *Freeway Studies: Panhandle Parkway and Golden Gate Freeway*. California Division of Highways and Public Works, 1966.
13. Issel, William. "Land Values, Human Values, and the Preservation of the City's Treasured Appearance: Environmentalism, Politics, and the San Francisco Freeway Revolt." Pacific Historical Review, Vol. 68, No. 4, November 1999, pp. 611-646.
14. King, John. "SF Has Startling Plan for Van Ness." San Francisco Chronicle, December 4, 2001, p. A-15.
15. Lathrop, William. "The San Francisco Freeway Revolt." Transportation Engineering Journal, March 1969. Special published copy, available at Koshland San Francisco History Center, San Francisco Public Library.
16. Leavitt, Helen. Superhighway – Superhoax. Garden City, NY: Doubleday & Co., 1970.
17. The Master Plan of the City and County of San Francisco. San Francisco Department of City Planning, 1963.
18. Nolte, Carl. "$600,000 Sought For New Study of Southern Crossing." San Francisco Chronicle, March 9, 2000, p. A-16.
19. Nolte, Carl. "Grand Central Restoration." San Francisco Chronicle, April 15, 1997, p. A-13.
20. Nolte, Carl. "Obituary – William C. Blake." San Francisco Chronicle, March 15, 1996, p. E-2.
21. Nolte, Carl. "Southern Crossing – Boulevard of Broken Dreams." San Francisco Chronicle, January 20, 2000, p. A-1.
22. Nolte, Carl. "Traffic Planners Baffled by Success." San Francisco Chronicle, September 13, 1996, p. A-1.
23. Report on Activities of the Department of City Planning, 1947-48. San Francisco Department of City Planning, 1948.
24. Rose, Mark H. Interstate. Knoxville: University of Tennessee Press, 1990.
25. A Transportation System for the Embarcadero Area. San Francisco Department of City Planning, 1974.

Dinyar Patel

Dinyar Patel is a senior majoring in International Relations, and for as long as he can remember, he has maintained a strong interest in city planning -- especially the peculiarities of San Francisco's urban environment. He is currently completing an honor thesis which examines imperial New Delhi and the political implications of its construction for the British Raj. While at Stanford, he has held internships with the US State Department in Thailand and India, and has studied abroad at Oxford and in Shanghai. He would like to thank the staff at the Koshland San Francisco History Center, San Francisco Public Library, for their assistance and help with his research for this article, and the San Francisco Planning and Urban Research Association (SPUR) for giving him access to their extensive transportation library.

Architecture & Film

Elevators:
Negative Urban Space, the Futurist City as Avatar, and Fritz Lang's *Metropolis*

This paper attempts to examine Fritz Lang's 1927 film Metropolis *through the lens of the Italian Futurist movement's preoccupation with urban space. Futurist theorists such as F.T. Marinetti called for the cities that would express, in concrete form, the Futurist's unique aesthetic of universal dynamism. The products of this theory, in the form of Antonio Sant'Elia's drawings, exhibit a preoccupation with "negative urban space." This is a term adopted from art practice that refers to spaces (e.g. streets, alleys, passageways) defined by being between the boundaries of contained spaces; negative spaces do not have walls of their own.*

Metropolis *presents a vision of a future metropolis that we can see as Futurist in appearance and character. But for any Futurist archetypes it may contain, Lang's future city also embeds a critique of the Futurist concern for negative urban space. In Lang's film, negative space is both the result and manufacturer of the oppression of the working class. The perpetual state of being in an emptiness, an in-between space, is an expression of the alienation and isolation produced by the modern metropolis the Futurist wanted to perfect.*

Casey Riffel

When F.T. Marinetti published "The Founding and Manifesto of Futurism" in 1909, the link between internal mind and external expression had already found its way into European artistic consciousness. In the decades bracketing World War I, the written word and the painted image were becoming tools of psychological and perceptual expression instead of strict mirrors. A portion of this new artistic development was due to the increasing urbanization of European life. But, as Marinetti believed, architecture as an art had failed to develop to the point that it reflected modern sensibilities, and thus the cities of modern Europe remained lodged in the past. For the Futurists, many aspects of contemporary architecture were contemptible for their nostalgia and fascination with the picturesque and the natural, such as the canals of "passéist" Venice and the art nouveau style flourishing in Paris.

Because of its distaste for all things traditional, the Futurist movement visualized its ideal metropolis as a direct projection of its desire for universal dynamism and the obliteration of history. In order to accomplish this, it required a theory of architecture that visualized the city as the concrete expression of the Futurists' mindset. The document codifying this architecture was published in 1914 as "The Manifesto of Futurist Architecture," and in it, Marinetti and Antonio Sant'Elia proclaimed an architecture that would create "our cities, which ought to be the direct and faithful projections of ourselves" (Sant'Elia and Marinetti, 2). The products of this theory, in the form of Sant'Elia's drawings, exhibit a preoccupation with space defined entirely by what I will call "negative urban space." This is a term adopted from art practice that refers to spaces (e.g. streets, alleys, piazzas) defined by being between the boundaries of contained spaces; negative spaces do not have walls of their own. One enters a negative urban space (hereafter, simply "negative space") only by exiting another space (1).

But a Futurist city was never built; instead, it is useful to see elements of Futurist thought embedded in other modernist texts. Fritz Lang's 1927 film *Metropolis*, one of the most towering achievements of the German studio system and one of the most famous films of all time, presents a vision of a future metropolis that we can see as Futurist in appearance and character. But for any Futurist archetypes it may contain, Lang's future city also reads as a critique of the Futurist concern for negative urban space. In Lang's film, negative space is both the result and manufacturer of the oppression of the working class. The perpetual state of being in an emptiness, an in-between space, is an expression of the alienation and isolation produced by the modern metropolis the Futurist wanted to perfect.

The City as Avatar

For the Futurist movement, the prototypical city was the quintessential backdrop of the quintessentially modern life. But the nostalgia for the incorporation of the natural into the urban

was completely absent from Futurist design. As Tisdall and Bozzolla point out, "the city had…replaced the landscape as the setting for the excitement of modern life" (Tisdall, 1978, 121). From this city, the Futurists dreamed of extracting all natural or picturesque elements. In his manifesto "Against Past-Loving Venice," Marinetti pleads that the "Grand Canals, widened and dredged, must become a great commercial port" (Marinetti, 1991, 64). Much of Venice's despised romance derived from the presence of the ubiquitous canals; they were Venice's metonymic face. Marinetti's great port imagined "wide roads built over canals that have finally been filled in" (Marinetti, 1991, 64).

Take, for example, Balla's *Street Lamp* [1] which takes a literal approach to Marinetti's rhetorical cry, "Let's murder the moonshine!" The lamp swells to occupy the frame, dominating the backgrounded moon in every respect. The close-up perspective on the lamp lends a comparatively imposing size to the modern electric light, overpowering an object which is both physically and culturally much more important. Thus the main murder taking place in Balla's painting concerns the moonshine itself. The overwhelming brightness of the street lamp seems to illuminate the moon, an inversion of roles most pleasing to a Futurist aesthetic. The moon, relegated to a corner and deprived of its own light, is cast in the incandescent glow of the modern city. But as much as the street lamp is projecting light outwards, it also appears to function by drawing light to itself. The individual photons are arranged in the spherical pattern associated with the emission of light, but the direction of their U shape implies a motion inwards, towards the center of the lamp. Balla imbues the lamp with its own gravity, expressing the pull of the modern city for Futurist vision.

1. Balla, Street Lamp. 1909.

Thus, we begin to see the relationship between the Futurists and the city that characterized their vision. But, as Boccioni's "The Street Enters the House" [2] demonstrates, early Futurism was still using its ideas "as a bludgeon to beat the old" (Tisdall, 1978, 123). The painting shows a harbor ringed by old houses, skewed and contorted in a Cubist style. Although the chaos and fragmentation certainly espouses the Futurist need for dynamism, Boccioni's rectangular houses still speak of the old Europe. The motion in the painting, akin to a vision of an earthquake, is "still the solid reality of square stone blocks. They envisaged a Futurist city life, but not the Futurist city" (Tisdall, 1978, 123).

By the time Sant'Elia joined the Futurist crowd and generated a view of actual Futurist architecture, Marinetti had given concrete expression to the vision already implicit in his ideas. In his additions to Sant'Elia's "Manifesto of Futurist Architecture," Marinetti proclaimed that "architecture must be understood as the endeavor to harmonize, with freedom and great audacity, the environment with man, that is to say, to render the world of things a direct projection of the spirit" (Sant'Elia and Marinetti, Proclamation 8). This proclamation creates a lens through which to view Futurists' visions of the city, which I will later apply to *Metropolis*. The "spirit" (of the Futurists) will cast its image directly on architecture, like Balla's street lamp lighting the moon. Viewing the city in its physical and spatial sense, as an entity defined by its structural components (buildings, streets, parks, etc.), Marinetti makes the assumption that architecture is a tool with which to change "the world of things"—that is to say, the city itself (2).

2. Boccioni, The Street Enters The House. 1911

Thus emerges the concept of city as a mirror image of the Futurist mindset, which effectively renders the city an avatar of Futurism's obsession with speed, dynamism, and historical destruction. I employ the term "avatar" somewhat ironically since it generally refers to the human embodiment of an abstract concept; in the Futurist example, we see a decidedly impersonal embodiment of abstract concepts, which are expressed primarily in manifestos—in a fiery rhetorical language designed explicitly to drive human emotions (3). Nevertheless, the term accurately reflects Marinetti's concern that the city, by means of its architec-

3. Sant'Elia, La Città Nuova, detail 1914

ture, be a concretization and physical articulation of the Futurist aesthetic.

Sant'Elia himself expressed this sentiment, before Marinetti's hand came into play, when speaking of "an architecture whose raison d'être lies solely in the special conditions of modern life, whose aesthetic values are in perfect harmony with our sensibility" (Sant'Elia and Marinetti, 5). The city is thus not just an avatar in general, but an avatar for the modern (and, implicitly, Futurist) way of thinking, and as such, must represent a complete break from history. Like Naples, any old city must be disemboweled and paved over; the rise of the Futurist city should obliterate the historical foundations from which it rises. "It must be as new as our frame of mind is new" (Sant'Elia and Marinetti, 5).

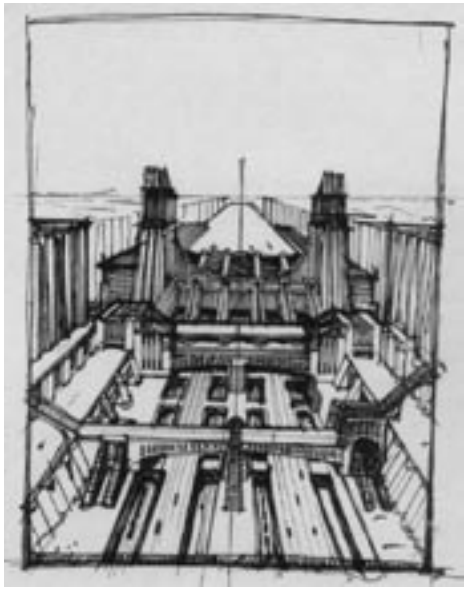
Negative Space in the Futurist City

Building on the city as avatar, it is useful to examine the visions of the Futurist metropolis to find the concepts that characterize the Futurist mindset. In addition to the universal dynamism, which finds expression in massive transportation hubs, vast communication networks, a multitude of bridges, and light building materials, the Futurist city contains a concern for negative space.

To adequately characterize negative space in terms of a city, it is necessary to speak briefly of buildings in the abstract. Visualizing an empty, flat plain, stretching to all the horizons under an empty sky, we see that there is one infinite and unbounded space. Out of this infinite space, we then create a building by erecting a series of connected walls covered by a roof of some sort. As Simmel suggests, the fundamental characteristic of a wall (along with a bridge) is division or separation. From the large pre-existing space, a building carves out a new, smaller, self-contained space which is necessarily a strict subset of the larger space. At the same time, walls necessarily have an outside. This outside, which consists of the non-partitioned remainder of the original space, can now be defined either as being outside the presence of walls or as not being inside the presence of its own walls; it is negative space (4). Two expressions of such negative space are integral to the Futurist city, where they become "negative urban spaces:" vertical space, as created by skyscrapers; and the streets.

The Futurist city projects upwards, into the sky. Marinetti calls to "raise the level of the city" (Sant'Elia and Marinetti, 10). This desire for the ascendant in the city relies on the buildings built "of reinforced concrete, of iron, of glass, of paste-board, of textile fiber . . . which make possible maximum elasticity and lightness" (Sant'Elia and Marinetti, Proclamation 1). Flexibility and lightness in architecture allow true skyscrapers to emerge, as Fritz Lang observed in New York City.

But for all their celebration of lofty towers, the Futurists were just as preoccupied with how the "city in the sky" represented a decisive break from the old. Just as "perpendicular and horizontal lines, cubic and pyramidal forms" are denounced for being "static, heavy, oppressive, and absolutely alien to our new sensibility," the new building materials are important as "substitutes" (Sant'Elia and Marinetti, Opposition 4; Proclamation 1). It is just as important to be high above the ground as it is to *not* be close to it. The desire to break away from gravity is analogous to the desire to break away from the past. Here we see Marinetti and Sant'Elia implementing their city as an avatar, creating a vision that flees from the earth at the same time as it flees from history.



4. Sant'Elia, La Città Nuova, train station and airport 1914

5. Sant'Elia, La Città Nuova, power plant 1914

But Sant'Elia's sketches for *La Città Nuova* do not present the interiors of these new cityscapes. Instead of presenting life from the inside, Sant'Elia chooses to focus on the exterior of the buildings; the actual mechanics of how these buildings support themselves is masked behind the smooth façades and opaque windows (5). At the same time, the buildings are remarkably unadorned, appealing to Marinetti's notion that "nothing is more beautiful than the steel frame of a house in construction" (Marinetti, 1991, 89). Nor are the views those of a person standing inside one building and looking outwards onto the city. The eye of Sant'Elia's sketches is disembodied, floating somewhere between street level and the tops of the buildings, al-

6. Sant'Elia, La Città Nuova, apartment buildings

ways positioned at just the right angle to give the correct perspective on the buildings themselves. But the sketches do not show a gaze looking into something; instead, they are looking upon something. The buildings, in fact, have no interiors, and the attention of the viewer is meant to pass airily through their glass crowns.

Consider one detail of *La Città Nuova* [3]. In this sketch, Sant'Elia presents a building of unspecified purpose. The building is much longer and taller than it is wide, capturing the elegance of the straight line; indeed, the building's elevators that jut out in at least four places can resolve into simple lines with distance. The narrow building also minimizes horizontal space. These external elevators are also present in a sketch of an apartment building [6]. As methods for transcending the limitations of gravity, the elevators became protagonists in Sant'Elia's city; they "must not hide like lonely worms in the stair wells; the stairs, become useless, must be done away with and the lifts must climb like serpents of iron and glass up the house-fronts" (Sant'Elia and Marinetti, 9). The emphasis is decidedly on motion through the vertical plane. It seems that negative space is strongly identified with vertical space.

Although the human element is lacking from *La Città Nuova*—the presence of any people, either directly or indirectly (in vehicles) generally exist solely to give a scale for the building's height [4, 6]—there exists a concern for the spaces in which people move between the buildings. Streets, despite any physical reality they may have in terms of concrete, are defined as the space between two buildings. Systems of transportation are central to the Futurist city, and Sant'Elia's detail [3] has no less than four different street levels, two of them comprised of bridges with open-air steel grid walls; all of the levels are open to each other. The sketch of the apartment building also displays two street levels [6]. The combined train station and airport [4] exemplifies Sant'Elia's assertion that "the street, which will no longer stretch like a foot-mat with the porters' lodges, but will descend into the earth on several levels, will receive the metropolitan traffic and will be linked, for the necessary passage from one to the other, by metal walkways and immensely fast escalators" (Sant'Elia and Marinetti, 9). The boarding platforms are below ground level and yet are completely open spaces connected with foot bridges and escalators.

7. Sant'Elia, La Città Nuova, buildings with airplane

The necessity of multiple-level streets reads as an extension of the Futurists' manic desire to break with history. Technically, expressions of negative space are present in Venice, the bane of Futurist thought. Buildings still rise above ground level, and streets still exist between them. But just as the Futurists perceived meager amounts of tall buildings to be defects, and dreamed of supple skyscrapers to represent their abandonment of history, they also took the streets and expanded them both upwards and downwards. Being able to move vertically between multiple streets metaphorically continues Marinetti's murderous designs towards history.

Combining the vertical spaces with the street, Sant'Elia devises what he terms a "tumultuous abyss" (Sant'Elia and Marinetti, 9). The image is striking in its accuracy; we are tempted to imagine the buildings as the walls of a chasm opened by an earthquake. The stone faces of the buildings' sides are blank, impassive surfaces, "without painting and without sculpture," and far below at the bottom flows a churning river of automobiles and pedestrians (Sant'Elia and Marinetti, 9). This explains Sant'Elia's removal of interiors from his sketches; the buildings are meant to resemble actual cliff faces.

This contrasts markedly with Boccioni's *The Street Enters the House* [2], where the open space of the harbor reaches out and, with its own hands, drags a person out from the rings of houses that crumble at the perimeter. Boccioni expresses a fundamentally urban experience: that of the invasion of the private by the anonymous public. Tisdall quotes Boccioni's catalogue for the 1911 exhibit: "The dominating sensation is that which one would experience on opening a window: all life, and the noises of the street rush in at the same time as the movement and the reality of the objects outside. The painter

8. Erich Kettlehut, sketch for Metropolis, cityscape, 1924?

9. Erich Kettlehut, sketch for Metropolis, city center, 1924?

does not limit himself to what he sees in the square frame of the window as would a mere photographer, but he also reproduces what he would see by looking out on every side from the balcony" (Tisdall, 1978, 43).

Here, the boundary between the inside and the outside, the positive and the negative, is tenuous and often breached. Boccioni feels that the essential activity of the street must sweep in like wind when the window is thrown open. Sant'Elia neither agrees nor disagrees with this sentiment; in *La Città Nuova*, the inner spaces are completely segregated from the outer (6). There can be no interaction because the outside is omnipresent, and the cliff walls that rise above the street are impermeable (7). In the new city, windows serve only to make the buildings more magnificent, instead of providing a gateway between the inner and outer worlds.

Metropolis as Futurist Vision

With its destructive and bellicose tendencies, the Futurist movement ended up as the unfortunate partner of Fascism. But it seized on several salient characteristics of the European experience in the early 20th century; the Futurist preoccupation with universal dynamism (basically the apotheosis of speed) masked an anxiety over the increasingly frenetic and mechanical pace of life stemming from rampant industrialization. Whereas Marinetti and Sant'Elia attempted to embrace and thus apotheosize modernity, Fritz Lang's *Metropolis* confronts industrialization and finds instead a nightmare of the mechanized worker. In creating his future dystopia, Lang and his production team realized a metropolis that, at least superficially, bears striking resemblance to the Futurist city.

In some of his sketches for *Metropolis*, art director Erich Kettlehut produced cityscapes which share many characteristics with Sant'Elia's sketches [8, 9]. The high glass-and-steel buildings stand above multi-level streets with raised bridges for both pedestrians and automobile traffic; there is even an airport built at the top of the tallest building [8]. But in some ways, Kettlehut's drawings are more in line with a Futurist aesthetic than Sant'Elia's; Kettlehut's streets are filled with both automated and human traffic, and there are even open shops where people enter and exit the ground floors [9]. These initial visions of what would become The City of the Sons certainly align us to a Futurist perspective on *Metropolis*.

The film itself presents a city which embodies many of Futurism's principles. The remarkable cityscape shots that Lang scatters throughout the film maintain a strong resemblance to Kettlehut's drawings, with the high, flat buildings and multi-leveled streets and raised bridges [11, 12, 13]. Just as Kettlehut incorporates motion into his drawings, Lang's city courses with the exalted universal dynamism that Sant'Elia's sketches lack. But there are two other principles which most strongly link *Metropolis* to Futurism, and, ironically, they are not visual.

First, *Metropolis* generates a city without a history. Even if the Upper City is not rebuilt by every generation, it is certainly a city to which time means nothing. At some indefinite point in the past, the Upper City was constructed, somebody first switched on the machines, and the Workers' City was carved out of the earth. Its inhabitants are ignorant of their past. Freder, before Maria's visit at the film's opening, knows nothing of the workers who toil underground. Initially, the workers do not have a mythology to explain their indentured present; this is precisely what Maria provides for them in her recounting of the story of the Tower of Babel (8). Even then, it is merely a

10. Erich Kettlehut, sketch for Metropolis, pyramidal buildings, 1924?

11. Still from Metropolis, The City of the Sons by day, 1927

12. Still from Metropolis, The City of the Sons by night, 1927

Christian allegory, with no basis in the actual history of the Workers' City. The only hint of past historical processes is in Rotwang's house, which was clearly built some time ago. Ironically, it contains the most advanced technology in the city.

The second link between *Metropolis* and Futurism is verticality. Sant'Elia concludes the body of his Manifesto with the decree, "let us sink squares into the ground, raise the level of the city" (Sant'Elia and Marinetti, 10). Indeed, this is exactly what happens in Lang's *Metropolis*. The only square given significant screen time is the one in the Workers' City where the children gather during the flood [14]. The city realized in Kettlehut's drawings is really just the tip of an iceberg; the world of air, glass, and steel is what I will call the Upper City. But below the ground lurks a complicated structure on which this Upper City rests.

As the plot progresses, the viewer, along with Freder, learns that the complex comprising the actual Metropolis consists of four levels. Above ground is the Upper City, and at the tops of the buildings there exist the pleasure gardens where Maria visits Freder (9). She leads him down directly below the surface, where lie the vast arrays of machines which power the Upper City. Below the Machine Level, connected by a series of elevators and rickety stairs, is the Workers' City. Once the narrative shifts to the leader of the Upper City, Joh Fredersen, the viewer gets a glimpse of the lowest level: the Catacombs. These levels are related only vertically, like strata.

Thus the concern for vertical space already developed in our discussion of negative space finds literal expression in the world of *Metropolis*. This proves to be another application of the city as avatar, in that the class hierarchy in Joh Fredersen's mind finds itself concretely manifested in the architecture and layout of the cities; vertical space is the avatar for class division. These four layers resemble four stories in any skyscraper in a Futurist city. But for Lang, who was much more concerned with the deleterious effects of machinery, the vertical distances serve only to express the disenfranchisement of the workers. In this sense, Lang opens up the smooth empty streets of Sant'Elia's sketches and peers down into the earth; he examines the entirety of the iceberg of which Sant'Elia's buildings are only the tip. A closer examination of the levels of the Metropolis will reveal how Lang's film raises issues swallowed in the Futurist's blind extermination of the past (10).

As in any building, these levels have a specific relationship: any given layer supports the one above it while simultaneously repressing and being dependent on those below it. The Upper City relies on the electric power produced by the Machine Level but refuses to incorporate any of the actual machines into the architecture above the surface. The Machines (and, implicitly, the Upper City) in turn require the presence of the human workers to operate them, while at the same time subjugating the workers to an impoverished existence in a blocky city of stone even deeper in the earth (11). Lang brilliantly expresses this relationship in the scene where Freder operates the clock-like machine [15]. Without a human to perform the extraneous task of lining up the hands, the machine will not work; after 10 hours of this mindless task, Freder is exhausted (12). The relationship between the two levels is parasitic, not symbiotic. Josef Bottlik's movie poster also demonstrates this relationship by alluding to the Atlas myth [16].

Underneath the Workers' City are the catacombs, where Maria preaches her doctrine of rebellion and revolution. Being the oldest, this level is crumbling, unoccupied, and dead. Interestingly, it is the only part of the Metropolis for which the viewer ever sees a map. The catacombs represent the suppressed spirituality and humanity of both the workers and the Upper inhabitants. They contain the history,

13. Still from Metropolis, traffic in the streets, sky-bridges, and airplane, 1927

14. Still from Metropolis, flood in the Workers' City, 1927

15. Still from Metropolis, Freder at a machine 1927

16. Josef Bottlik, UFA Poster for Metropolis, Berlin, 1927

mythology, and hope that Maria is able to tap when inciting her rebellion; they have the same effect on the workers and on Freder. On the other hand, the accessibility of this vivacity increases with physical proximity to this deepest level. The workers are the first group Maria converts; second is Grot the Foreman, who controls the Machine Level; finally, by the conclusion, Joh Fredersen himself is converted (13). The power of the catacombs slowly works its way towards the surface as the movie progresses.

All the movement in the film is directed in the vertical plane, both thematically and literally. Maria must ascend to retrieve Freder, who in turn descends into the earth, only to return to the surface after the flood. Joh Fredersen also descends in order to glimpse Maria's preaching; the Robot moves repeatedly between the levels, inciting both crowds to riot. Most important is the upward surge of the workers once moved to action. They overflow the tiny elevators which slowly rise up to the Machine Level. Freder and Rotwang, who have been moving towards each other during the movie, enact their struggle on top of the cathedral. Having defeated the inventor, Freder is able to descend again to the level of the square.

Lang's film transforms the Futurist development of the city as an avatar into a vision of the machine actively pressing modern man into the earth. The skyscrapers and multilevel streets express only the alienation of all the Metropolis's citizens. Consider the "tumultuous abyss," which in Sant'Elia's formulation expresses the vitality of the city. In *Metropolis*, the activity in the abyss is only the frothing madness of a crowd incited to riot by the Robot or the horde of children fleeing from the flooding of their homes. The children, led by the real Maria, flee upwards on rickety stairs to escape the rising waters. Here, negative urban space is something to be climbed, an obstacle that must be overcome in order to survive.

Metropolis uses its physical city as a metaphor for class struggle and the repression (and eventual recovery) of history. The same system of towering buildings, elevators, and bridges that make possible the luxurious life of the Upper City also reads as a critique of the Futurist mindset. The history of the Metropolis is both suppressed and recovered by motion through the vertical spaces of the city. It is clear that, for Lang, only by ample and frequent communication between the various levels of the Metropolis will any semblance of happiness occur. Transgression across the vertical boundaries (Freder's assimilation into the Worker culture, the ascendance of both Maria and her doppelganger Robot into the Upper City, the flood that forces the children out of the Workers' City) is the only means of unification. In this sense, the elevators metaphorically represent the greatest hope. Although Lang questioned the clean simplicity of the Futurist vision, he nonetheless recognizes that some of the elements the Futurists advocated most vehemently may serve to level society. It thus seems possible that some of Futurism's ideas have the potential to destroy the things they support, although, in the end, this might have been exactly what Marinetti wanted.

Author Notes

1. A trace of this concept exists in contemporary idiom, where one can "go out into the world." We see that the juxtaposition of "out" and "in" hints that although someone might be entering a space ("the world"), he can only do so by leaving whichever space he currently occupies.
2. This assumption, of course, rests on the assumption that urban planning, as the arrangement of these structures, has a minimal effect

on the city. This assumption is probably naïve, but my paper will follow Marinetti in exploring a one-to-one relationship between architecture and the city, as an investigation into the actual relationship between architecture, urban planning, and the city would require another paper.

3. As will be discussed later, the inhabitants of Sant'Elia's La Città Nuova are virtually nonexistent.

4. This is not, for our purposes, an important distinction; it is essentially a matter of perspective. Defining negative space as outside the walls of another space emphasizes the separation between the two spaces. Defining it as its own space lacking bounding walls chooses to stress the infinitude inherent in negative space; in this sense, only one gigantic negative space exists in the universe.

5. This is an interesting contrast to other visions of the modern metropolis, where the interiors of old buildings often show like the ribs of a starving man. Consider Rilke's The Notebooks of Malte Laurids Brigge, where a gutted house in Paris shocks and terrifies the narrator.

6. Here arises a point of debate over the true Futurist bent of Sant'Elia's work. The young architect was recruited by Futurism as Sant'Elia joined the movement (Tisdall, 1978, 125-8). As Tisdall notes, Boccioni's vision of the violent interpenetration of the positive and negative spaces is representative of Marinetti's vision in the "Technical Manifesto": "Space no longer exists: the street pavement, soaked by rain beneath the glare of electric lamps, becomes immensely deep and gapes to the very center of the earth" (Tisdall, 1978, 42). As much as this echoes Sant'Elia's "tumultuous abyss," Marinetti still felt compelled to alter Sant'Elia's "Message" from the La Città Nuova opening into a manifesto. And Sant'Elia's absolute division of positive and negative space certainly ignores Boccioni's ideas. But a full investigation of these issues must remain for another paper.

7. This partitioning of space extends even within different negative spaces; the city "was a complex in which buildings, mechanized and pedestrian traffic, were to share a space but function separately" (Tisdall, 1978, 131).

8. I abstain from dealing with the romantic ideals expressed in the plot of Metropolis, as the actual mechanics of how the revolts occurs, who is implicitly blamed, and the eventual reconciliation of the workers with the children of the Upper City are not relevant to my argument (in addition to being somewhat absurd or naïve).

9. These pleasure gardens would certainly be the place for someone like Marinetti, who was wealthy and lived in decidedly non-Futurist surroundings.

10. A Marxist perspective reveals an interesting comparison; if all history is simply class struggle, then the Futurists would, by exterminating history, also eliminate class struggle. Lang, however, understands that even a mechanical utopia of speed and movement must still be supported by the types of machines the Workers operate.

11. The Workers' City has many of the qualities that the Futurists denounced—especially the building materials.

12. This is precisely the reason Rotwang creates the Robot—to eliminate the Machines' dependence on men. If his scheme had succeeded, then the relationship would have been broken and the Workers' City could have been destroyed without consequence.

13. This leaves out Freder and Rotwang, who, as the respective hero and villain, are presented as exceptional cases.

References

L'opera completa di Boccioni (1969). Milan: Rizzoli Editore.

Marinetti, T. F. (Flint, R. W. and Coppotelli, A. A., trs.) (1991). Let's Murder the Moonshine: Selected Writings. Los Angeles: Sun and Moon Classics.

Meyer, E. C. (1995). The Work of Antonio Sant'Elia: Retreat Into the Future. New Haven: Yale University Press.

Tisdall, C. and Bozzolla, A. (1978). Futurism. New York: Oxford University Press.

Sant'Elia, A., and Marinetti, T. F. "Manifesto of Futurist Architecture". Class Handout.

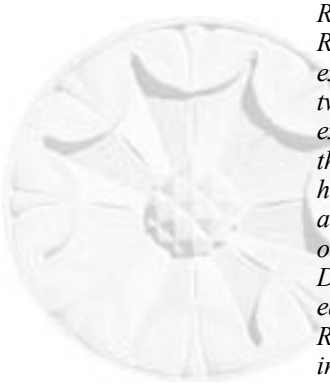
Simmel, G. "Bridge and Door" from *Simmel On Culture,* Frisby, D., Featherstone, M. (eds.). London: SAGE Publications, 1997.

Casey Riffel

Having grown up in rural Snohomish, Washington, Casey Riffel has a hankerin' for the big city. He plans to write his honors thesis on simulations of nature in present-day New York City. This paper is the project that sparked his academic interest in cities. June, 2005 will see Casey receive a B.A. in Interdisciplinary Studies in Humanities with a Creative Writing minor and an M.A. in Sociology. He thinks student publications are really something else and should always be funded. Nothing at all would be possible without Casey's friends over at Seaforth Boat Rentals (namely Justin, Peter, and Genny). He would like to thank his mentors Paul Robinson from the Department of History and Elena Coda from the Department of Italian.

Biophysics

The Study of RNA Polymerase Pausing by Optical Traps



RNA polymerase (RNAP) is the crucial enzyme responsible for the DNA-directed synthesis of RNA, known as transcription. RNAP's movement is discontinuous, which can directly affect the expression of any gene in an organism's genome. This research studies RNAP transcription along two biochemically engineered DNA templates. Repeated pauses maximize the pause data per experiment because one RNAP transcribes identical pause sequences multiple times. We tracked the movements of individual Escherichia coli *RNAPs with an optical trap, a device that utilizes a highly focused infrared laser to manipulate single molecules. Optical trapping experiments give accurate measurements of RNAP position on an order of several base pairs, and allow for the observation of rare events that would otherwise be averaged out in large populations of molecules. Despite the identical nature of the pause sequences, our data suggest RNAP behaves variably at each one, sometimes not pausing at the expected site at all. The variability might arise from the RNAP repeatedly transcribing the pause sequences, thus slowing transcriptional motion and thereby increasing the probability of an RNAP pause at subsequent sites.*

Becky Wong

RNA polymerase (RNAP) occupies a fundamental position in the biology of living organisms. It is responsible for DNA transcription, the first step of genetic expression, and is therefore highly regulated by the cell. If transcription is not regulated, cellular processes can become unsynchronized, possibly leading to uncontrolled cell growth and cancer. Transcription is also regulated to ensure the fidelity of the RNA transcript. If RNAP incorporates incorrect bases and induces errors into the nascent RNA, defective proteins will be made that can be lethal to the cell. In addition, because the function of RNAP is so integral to life, certain antibiotics target RNAP enzymes in disease-causing bacteria. The antibiotics disrupt the enzymes' processivity and kill the infectious bacteria.

RNAP pause types and mechanisms

The goal of this research is to study the pausing motion of RNAP as it transcribes DNA. RNAP transcription is not a continuous process; it is punctuated by temporary pauses in movement along the template DNA. During pauses, the RNAP takes 100-1000 times longer to build the RNA chain[1]. Pauses can therefore range from a few seconds to a few minutes duration. The cell implements different pausing mechanisms based on the conditions and nutrient requirements in the cell. For example, in *Escherichia coli,* pausing might couple transcription and translation by synchronizing the interaction between RNA synthesis and ribosome movement in the coding portions of mRNA. This coupling may well regulate the attenuation and termination of bacterial operons, a genetic unit of related genes and control sites[2].

Specific sequences of DNA have been known to induce RNAP pausing, and different sequences have varying probabilities of RNAP pausing. It is thought that there are two mechanistically different types of sequence dependent RNAP pausing (Figure 1): Class I pauses, such as the *his* leader pause sites, which are dependent on the formation of a 5-basepair-stem and 8-nucleotide-loop RNA hairpin structure; and Class II pauses, such as the *ops* (operon polarity suppressor), which involve the RNAP backing up on the DNA and blocking the active site with nascent RNA[3].

Experimental methods for studying RNAP pauses

Previously, RNAP pauses have been studied by biochemical experiments. These assays have been done in bulk solution that involve trillions of RNAP molecules transcribing simultaneously. Under these conditions, data is collected from so many molecules that fine details of motion are averaged out and cannot be distinguished. In contrast, single-molecule experiments measure motions of single molecules. These studies help to identify the uniqueness of individual enzymes and distinguish between subpopulations of RNAP molecules. Research at the

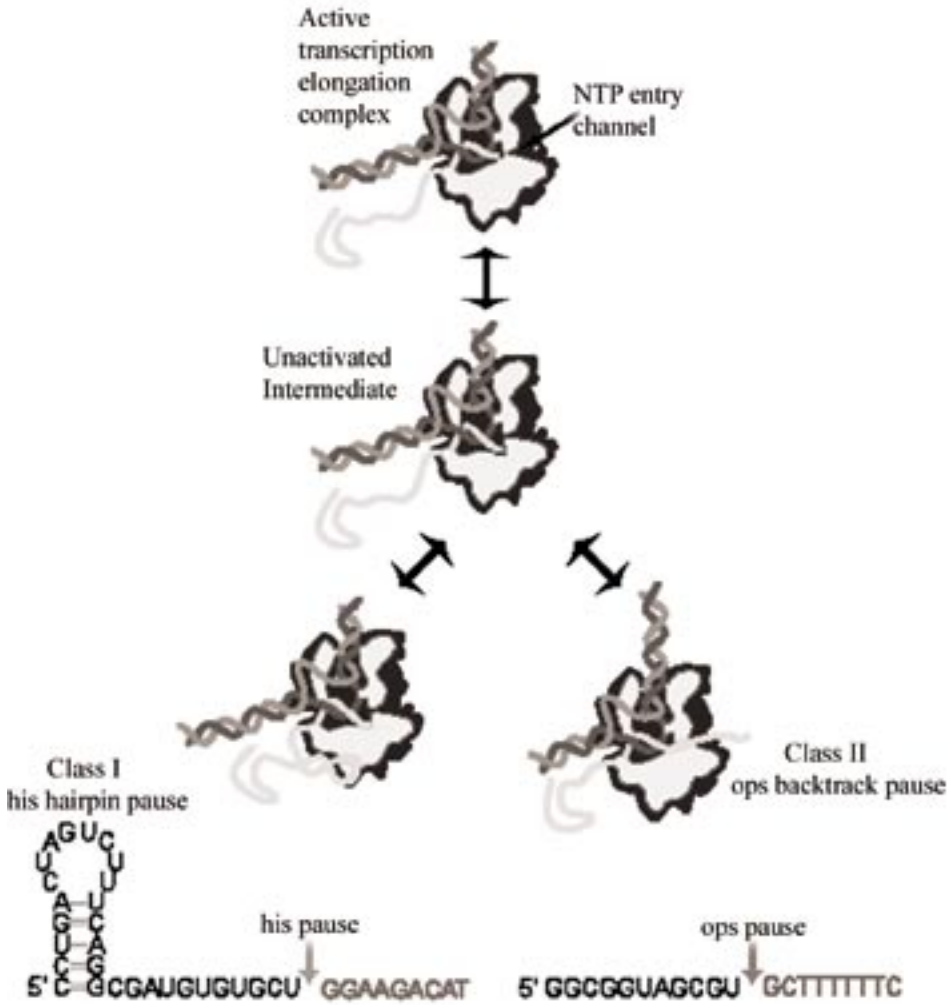


Figure 1 Model for RNAP pause states.

RNAP bound to the DNA is actively transcribing the coding strand to generate messenger RNA. From the active transcription elongation phase, RNAP enters an unactivated intermediate state and undergoes small rearrangements in the active site to form two different types of stabilized and sequence dependent pauses. The Class I *his* pause (left) occurs with the formation of an RNA hairpin in which the single stranded RNA bends and base pairs with itself. The hairpin in the nascent transcript interacts with the polymerase, disrupting its motion. The pause occurs before the downstream DNA. In the Class II *ops* pause (right), the RNAP moves backward along the RNA-DNA hybrid so that the RNA transcript is excluded from the NTP entry channel of the polymerase.

single molecule level elucidates properties of individual polymerases, such as pause probability and pause duration. From this data, we can investigate whether particular RNAP molecules tend to pause more frequently or longer than others.

Recently, new techniques have been developed to study single RNAP molecules moving along the DNA[4]. These techniques have high temporal resolution and allow for the study of low probability events from single molecules which would otherwise be averaged out in bulk studies. One of the main single molecule methods is optical trapping, pioneered by Ashkin *et al.* [5]. The technique uses a focused laser beam to manipulate microscopic particles without mechanical contact. By applying this technology in conjunction with biochemical techniques, we can study transcription with a level of detail that is impossible to achieve with conventional bulk experiments alone.

Advantages of concatenated pauses studied in optical trap experiments

This research studies RNAP transcribing a DNA template of concatenated pauses, or a string of repeated and identical pause sequences. This method offers many promising insights into the polymerase's pausing mechanisms. Previous optical trapping experiments have shown that random, stochastic RNAP pauses occur while the enzyme moves along the DNA[6]. The concatenated template will also allow us to study the behavior of an individual enzyme as it encounters the identical sequence of DNA multiple times. The mechanism by which a specific RNAP acts subsequent times to transcribe the same stretch of DNA has not yet been determined. This information will help elucidate the theory of molecular individuality. The idea theorizes that each molecule will act differently from another molecule even though they are identical in sequence and structure.

Using optical traps to study the pauses of RNAP gives us specific data on the transcriptional motion of the enzyme. The trap works because the focused infrared laser attracts particles toward its focal region. Thus, trapping the bead connected to the RNAP will allow us to apply a range of forces on the enzyme as it transcribes the DNA template and encounters long-lived pause sequences. The behavior of the polymerase under an applied force will provide clues regarding the mechanism of the pauses. For example, the formation of a hairpin structure in nascent RNA has been known to induce pauses in the *his* pause sequence[7]. Although it is unclear how the hairpin causes a pause, there are a few theories. One idea is that the hairpin interacts allosterically with the RNA strand thereby influencing the binding of molecules at other sites on the RNAP and disrupting transcription. Another possibility is that the hairpin structure might physically displace the polymerase from the

DNA and nascent RNA. The application of force on the polymerase allows us to distinguish between these two theories.

If the RNA hairpin dislodges the RNAP from the DNA and RNA transcript, the pause is mechanical and force-dependent. This should result in a greater effect of optical trapping forces on RNAP's pausing efficiency. In the case of the *ops* pause sequence, the backtracking motion of the RNAP at a pause site should be highly force dependent. The optical trap experiments could either pull the RNAP forward toward the downstream DNA or pull the RNAP backwards toward the upstream DNA. This data should help us determine the translocated distance and illuminate the finer mechanisms involved with entering and exiting the pause state.

In this study, we attempted to make a concatenated template of DNA with multiple, identical stretches of DNA ligated together. Instead of tracking the motion of RNAP transcribing just one pause site, we can study transcription of eight concatenated pause sequences of DNA. Thus each run of the polymerase is effectively eight single-pause experiments in one. The concatenated templates help us characterize the pausing and velocity of RNAP-mediated transcription with resolutions approaching the order of several base pairs.

Materials and Methods

Ligation of pause sequences and two flanking regions of DNA

To construct the DNA template, eight separate oligonucleotides (three complementary pairs) were used. The pieces ranged in size from 73-78 base pairs. Each of the complementary oligos was annealed, creating complementary overhanging ends. The sticky ends of each of the three annealed oligos were ligated in a specific orientation of Segment A, Pause, and Segment C. Furthermore, the (A-Pause-C) sequence will be referred to as a pause unit. Each of the pause types, *his* and *ops,* were made into separate pause units, with the pause sequence flanked by the two other sets of complementary oligos. Single pause units were ligated into a cloning plasmid (pCR Blunt from Invitrogen), transformed into *Escherichia coli* Stbl2 cells (Gibco BRL), and selected on LB agar plates supplemented with Kanamycin antibiotic resistance (50 μg/ml).

Preparing the pause units for cloning

The pause units were concatenated to form a repeating sequence of eight pause units using a protocol established by Carrion-Vazquez *et al.*[8]. The method capitalizes on the unique qualities of two restrictions sites, BamHI and BglII. In addition to being palindromic restriction sites, which ensures orientation specificity, restriction digests at BamHI and BglII sites leave complementary, compatible cohesive ends used to screen for desired ligation products. Using the iterative process of digesting, ligating and cloning, we multiplied the number of repeated pause units in a single vector. The first round of cloning yielded a 2mer (two pause units), the second round produced a 4mer, and the third round gave us the desired 8mer. The concatemer of eight repeated pause units was cloned into the pRL732 plasmid behind the strong T7A1 promoter in order to perform transcription assays.

Constructing labeled template for optical trapping experiment

In order to construct transcription templates for the assays in the optical trap, we needed a linearized piece of DNA with a labeled end to attach to the microscope slide. Each of the *his* and the *ops* templates were digested with ScaI restriction enzyme to linearize the plasmid. Then, a terminal transferase reaction was performed (Roche-DIG Oligonucleotide 3'-End Labeling Kit, 2nd Generation) to label both 3'ends with a Digoxigenin-ddUTP. Since the microscope slide is coated with anti-digoxigenin, the digoxignin on the DNA attaches the template onto the surface of the slide through an antibody linkage.

Two types of experimental geometries can be produced from each of the *his* and *ops* transcriptional templates. Digestion of the DIG-labeled DNA with Eco109I restriction enzyme removes the upstream 3' end of the DNA, leaving the downstream end of the DNA tethered to the microscope slide. Thus as the RNAP transcribes the DNA, the optical trap pulls the RNAP in the opposite direction that it would normally move as it transcribes DNA. In the hindering load regime, the DNA tether becomes shorter as it transcribes DNA (See Figure 2).

In a similar fashion, digesting the doubly-DIG-labeled DNA with SapI cleaves the downstream 3' end label of the DNA and consequently allows for the optical trap to create an assisting force on the RNAP as it moves along the DNA template. In this case, the DNA tether becomes longer as the polymerase transcribes DNA.

Optical Trapping

To perform optical trapping experiments, the DNA was attached to the coverslip of a microscope slide via a digoxigenin / antibody linkage. A 500 nm polystyrene bead was attached to RNAP by a biotin molecule and avidin protein linkage, a strong noncovalent bond, and the RNAP was bound to the DNA. Once ribonucleotides (rNTPs) are added, RNAP starts transcribing and its position on the DNA template is recorded as a function of time. Using the optical trap, an effective force can be calculated from the template versus time position of the RNAP. By collecting data at varying forces and calculating the efficiency of pausing as a function of force, we should be able to distinguish between the allosteric

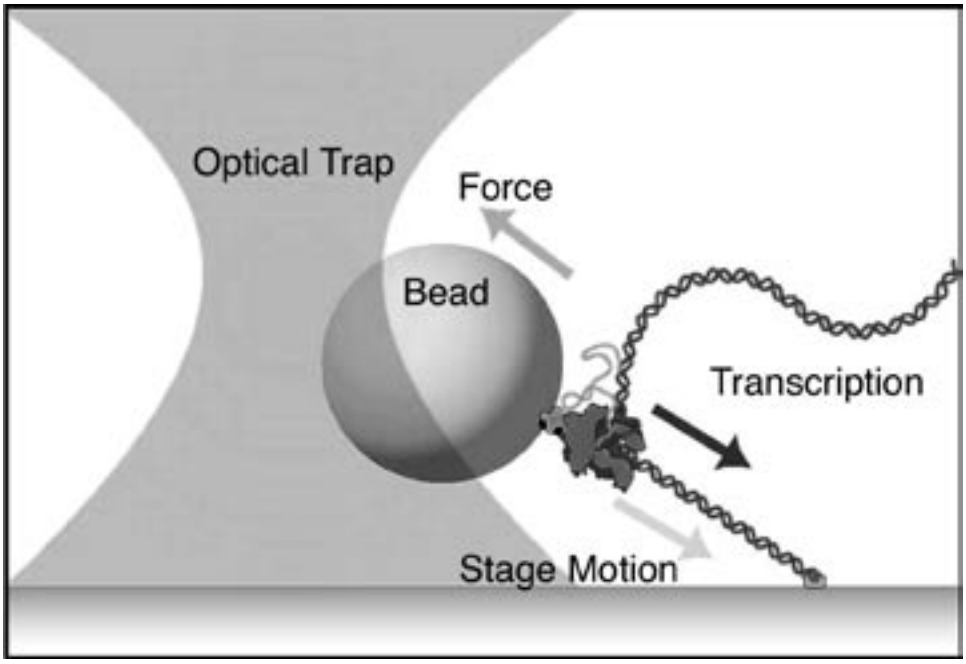


Figure 2 Schematic representation of a hindering load optical trapping assay (not to scale).

The light from the focused infrared laser, known as an optical trap, exerts a force on the 500 nm polystyrene bead that grows linearly as the bead moves from the center of the trap. This force holds the bead at a predetermined position from the trap center. The RNAP is attached to the bead via a biotin / avidin linkage and is bound to the double stranded DNA and the elongating RNA (gold). The downstream end of the DNA is attached to the microscope slide surface via digoxigenin and anti-digoxigenin. The microscope slide is mounted on a piezoelectric stage. The optical trap exerts force in the opposite direction of the polymerase's transcriptional motion; to compensate for the enzyme motion, the stage moves in the horizontal and vertical directions to maintain the bead's position in the trap and thus exert constant force on the bead.

and mechanical models of the RNAP pausing for the Class I *his* pause and observe backtracking events with the Class II *ops* pause.

The optical trapping experiments were performed in a low vibration, temperature-regulated clean room at 21.1 ± 0.5ºC. The optical trap used to collect data was fundamentally the same instrument as described in previous work from the lab[6]. A feedback-equipped, piezoelectric stage (Physik Instrumente, Karlsruhe, Germany) was used to allow three-dimensional positioning in order to maintain constant force on the bead. Further modifications included adding a more powerful and stable ND:YVO_4 laser at 1064nm, removing the Wollaston prisms, and replacing the interferometric detection system with a position sensing diode. Custom software was written in LabView (National Instruments) to run the automated microscope stage. The instrument has allowed the collection of data with motion resolved temporally at 1/50th of a second, spatial precision of ~1nm, and accuracy within 70 nm.

Results and Discussion

In summary, we constructed two types of templates with different repeating pause sequences: the *his* pause and the *ops* pause. Each of the types of pause sequences have been biochemically engineered to allow the RNAP to transcribe in either the assisting load or the hindering load regime in the optical trap. These four DNA templates offer a multitude of different experiments and opportunities to probe the mechanism and force dependencies of RNAP's pauses along the DNA.

The templates of the *his* and the *ops* concatamer of pause sequences were constructed as shown (see Figure 3). Sequence analysis confirmed that the DNA templates have eight concatenated pauses, and gel electrophoresis assays indicated the expected template lengths of 6.8 kilobases for the *his* pause template and 6.7 kilobases for the *ops* pause template.

The DNA sequence has been modified with the appropriate digoxigenin labels for the hindering load and the assisting load geometries under the optical trap. Recent data includes some partial traces of RNAP transcribing the region of DNA with the concatenated pause. The traces are considered partial because RNAP position measurements did not start precisely when the RNAP began transcribing the DNA.

We have recorded traces of RNAP transcribing the DNA sequence of concatenated *his* pauses under a 5 ± 0.8 pN hindering load from the optical trap (see Figure 4). The partial traces detect the position of the bead connected to the RNAP. The y axis indicates the template position (in base pairs) of RNAP transcribing the DNA as a function of time. The forward motion of the polymerase is indicated by the decreasing base pair position on the template over time. The RNAP moves from the end of the template toward the beginning, and the tether length of the DNA connected to the slide gets shorter. The motion is discontinuous, as transcription is interrupted by pauses. These pauses are indicated by the horizontal segments in the trace. RNAP motion is affected by the availability of free rNTPs which are incorporated into the elongating RNA transcript. Thus, these initial experiments were done under saturating 1mM concentrations of A, U, C, and G nucleotides. The high levels of rNTPs allowed for RNAP transcription to be observed.

Based on our sequence analysis, the *his* pause template should have its first pause when the RNAP has reached the template position of 2861 bp with the pause sequence repeating every 239 bp. The grey bars indicate possible locations of expected pauses. Shown in

the figure are two separate traces of the RNAP molecules transcribing the *his* pause DNA template. Although there is a positional uncertainty error of up to 200 bp, the traces suggest regularly spaced pauses of eight concatenated sequences. Periodic pauses are observed and the two traces appear to have some correlation in terms of aligning the pauses with each other.

It is interesting to note the variation between the traces. Although both traces are from the RNAP transcribing the same sequence of the DNA, different RNAP molecules behave differently at the same pause site. For example, the RNAPs did not always pause at the expected sites. At the template position around 1600bp, there is an expected pause site. Surprisingly, however, there did not appear to be a pause in the red trace, but there was a pause in the blue trace. This discrepancy between traces supports the idea of variable pause efficiencies, in which a sequence of DNA might have only a certain probability of inducing the RNAP to pause. Another explanation for the different pausing behavior at around 1600 bp is that with the blue trace, the long pause at 1900 bp might have predisposed the molecule for more pausing later in the template. A long pause might sufficiently slow the polymerase's transcriptional rate and allow it to enter the next pause state more easily.

The traces indicate various durations, or pause lifetimes, between different RNAPs as well as between different pause sites with the same RNAP. For example, at the template position around 1900 bp, the red trace indicates RNAP spent a few seconds in a pause, but in the blue trace, the RNAP spent orders of magnitudes longer in the pause. Other factors might have contributed to the polymerase in the blue trace displaying an exceptionally long pause at 1900bp. Previous studies in our laboratory have reported non-sequence-dependent pauses, separate from the sequence-dependent pauses studied in this project. Nucleotide

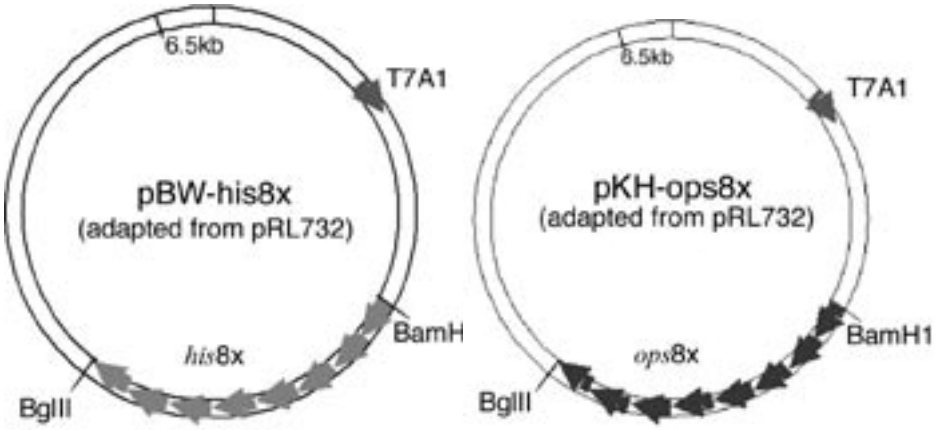


Figure 3 Diagrams of the *his* and the *ops* concatemer of eight pause sequences.

The plasmids are both over 6.5 kilobases long and have a T7A1 promoter. The pause sequences are concatenated, repeating eight times, occurring every 239 bp for the *his* pauses and every 227 bp for the *ops* pauses. Each of the eight pause sequences is flanked by two restriction sites, BamHI and BglII.

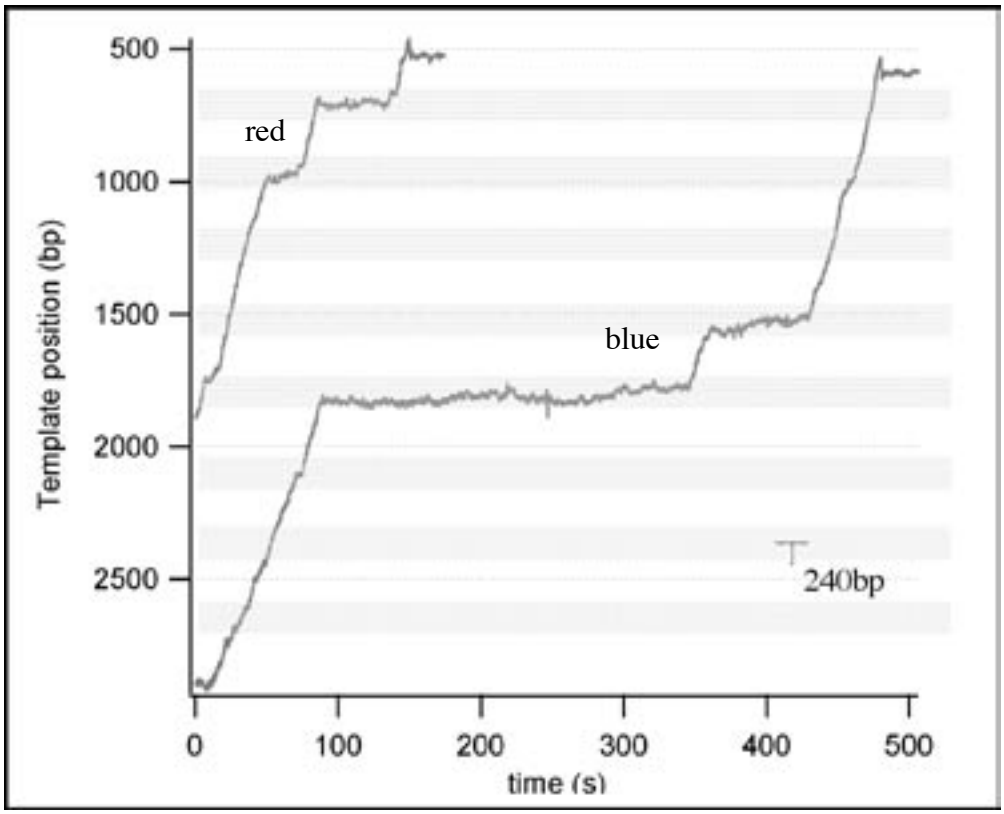


Figure 4 Traces of RNAP transcribing the DNA template

The y-axis gives the position of a single RNAP molecule transcribing a portion of a 4 kbp DNA template recorded as a function of time. RNAP movement is observed with 1 mM rNTPs concentration and hindering force of 5pN against the direction of transcription. There is a discontinuous motion punctuated by pauses on various time scales. In the hindering load regime, the polymerase moves farther along the DNA, the template position decreases, and the DNA tether length decreases.

misincorporation events in the growing RNA transcript can induce long-lived, non-sequence dependent pauses[9]. This type of pause induced by an error in the RNA could have contributed to the long-lived pause observed at around template position 1900bp in the blue trace. Also, non-sequence dependent pauses can occur ubiquitously throughout the template and might explain some of the early pauses occurring at template positions shorter than 1000 bp and the predicted start of our concatenated pause sequences[6].

Data have been collected under a limited range of conditions. Most of the data have been recorded at saturating rNTP concentrations (1mM of each A, C, G, and U nucleotides), and we have observed transcription punctuated by pauses. We plan to lower the rGTP concentrations to enhance pausing at our pause sequences. Furthermore, there will be experiments using the other geometry in which the RNAP will be subjected to an assisting load as it transcribes the DNA in the optical trap. In total, there will be four different combinations of assisting and hindering forces using two different types of pause sequences, the *his* and the *ops*.

Our findings suggest pausing at the expected pause sites, as well as random pauses between the known, predicted pause locations. There are multiple types of pauses that can be observed in any single trace of the RNAP transcription. Because each trace shows the motion of a single RNAP transcribing the pause sequences, a large number of traces need to be obtained for each combination. The pausing statistics will help to determine the probability of the RNAP entering a pause state when it encounters a pause site and the average lifetimes of the pauses. With the DNA template of identical and repeated *his* pause sequences, we have found that the RNAP pausing varies between successive pause sites.

References

1. Yager, TD, von Hippel, PH. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Washington, D.C.: American Society for Microbiology, 1987:1241-1275.
2. Chan CL, Landick R. Dissection of the *his* Leader Pause Site by Base Substitution Reveals a Multipartite Signal that Includes a Pause RNA Hairpin. *J Mol Bio* 1993; 233:25-42.
3. Artsimovitch I, Landick R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *PNAS* 2000; 97:7090-7095.
4. Schafer DA, Gelles J, Sheetz MP, *et al.* Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature* 1991; 352:444-448.
5. Ashkin A, Dziedzic JM, Yamane T. (1987) Optical Trapping and Manipulation of Single Cells using Infrared Laser Beams. *Nature* 1987; 330:769-771.
6. Neuman K, Abbondanzieri E, Landick R, *et al*. Ubiquitous Transcriptional Pausing is Independent of RNA Polymerase Backtracking. *Cell* 2003; 115:347-447.
7. Mooney RA, Artsimovitch I, Landick R. Information Processing by RNA Polymerase: Recognition of Regulatory Signals during RNA Chain Elongation. *Journal of Bacteriology* 1998; 180:3265-3275.
8. Carrion-Vazquez, M. *et al*. Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci USA* 1999; 96:3694-9.
9. Shaevitz, JW, Abbondanzieri EA, Landick R, *et al*. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* 2003; 426:684-687.

Becky Wong

Becky Wong is majoring in biological sciences and pursuing a co-terminal masters degree in chemical engineering in 2004. Her love for science began in high school as a lab assistant at Mendel Biotechnology, Inc. At Stanford, she spent three years conducting single-molecule biophysics research in the laboratory of Dr. Steven Block and received a URO Major grant in 2002. The Block Lab members were instrumental in the success of her project, with special thanks to Kristina Herbert, Ravi Dalal, Dr. Keir Neuman and Professor Block. Becky would also like to thank her parents and older brother of Castro Valley, California for their continued support.

Knowledge, Sensitivity, and HOPES

by Albert Keung

SURJ is devoted to exploring original undergraduate field and laboratory research, but it also strives to highlight the real world, social, and personal implications of scientific research that are a consequence of academic exploration. The Huntington's Outreach Project for Education at Stanford (HOPES) is a student-run website (http://HOPES.stanford.edu) founded to provide the public with accurate, scientific information about Huntington's Disease in a manner comprehensible to the lay person. According to the HOPES website, Huntington's Disease (HD) is a genetic disease that "causes deterioration of the nerve cells in the brain, prompting significant changes in one's ability to think, feel, and move." The primary mission of the HOPES team is to describe and explain the science behind this disease while remaining aware of sensitive HD-related issues. Unlike Stanford students involved in other research endeavors, the HOPES team focuses their efforts on reading research papers, understanding new studies on HD, and distinguishing between supported science and hearsay. After compiling the most current information, they tackle the difficult task of communicating the latest HD research in a clear and sensitive manner. The HOPES team strives to be responsive and mindful of the fear, frustration, and hope that mark the lives of those affected by HD. Their work demands strong literature research skills, clear writing ability, and, most importantly, sensitivity to those affected by the disease.

The HOPES Team
Back row, left to right: Tucker Cunningham, Ruth Lo, Shawn Fu, Kaizad Cama.
Front row, left to right: Taylor Altman, Matt Stenerson
Not pictured: Melissa Schapiro.

HOPES was founded in 2001 after Professor William Durham, the Bing Professor in Human Biology, Dept. of Anthropological Sciences, was contacted by a friend with a family member affected by HD. The family was frustrated at their inability to understand their doctor's explanation and the scientific literature's explanation of the disease. Professor Durham emailed his friend back, breaking down the details of HD in his explanation. The family's positive response to his email prompted Durham to create a widely available and comprehensive resource that individuals and families affected by HD could understand without needing a background in biology. With private funding and the support of the Human Biology internship program, HOPES was born. It is currently under the faculty supervision of Professors Durham, Joanna Mountain, and Ronald Barrett and is based in the Anthropological Sciences building.

The story behind the inception of HOPES gives just another example of how information, such as the term "genetic disease," along with the vocabulary used in scientific literature, is incomprehensible to the majority of the world. HOPES' mission is to break through the barrier of specialized language and provide accurate and scientifically rigorous information on HD in a style understandable to the common person. Professor Durham elaborates by saying, "We make no prior assumptions. We try to walk people from first premises. It makes [HD] less fearsome, less threatening to understand what's behind [HD]."

In addition to the accessibility of HOPES' language and information, the very nature of HOPES as an online resource is important, as realized by Melissa Schapiro, a HOPES senior majoring in biology and psychology. From her previous studies of autism and from talking to parents with autistic children, Schapiro found a "huge divide between what the scientific world knows and what the general public knows." Furthermore, she learned that most parents did not read peer-reviewed research journals but instead relied heavily on websites and Internet discussion groups. Part of what motivated her to join HOPES was a desire to combat the distorted and often incorrect online information about medical conditions. Schapiro states that, "People deserve correct, unbiased scientific information that they can understand even if they don't have a degree in science, or any college degree at all. HOPES provides a valuable resource to the Huntington's community because it offers carefully researched articles

that patients and their families can actually understand."

With its well-organized and easily navigable structure, the HOPES website reflects Schapiro's sentiments of accessibility. It includes helpful features like hyper-linked words in articles which can take the reader to glossary definitions or other web resources when clicked. Clear graphics help illustrate concepts like genetic inheritance and nerve cell biology. Interactive Flash components, like the "Brain Tutorial," further enhance the learning process by teaching through a storybook format with step-by-step illustrations. HOPES is also working to translate the website for non-English-speaking readers.

Internet technology is only one aspect of HOPES' approach to creating an informative yet accessible resource. The HOPES team also aspires to reach a broader audience, which includes young children whose parents or loved ones may have HD. Matt Stenerson, a former undergraduate student leader of HOPES, is currently working on a Flash online storybook called "For Kids," aimed at helping young children deal with the social and personal aspects of HD. "For Kids" tells the story of a child who is grappling with the social and emotional effects of having a father afflicted with HD. The story shows how education about HD empowers the child to deal with his feelings and the painful situation of having a family member diagnosed with HD. Stenerson also anticipates educational Flash games and puzzles in the future to help younger children learn about HD.

Stenerson, currently a HumBio course assistant, has been touched in many ways while working with HOPES. One of his most memorable moments was a late night correspondence with a woman whose husband had HD. Stenerson was surprised that she was still awake at 3 AM writing e-mails. She explained that she could only fall asleep each night by devoting her life to learning about HD and corresponding with others about the disease. The effect HD had on her and how it defined her life touched Stenerson and opened his eyes to the profound effects of the disease. He reflects that, "It wasn't just neurons and cells, but real people and families that were being affected by this disease." His aspiration to practice medicine was strengthened as he discovered the very personal, real world effects that HOPES and scientific knowledge have had in helping patients and their loved ones. Stenerson shared his feelings about HOPES and how it changed him: "It regenerated my hunger to achieve that goal [of being a doctor]. It gave me a real sense of worth, a feeling that I had actually put my education to good use."

Both the professors and the students working on HOPES identified numerous ways in which the project changed them. Beyond the academic learning in biology, epidemiology, and computer science, the HOPES students learn to distinguish hearsay and urban legends from peer-reviewed and supported science. They also develop teamwork and leadership skills from group interactions like peer-editing. Durham views the act of creatively collaborating and bouncing ideas off peers as an invaluable aspect of HOPES. Perhaps most important to Durham, students leave HOPES with a strong sense of responsibility and develop the maturity and ability to deal with delicate and complex issues in a sensitive way. "Knowledge is also responsibility," Durham said. "There is enormous responsibility to come face to face with these issues [involving HD], to explain the issues to the interested readership of the site, to not hide any issues, but to bring them up in a way that a painful message can be heard and appreciated and maybe rendered less painful by bringing it gently."

From the perspective of student experience, Durham views HOPES as a "kind of a prototype. It shows the kind of project students can direct and sustain, putting their education to work even as they are themselves working at that education." He also emphasizes the interdisciplinary nature of HOPES. He and the rest of the HOPES team believe students in any field can learn a great deal from working on the project and will contribute to an invaluable HD resource for the thousands affected by the disease. The HOPES team encourages interested students of all disciplines to visit the HOPES website for more information on how to join and contribute to the project.

Because of its many positive aspects, HOPES is a great example of what service projects can be. It provides a valuable and helpful resource for those affected by HD, and in creating this resource, students learn a unique approach to knowledge while simultaneously developing academically and personally. Team members graduating and leaving Stanford have been strongly influenced by their experiences in HOPES, and what they have learned will undoubtedly shape their future lives. The social impact of HOPES through both the website and its team members is immeasurable. As it gains broader notice, HOPES should serve as a paradigm for knowledge in action, applying science and education to benefit society.

PLoS: Leading the Open-Access Revolution

by Charles Feng

In October of last year, Patrick Brown, a professor of biochemistry at the Stanford Medical Center, stirred up a storm in the scientific world when he launched the Public Library of Science (PLoS) Biology (http://www.plosbiology.org). Like any other science journal, it contained research articles, editorials, features, and reviews. Why, then, would its website receive hundreds of thousands of hits the first day it opened, nearly bringing down the server? The research was ground breaking: the first issue announced an implant that could control monkey behavior and the development of genetic tests that discovered a new Asian elephant species. The caveat, however, was that PLoS Biology could only be accessed by those with a computer and an internet connection, but readers would not have to pay the exorbitant fees that usually come with reading original research articles.

A site license for an annual subscription to *Nature* is $1,280 a year (Stanford University would qualify as a site), chump change for a university with an endowment approaching $8 billion. Yet for many people off-site, *Nature* does not offer individual subscriptions, though individuals can purchase a glossy reprint of an article for thirty dollars each. In contrast, open access rests on the idea that everyone, from world-class scientists to grade-school students, can read materials from the comfort of their own desktops. Even though the articles are targeted for a very specific demographic (mostly research scientists), the allure of PLoS is the sense of egalitarianism that free access to knowledge bestows upon the populace. A woman with advanced carcinoma can decide which treatment will best suit her; a young teenager, determined to learn about the heavens, can launch a successful career in astronomy.

When Brown, Harold Varmus, and Michael Eisen founded PLoS in 2001, they posted a brief, pointed letter online, warning that scientists would not subscribe to, publish in, or review any scientific journal that did not make its full contents available after half a year. The letter, with the signatures of over 30,000 scientists from around the globe, seemed to be a harbinger of change. Although the gauntlet had been thrown, the Goliaths barely blinked an eye. *Science* agreed to release its articles after one year, while *Nature*, with a few exceptions (such as the mouse genome paper), held to the status quo, allowing access only to subscribers. The journals realized that, for many scientists, publication within their pages was far more important than the lofty ideals outlined in PLoS' letter.

According to Mary Buttner, digital resource manager at Lane Medical Library, the reason *Science* and *Nature* subscriptions are so expensive is because established journals, ingrained with the print model, are spending a fortune on technology support—site maintenance, programmers, and the like—in order to keep their websites up to date and to maintain their simple accessibility. Although each or its article is saturated with advertisements and pop-ups on all sides, the *New York Times* website, updated every few minutes, is barely breaking even. "Technology is expensive, and people still want to make money," said Buttner.

PLoS, on the other hand, hopes to become self-sustaining by implementing a novel business model. Already endowed with $9 million in seed money from the Moore Foundation, which is being used to defray operational costs for the first four years, PLoS is also charging authors $1,500 to have their articles published. The fee is small compared to costs of running experiments, which can run into the hundreds of thousands of dollars. Moreover, authors already pay fees for longer articles and color printing in established journals. Also, some institutions, such as the Howard Hughes Medical Institute and the University of California at Berkeley, have already agreed to pay this fee for scientists within their institutions**.**

The goal now is to develop the presence of PLoS to the point where publication in its pages signifies preeminent findings. A number of steps can be taken to more quickly establish PLoS within the scientific community. For example, the National Institute of Health (NIH) can recommend that its scientists publish in open access journals. After all, if the NIH is already funding a large amount of their research, the authors should not be forced to pay what are essentially middlemen, journals such as *Nature,* just so they can have access to previous research. Another step is to correct the notion that only older, established scientists will publish in open access media because PLoS does not have the name-brand recognition that is critical to gaining tenure. Buttner, however, believes that PLoS will garner respect soon enough: "Very soon PLoS will be competing with *Nature* and *Science*."

In the meantime, PLoS is expanding its horizons and is currently in the midst of launching PLoS Medicine. As the established journals quietly wait to see what happens, PLoS is busy moving onward, confident that its publication model will ultimately force other journals to embrace open access. "PLoS is the future, and the future is very soon," Buttner said.

Charles Feng is a senior majoring in human biology.

Overseas Research in the Netherlands: The Cultural Memory of World War II

Brittan Heller is a senior majoring in English, co-terming this year, and pursuing an undergraduate honors thesis through the Interdisciplinary Humanities Department. Sarah Allen sat down to talk to Britten about her past summer's overseas research experience in the Netherlands.

Sarah: Could you start out by talking about what you did overseas, where you were, and what you were studying?

Britten: I have always been really interested in World War II…and I knew I wanted to study literature because it's what I love, so I did an interdisciplinary honors thesis. It was able to incorporate aspects of not only literature, but also history and art and cultural theories of memory and representation, which I find really fascinating. I developed this project to study the cultural memory of World War II in the Netherlands. [I] went to the town of Leiden, a suburb of Amsterdam, [and] stayed there for ten weeks. My project meant that I had to travel, so I got a train pass and went all around the country. I was able to travel from the very top [of the Netherlands] all the way to the borders of Germany and France.

S: Why did you choose the Netherlands?

B: I picked the Netherlands because nobody has really studied it. I decided to go there, and find out what the cultural myths were, what people actually believe happened, how that compared with the history, and how those developed over time. I traced through the war memorials and what the designs of the memorials said, what the texts of the memorials said, and how they evolved from right after the war until the present day. I also used museums, but I didn't use museums for their content. I read them like texts. I read their layout, how things were situated, the colors [used], the presentation and what they said about the topic.

S: While you were there, were you interviewing people?

B: Yes, I had to go through the human subjects protocol [and] an online tutorial, submitted a protocol to a committee, then they reviewed it and sent back comments and questions. Since my project was potentially involving Holocaust survivors, they wanted to make sure that it was really safe, because [interviews] could put people under severe psychological trauma. I made a pamphlet of resources in case anyone became upset or distressed talking about anything related to the Holocaust, and got it translated into German, Dutch, and English. I actually had to use it once. The curator of one of the national museums became very upset and actually starting switching between German, Dutch and English. It was a little scary.

S: How did you benefit from researching overseas? What did you accomplish that you couldn't have accomplished here?

B: Personally, I think the experience of living overseas was more beneficial to me than any other type of research experience I've had, because it put me out of my normal environment. It made me very aware of the cultural and social heuristics that I use to navigate everyday life, because they become very evident when you're put in a different place and they don't apply anymore. It also made me very politically aware about what it means to have an American identity. People would come up to me and say, "After 9-11, why does America do this? Does America hate Islamic people?" It made me think very critically in ways that I wouldn't have if I had just stayed here.

Memorial, Kamp Westerbork: These bricks are part of a memorial on the Appelplatz, or area for daily line-up. There is one for every person who was transported through the camp and killed. The metal symbols on the tops represent the three major populations processed through the camp.

Perspectives on Overseas Research

by Chris Couvelier

There is no doubt that Stanford is one of the preeminent research institutions in the world, providing its students with unparalleled resources and opportunities to foster their diverse research interests. One of Stanford's research outlets that has maintained a remarkably low profile over the years is its overseas research opportunities. Recipients of awards, such as URO grants and Chappell-Lougee scholarships, often opt to conduct research projects, not on campus or even within the United States, but rather overseas. Overseas research invariably results in immensely rewarding experiences on both an academic and a personal level.

Carly Schuster and Nadiya Figueroa are two Stanford students whose research projects epitomize both the advantages and obstacles offered by research abroad. Long fascinated by Argentina's unique economic and political situation, Schuster, a junior, elected to complement her study abroad in Argentina with a research project. To that end, Schuster used her $2,500 Chappell-Lougee scholarship to examine microlending to women in the Argentine altiplano, a subject with ties to her independently designed major, Women and Development. Schuster first made arrangements to study in Buenos Aires through Stanford's Overseas Study Program (OSP), intending to "get a feel for the political situation and make vital contacts while being safely taken care of in the Stanford facility." To Schuster's dismay, she arrived in Buenos Aires to discover that the non-governmental organization (NGO), with which she had hoped to collaborate on her research, no longer existed, requiring her to actively search for other contacts and organizations. After her study in Buenos Aires, Schuster spent the summer in a rural region of northern Argentina, near its shared border with Bolivia. Schuster was able to closely examine the altiplano's unique system of microcredit—or what Schuster described as "a system of small loans aimed at women to foster grassroots economic development." Schuster observed a series of meetings between the microlending NGO and a female artisan cooperative in which terms of financial transfers were worked out. The strategic and political nuances of these interactions will serve as the focal point of Schuster's honors thesis, currently in progress.

For senior Nadiya Figueroa, a History and Cultural and Social Anthropology double-major, the decision to conduct research overseas was an easy one. A native Jamaican, Figueroa was disappointed by the somewhat surprising fact that Stanford did not have professors specializing in Caribbean studies. Rather than give up on her academic curiosity about the region, Figueroa proactively crafted a research project that permitted her to explore her academic interest in the Caribbean, specifically in her Jamaican homeland. Using her URO major grant, Figueroa sought to study the elite members of the Jamaican Diaspora, their motivations for leaving Jamaica, and how they perceived themselves contributing to Jamaica. Though Jamaica is a country of only 2.5 million people, some estimates suggest that between 2 and 3 million additional Jamaicans live outside the country. More troubling to Figueroa was the fact that 70% of college-educated Jamaicans no longer lived on the island, a phenomenon known as the Jamaican brain drain. Figueroa hoped her research would shed light on how modern Jamaica could harness these emigrants' intellectual and economic capital. For Figueroa this research project was especially gratifying. Her research directly spoke to her as a member of the Diaspora contemplating how to contribute to her homeland in the future. After the lengthy process of compiling a subject population, with the assistance of the Jamaican Consul General in New York, the Jamaican Ambassador to the United Nations, and family friends, Figueroa had the chance to conduct numerous interviews with Jamaicans educated in North America. Though the interviews were largely held in Jamaica, Figueroa also traveled to San Francisco, New York, and Philadelphia to interview prominent members of these cities' Jamaican populations. Figueroa found that interview-

These craft products are representative of the goods circulating in the informal economy of the Argentine altiplano, and are a point of contention between the NGO and the artisans.

ees frequently related to her and became sentimental. She recalls an interviewee who broke into tears while recounting his experience as a member of the Diaspora. Having conducted such emotional and informative interviews, Figueroa has found the entire research process very meaningful. Like Schuster, she intends to incorporate the findings gleaned from her extensive interviews into her honors thesis.

The unique experiences of Schuster and Figueroa serve as testaments to the meaningful and diverse nature of the overseas research opportunities offered by Stanford. Although the topics and locations of their research differed, their experiences highlight many common advantages and benefits that research abroad can offer. Foremost, overseas research offers an unparalleled degree of proximity to research contacts, making for better interviews compared to email or telephone correspondence. Schuster stated, "although I probably would have been able to establish some contact via telephone, to be able to sit down with an Argentine grandmother and discuss her llama herd was simply incomparable to collecting information from Stanford."

Drawing on her experiences in Jamaica, Figueroa concurred, saying that "doing interviews in person and doing research on location is an inspiring and stimulating process which has the power to shape one's research in a way that other contact cannot. Research on location gives the researcher a sense of purpose and of context." Being on-site to conduct their research enabled both researchers to access firsthand information and experiences that inspired them on academic and personal levels.

However, just as overseas research offers unique benefits to the researcher, it also presents a host of obstacles that typical on-campus or domestic research would almost certainly not entail. Most obviously, a language barrier might prove a formidable obstacle, preventing the forging of contacts and severely slowing the interview process.

Proficient in Spanish, and having just spent a quarter in Buenos Aires, Schuster reflected, "though I typically managed to understand 80% of what was being said in an interview, different dialects and colloquialisms made complete comprehension nearly impossible. There was still a valuable 20% that I was missing out on." Additionally, anti-American sentiment and preconceived notions of American culture often made setting up appointments and arranging interviews difficult.

Arriving in Argentina a mere four days after the announcement of the war in Iraq, Schuster observed ambivalence in the Argentines, "Although the groups I was working with wanted to have the Stanford name attached to their cause, they were deeply suspicious of the West and of a white person poking around for information."

Though part of the allure of overseas research lies in its physical detachment from Stanford's campus, the importance of having a good advisor to stay in touch with while abroad cannot be overstated, according to both Figueroa and Schuster. Figueroa's advisor made a concerted "effort to maintain email contact, recommending books and magazines which she felt would be pertinent to the research being done." In slight contrast, Schuster's contact with her advisor while overseas was not nearly as frequent, mainly because the closest internet connection was three hours away in neighboring Bolivia. Schuster described her advisor as "hands-off in a positive, supportive way—steering [her] in the right direction whenever possible." Having an advisor who takes a genuine interest in the research being done provides not only a reassuring link to Stanford when abroad, but also a sense of security which, said Figueroa, "allowed me to become immersed in what I was doing without worrying."

While these overseas research projects have already proven to be highlights of their researchers' respective Stanford careers, it is evident that their experiences will also figure prominently in their futures. Schuster had long considered development studies an area of interest, and her research experience in Argentina solidified her desire to pursue this topic in her academics and her future career. Figueroa described her research project in Jamaica as having provided her with "interesting insight through its reflexive nature," insight that has reaffirmed her will to pursue a Master's in International Relations in Jamaica and to ultimately work there.

Though a great deal of discipline, organization, and personal initiative is required to plan an overseas research project, and though there are many obstacles, the experiences of both Schuster and Figueroa leave no doubt that conducting research in another country is a phenomenal experience that Stanford students can benefit a great deal from. Both researchers speak glowingly of their experiences and encourage their fellow students to take advantage of this program. In her final reflections on her overseas experience, Figueroa spoke to the amazing power of research abroad by commenting that a student has "only four years at Stanford, and so much of that time is spent studying. To be able to construct a project someplace other than Stanford truly can broaden horizons, spark interests, and create experiences that really will last a lifetime."

Image and Ethics in Biomedical Enhancement

by Ben Howard

"But I don't want comfort! I want God, I want poetry, I want real danger, I want freedom, I want goodness. I want sin."
"In fact," said the Controller, "you're claiming the right to be unhappy."
"All right then," said the Savage defiantly, "I'm claiming the right to be unhappy."
-Brave New World (Huxley, 230-240)

Perched at the pinnacle of technological advancement and biomedical prowess, Western medicine finds itself reaching further and further beyond the traditional role of healing and into new realms of human enhancement. Closely associated with recently developed technologies for disease treatment are unheralded powers in the improvement of otherwise "healthy" human beings. From the already widespread use and promotion of cosmetic surgery to the selective potential of pre-implantation genetic diagnosis, medicine is faced with the power to enhance human life at its most fundamental biological level. Pharmacological substances, so effective in the treatment of destructive neurological disorders, have the potential to pacify human behavior, sharpen mental acuity, and reverse the processes of aging. As healing gives way to enhancing, the very definitions of health and disease become problematic, as do the questions of medicine's goal. With no clear philosophical understanding of the human being to ground a guiding ethics, our science leads us on into an age in which the very biological image of the *Homo sapiens* is open to increasing manipulation.

To be sure, the drive toward enhancement should not be viewed as some terrible modern aberration brought about by technologically driven capitalist society; the human practice of self-transformation transcends boundaries of class, culture, and history. From the cosmetic skin dyes of ancient Egyptian royalty to the surgical (and potentially genetic) alterations of today's image-driven elite, human beings seem universally characterized by a predilection for enhancing their outward appearance. Furless and inventive, we could even be described as evolutionarily adapted to such freedom of expression. Yet despite the wide range and variability of culture and ritual across human groups, the practice of enhancement in nearly every case is inherently guided by an ideal, an image of the flourishing, beautiful, "healthy" human. For many cultures, that ideal integrates biology and being, infusing meaning into the lived experience of the human organism.

So what image of health serves as *our* guide? This fundamental question grounds any ethical undertaking of the problems of biomedical enhancement, and of medicine as a whole. Part of our image comes from the practice of modern biology, with theoretical roots tracing back to Descartes' separation of body from soul and extending through to the evolutionary psychology of today, in which the organism is best understood in terms of reduced and impassionate mechanism, a complex aggregate of biological parts to be explained in functional terms. While this image is key to our medical approach to understanding the human organism, our concept of health draws on broader standards than pure mechanism. Whole specialties of medicine are devoted to correcting problems that are not inherently physiological malfunctions; the cosmetic correction of cleft palates illustrates this larger social conception of health. Thus, it seems difficult to ascertain what ideal, what image, guides such a conception. Is it one of human freedom, of an autonomous agent acting in a pluralistic democracy, or is it, as many cosmetic surgeons would argue, one of human "happiness?" Huxley's *Brave New World* reminds us that our humanness goes much deeper than mere happiness. It seems that by raising these fundamental human questions, issues of enhancement become problematic; to differentiate between healing and what is beyond healing, we must reexamine our foundational understanding of health and the human being.

Any such reexamination should not only consider the facts of evolved biological human nature but also integrate a rational interpretation of the lived human experience. As the philosopher Hans Jonas articulates, the uniqueness of animal (and human) life lies in its *mediacy* (99-107). Unlike plants, we are not in constant and immediate contact with our source of nourishment. Instead, we are distinctly separated from our environmental context, directed outward toward our objects of desire, separated from them both spatially and temporally. We know the sharp distinction between subject and object and the sharper pain of unfulfilled desire; we know suffering. And while mediate distension thrusts us out into the world, it simultaneously increases our relational dependence and emphasizes our inescapable mortality, a concept fully grasped by human sensibility.

In expanding the biological account of the human being to include a phenomenological and existential hermeneutic, the beginnings of a more integrated image of the human be-

ing can be approached. As Jonas' teacher Martin Heidegger shows, we find ourselves characterized by a constant immersion and encounter in which we carry out our synthetical interpretation. Seeing, interpreting, naming, understanding, and giving meaning to the world in which we live, we mediate our existence into an ordered and coherent structure. Such gathering, or *logos*, implies that our nature is to be thrust into the possibility of that world, continually living ahead of ourselves in our meaning-giving planning and projection. By such "thrownness" we are opened up to our inherent possibility, and in such openness, such mediacy, we construct our significance.

Perhaps the most striking aspect of such mediacy is that we do not experience it alone. From the moment of conception to our last dying breath, we are immersed in a web of relational interdependence. One cannot truly understand a person without realizing where they come from, how they relate to their family, who their friends are, who their enemies are, whom they love, how they love, whom they make love to. Such mediate relationality resonates with a biological interpretation of human life; from upright posture and vision-guided movement to language capacity and intentional symbolic brains, our biological anatomy itself forms a delicately balanced embodiment that directs us into relation. Biology and evolution have resulted in a fully relational species, and an ethical understanding of our species should take this into account.

What might an ethics guided by such relational mediacy look like? One could argue that it would take upon itself the task of discerning the way of life in which that deeply relational mediacy could flourish. Such an ethics could strive for "health" in terms of fulfilled relation, in which the dignity and integrity of each individual is understood in the context of that person's multifaceted relational being—not in terms of satisfying an immediate desire for individual "happiness." And given the inherent embodiment of such relationality, the essentially biological and relational practice of medicine would be the archetypal field in which to carry out and develop such an ethics.

Returning to the original and pressing question of biomedical enhancement, we can begin to construct a provisional framework to guide our practices. With the biological, rational, and experiential image of the human being as mediate and relational, we would see enhancement techniques as affecting human beings in the most profound manner. Practices that hold true to the integrated and mediate image of human being as a relational whole, and that perhaps even encourage further depth of meaningful interaction, could resonate with the grounding of such a provisional ethics and with the age-old human drive toward health and enhancement. Conversely, procedures and technologies that go beyond such an image of health to radically alter human performance, behavior, and biology, could arouse suspicion and warrant greater caution. By breaking down our relational and mediate dynamics, they might threaten our most profound sources of meaning, thereby endangering the foundation of our very humanness.

To be sure, the understanding of health and the human being offered here is but a provisional one. The fundamental questions raised require considerable attention and rigor, and they approach their fullest potential when addressed by multiple disciplines. In developing any compelling image of the human being with which to guide our biomedical practice, we must consider scientific, sociological, psychological, philosophical, and religious perspectives. To better approach the complex issues of health and biomedical enhancement, the difficult and profound ethical questions of human image and human meaning must be reopened in a new light.

Ben Howard is a Religious Studies senior with a minor in Biological Sciences.

The author thanks Dr. William Hurlbut (Human Biology) and Professor Thomas Sheehan (Religious Studies) for challenging and encouraging the ideas expressed herein.

References

Huxley, Aldous, Brave New World. Perennial Classics, 1998.

Jonas, Hans. To Move and Feel: On the Animal Soul. In: The Phenomenon of Life. Evanston, Illinois: Northwestern University Press, 2001: 99-107.

BIO-X
STANFORD UNIVERSITY

Bio-X: A Celebration of New Possibilities

by Chris Couvelier, Albert Keung, Jordan Miller

In the spring of 1998, a team of Stanford faculty decided to create a program that would facilitate interdisciplinary study in the biosciences. A faculty planning committee was assembled to further develop and plan the program, which has since become known as Bio-X, one of the most radical and ambitious scientific research programs in recent history. At the heart of the Bio-X program is a strong belief in the power of interdisciplinary research. Though the program itself explicitly strives to pursue biosciences research, the way in which the program is structured welcomes and facilitates the "intermingling" between biology and the fields of chemistry, physics, engineering, and computer science, along with law, business, and ethics, as they apply to the life sciences.

Professor Matthew Scott (right), chairman of Bio-X's Scientific Leadership Council, described the interest surrounding the program, saying, "Biology has matured to the point where [other areas] of expertise are highly applicable and valuable in unraveling biological problems. At the same time the opportunities to engineer things that might be useful have increased with the understanding of how things in biology work. The people in these other fields coming to biology are excited about what they see. The biologists are excited about having them come to the field."

More than 250 Stanford faculty from sixty departments have chosen to affiliate themselves with Bio-X, participating in events, sharing core facilities, and competing for seed money grants awarded through the Bio-X Interdisciplinary Initiatives Program. Scott believes that Bio-X life sciences breakthroughs will not only lead to a greater understanding of science and technology but will also have practical and beneficial social impacts. He sees collaborations in science and technology leading to numerous improvements like non-invasive diagnoses, less damaging surgery, artificial tissue engineering to restore sight, and improved child health care.

From a social perspective, Bio-X provides clear benefits in many areas of life, and it also provides an invaluable resource and academic environment for students on campus. Stanford students are poised to reap the benefits of the cutting-edge research and educational programming to which Bio-X is already giving rise. A graduate-level program in the new Department of Bioengineering has just been started, with plans for an undergraduate program in a few years. Even though there is no official undergraduate program yet, many Stanford undergraduates are already involved with

Bio-X.

Scott, a proponent of early exposure to research in students' undergraduate years, discussed undergraduate work in Bio-X: "There are many undergraduates working in labs affiliated with Bio-X, not necessarily because Bio-X put them there, but because there's a lot of excitement about these kinds of science going on in both this building [the Clark Center] and other Bio-X buildings…The idea of a young person working with someone who is interested in seeing that young person flourish, that's a good thing. The opportunities for learning and even publishing are enormous if they take it seriously and spend a lot of time at it."

Some of the Stanford students working in Bio-X laboratories are located in the Bio-X program's symbolic core and showpiece, the James H. Clark Center (left). The structure boasts a myriad of state-of-the-art laboratories, two supercomputers, many core facilities, an auditorium, and seminar rooms. The building currently provides functioning workspace for approximately forty faculty members from twenty-five departments.

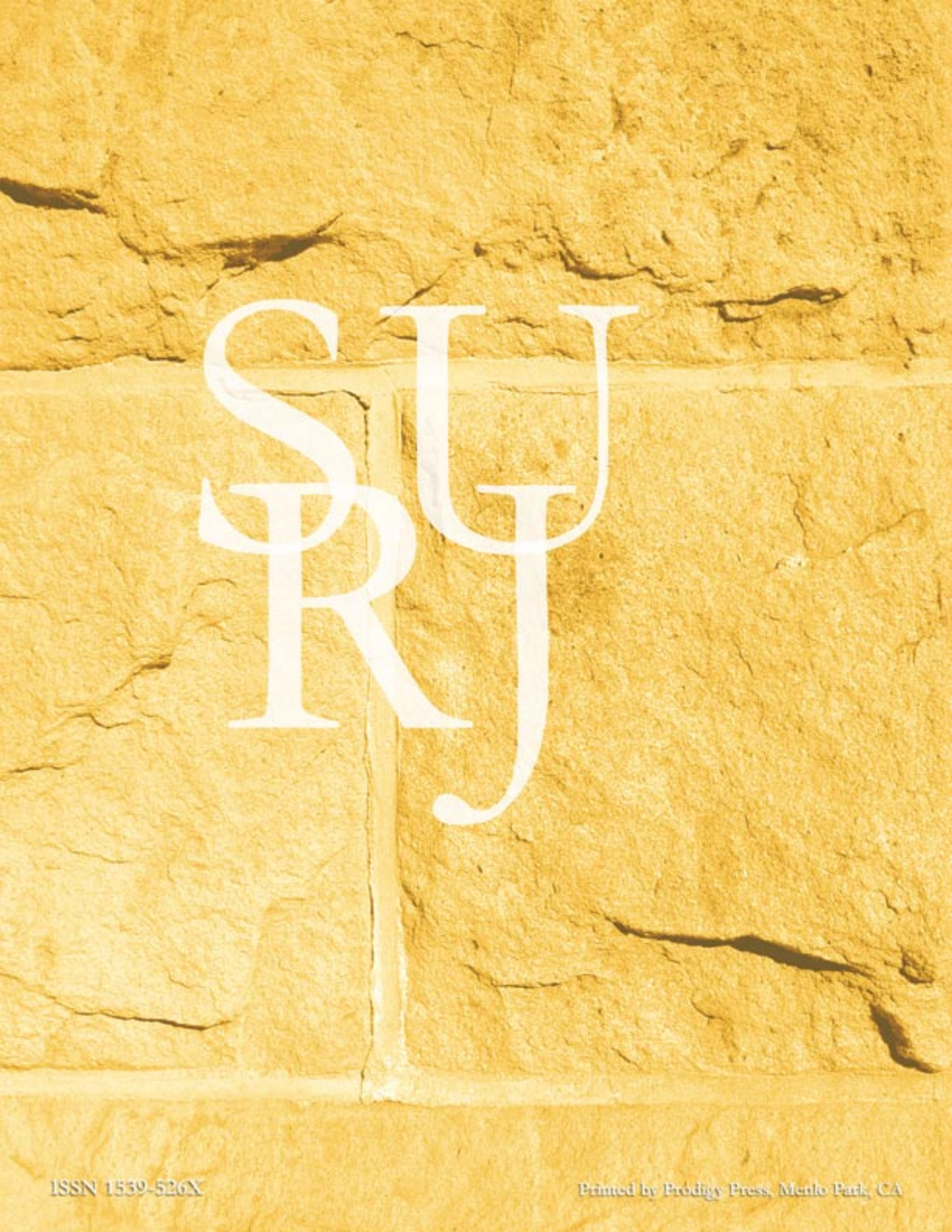
The rapid progress in the successful development and implementation of the Bio-X program has been remarkable. Even more exciting is the fact that the program is still only in its nascent stages. The tremendous sphere of possibilities for Bio-X spans improvements in healthcare, environmental conservation, a deeper understanding of science, and a novel approach to thinking and education. Bio-X provides vast room for growth and invaluable opportunities for undergraduates to explore their interests in an interdisciplinary setting. The rising generation of scientists and engineers will undoubtedly learn and benefit from Bio-X and its celebration of new possibilities.

We Would Like To Thank The Following Individuals And Departments For Generously Supporting The Stanford Undergraduate Research Journal This Year:

Associated Students of Stanford University
Creative Commons
Office of Student Activities
Office of Undergraduate Admissions
School of Engineering
The Writing Center
Undergraduate Research Programs

Gene Awakuni, Vice Provost for Student Affairs
Elizabeth Bernhardt, Professor of German Studies
Kristin Black, Lecturer in the Biological Sciences
* John Bravman, Vice Provost for Undergraduate Education
Susie Brubaker-Cole, Director of Undergraduate Research Programs
Nigel Bruce, Stanford Alumni Association
Eve Clark, Professor of Linguistics
Karen Cook, Cognizant Dean for the Social Sciences
* John Hennessy, President of the University
Michael Keller, University Librarian
Lawrence Lessig, Professor of Law
Andrea Lunsford, Professor of English and Director of Program in Writing and Rhetoric
Hilton Obenzinger, Associate Director of Undergraduate Research Programs
Susan Olzak, Professor of Sociology
Douglas Osheroff, Professor of Physics
David Palumbo-Liu, Professor of Comparative Literature
Jim Plummer, Dean for the School of Engineering
Jared Scherer, Assistant Director of the Office of Student Activities
Laura Selznick, Special Asst. to the Vice Provost for Undergraduate Education for Diversity Outreach

* Provided key budgetary support



ISSN 1539-526X

Printed by Prodigy Press, Menlo Park, CA