Volume 20 Fall 2010

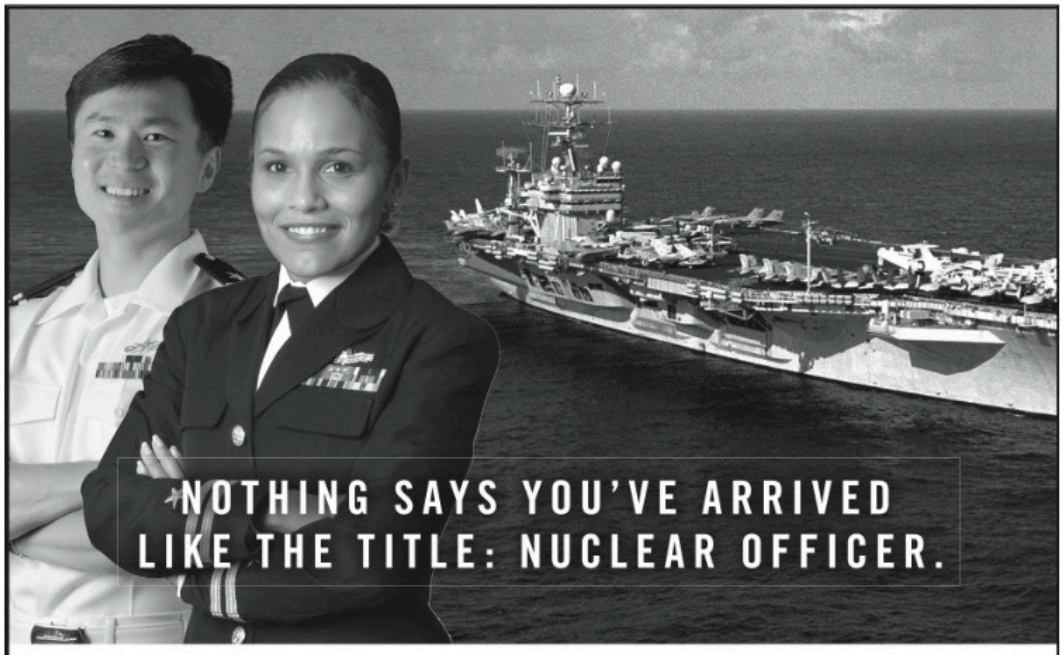
MURJ

Massachusetts Institute of Technology
Undergraduate Research Journal

Detecting the Unfolded Protein Response in Drosophila

p. 66

Massachusetts Institute of Technology



Put yourself in truly elite company. From day one. In the Navy Nuclear Propulsion Officer Candidate (NUPOC) program. **Get up to $168,300* in financial support as a student.** Enjoy an impressive salary. Extraordinary benefits. As well as world-class technical training. Command a nuclear-powered aircraft carrier. A stealth submarine. And teams of Sailors. Ready to gain experience beyond your years? Learn more.

WANT TO LEARN MORE? CONTACT YOUR NAVY RECRUITER TODAY.
800.792.9099 | M_MILL_CNRC_LPT_NEWENGLAND@navy.mil

*Depending on location.

4D2DADGOF11

AMERICA'S NAVY

A GLOBAL FORCE FOR GOOD.™

MURJ

Massachusetts Institute of Technology
Undergraduate Research Journal

Contents

UROP Summaries

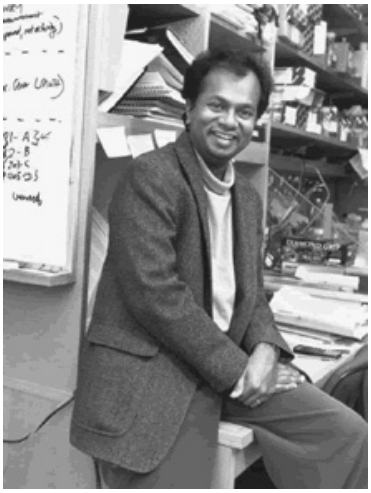
Web Abstracts

Reports

Ram Sasisekharan
Departments of Health Sciences & Technology and Biological Engineering
Director and Edward Hood Taplin Professor of HST

Massachusetts Institute of Technology
77 Massachusetts Avenue, Building E25-519F
Cambridge, Massachusetts 02139–4307
rams@mit.edu

December 2010



Dear MIT Community,

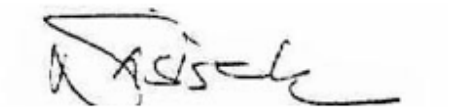
It is an honor and a pleasure to introduce this issue of MURJ. Although my own undergraduate career was not spent here at MIT I have had the satisfaction and privilege of engaging with the bright minds that populate our student body as a Post-doctoral Fellow, Professor and now Director of Health Sciences and Technology. I need look no further than the wonderful Undergraduate Research Opportunities Program (UROP) scholars that energetically apply themselves to varied projects at the very cutting edge of innovative research and invention to be powerfully reminded of the unique value of engagement in project work by our undergraduates.

From the discovery of a comet by an EAPS UROP in 1987 and numerous UROP alums who have gone onto astronaut careers with NASA to the engagement of several UROPs on the Peabody Award-winning Design Squad series, undergraduate researchers have sustained and increased their visibility in an expanding number of roles and arenas. Today MIT's research takes place around the world and at many levels and undergraduates are at the forefront of many such initiatives, participating in programs such as the Singapore-MIT Undergraduate Research Fellows (SMURF) Program, and interacting with a diverse array of students, faculty and communities in the greater Boston area and beyond.

The cross-disciplinary nature of the work undergraduates engage in as UROPs is paralleled countless times over each day here on campus by students from a wide range of theoretical and experimental backgrounds who apply themselves with ingenuity and zeal to work that is not readily pigeonholed into one department or focus. I have marveled at the increased reach of MIT into the global arena with each passing year and this is particularly notable amongst our youngest community members. MIT's undergraduates do not think inside the box, nor do they satisfy themselves to foster collaborations and inventions solely at a local level. This group defines not only what the Institute is today but a bright vision of what it can be—and do—tomorrow.

The distinction surrounding the creative and determined spirit that MIT students embody grows with each class as they develop their own style and method of spreading the impact and benefits of the work that they engage in on campus. Our undergraduates are prepared to tackle the big challenges that preoccupy us today including global health concerns and sustainability and I for one cannot wait to see their innovations emerge as they do so.

Sincerely,



Ram Sasisekharan
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering
David H. Koch Institute for Integrative Cancer Research

Massachusetts Institute of Technology

UNDERGRADUATE RESEARCH JOURNAL
Volume 20, Fall 2010

Printed at Universal Millenium

PSB 07-02-0097

MURJ Staff
MIT Undergraduate Research Journal

Massachusetts Institute of Technology
murj.mit.edu

December 2010,

Dear MIT Community,

We are honored to present the 20th issue of the MIT Undergraduate Research Journal (MURJ), a biannual publication featuring the stunning undergraduate research work at the Institute. For 10 years, we have been encouraging undergraduate research and providing exposure to the unique scientific endeavors at MIT. Our newest issue represents our commitment toward the important role of undergraduate research in a student's education, and we are amazed by the quality and brilliance of the reports showcased here. Spanning an impressive range of fields, the research presented here involves such fascinating topics as mapping the unfolded protein response in *Drosophila melanogaster* and a statistical analysis of the fairness of political elections. Proud of the exceptional and inspired research at the Institute, we are especially excited to share these articles with the MIT community.

In addition to presenting undergraduate research, our staff is also dedicated to highlighting the science of the overall community. As this semester marks the opening of the new David H. Koch Institute for Integrative Cancer Research at MIT, we sat down with its director Dr. Tyler Jacks to discuss MIT's dedication to fighting cancer and bring readers a behind-the-scenes tour of the facility. We also feature the gravitational wave research of Professor Nergis Mavalvala, a 2010 MacArthur Fellow and an example of the brilliance at the Institute. We hope these feature articles will inspire you and provide a glimpse of the fascinating work within the community.

Our hope is that this journal is a representation of the intellectuality and spirit that the undergraduates fill the Institute with. This issue has been made possible because of the hard work of many students, departments, and faculty at MIT. We would like to thank all our executive members and associate editors for their tireless work in reviewing submissions and ensuring the utmost quality of work. Moreover, our staffers in the News and Features departments have shined with their immense creativity and extraordinary effort, without which this incredible issue would not be possible. A special thanks goes out to the Office of Undergraduate Advising and Academic Programming, MIT Student Activities Office, and the MIT Publishing Service Bureau for their continued support. And lastly, a special thanks to the undergraduate authors for sharing their research with us and the larger community. We hope you enjoy the issue!

Sincerely,

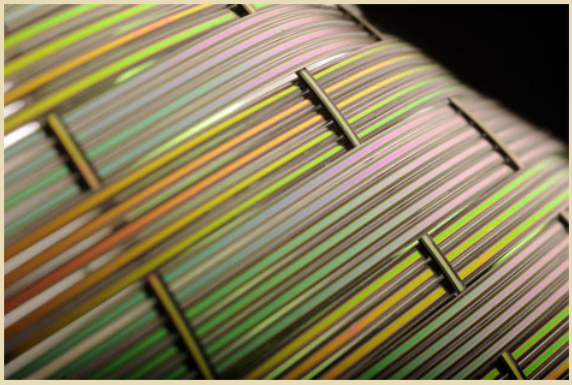
Omar Abudayyeh
Co-Editor-in-Chief

Ana Lyons
Co-Editor-in-Chief

Evelyn Wang
Co-Editor-in-Chief

Science News In Review

Vibrating Fibers with Extraordinary Properties



Fibers that can pick up and emit sounds or transmit acoustic energy into electrical energy are in development at Professor Yoel Fink's lab at MIT RLE..
Credit: Courtesy of Professor Yoel Fin

In the very near future, the fibers in an evening gown could have the capacity to produce—and even record—the same melodious vibrations of violin strings. Professor Yoel Fink at MIT's Research Lab of Electronics has been working to develop extremely sensitive fibers, capable of being woven into fabrics that can be used to perform such feats. One of several types of fibers produced in Fink's lab, this particular variety can act as a microphone to pick up acoustic signals, and as a speaker to emit them. The fibers, with strands slightly thicker than hair, contain piezoelectric material and can translate acoustic energy into electrical energy and vice versa with the aid of electric pulses. Cloth made from these fibers would be able to function as a personal music player or a device to monitor blood and oxygen flow in the body. These fibers have already been woven into fabrics to test their uses in practical settings. As this research proceeds, we may soon see its effects in the form of a new brand of clothing on the market—a type of clothing that can revolutionize our lives.

—P. Thaker

Developments in Cancer Research and Tumor Growth

A study published in the British Journal of Cancer in late August has found a possible connection between AMP-activated protein kinase (AMPK) and the pathway MAPK3/1 in colorectal cancer. AMPK plays an important role in the energy balance of a cell, as it essentially acts as a sensor, and MAPK3/1 plays an important role in the cell's differentiation, proliferation and survival. In the past, it has been found that AMPK can kill cancer cells and suppress tumor growth, and that the suppression of AMPK itself can lead to cancer. However, our ability to form a prognosis based on AMPK levels has never been realized. It is therefore important to study this protein kinase in more depth. This study did just this by investigating more carefully AMPK's specific role in human colorectal cancer. The scientists looked for evidence of phosphorylated AMPK (p-AMPK) and p-MAPK3/1 in 718 cases of this kind of cancer, and found that expression was related to survival in p-MAPK3/1-positive cases, but not in p-MAPK3/1-negative cases. This suggests that there is some kind of link between AMPK and MAPK concerning how tumors grow and behave. The results of this study have shed some light on the mystery that is tumor growth and may lead to more effective treatments in colorectal cancer.

—A. Ho

The Importance of Water: A Structural Perspective

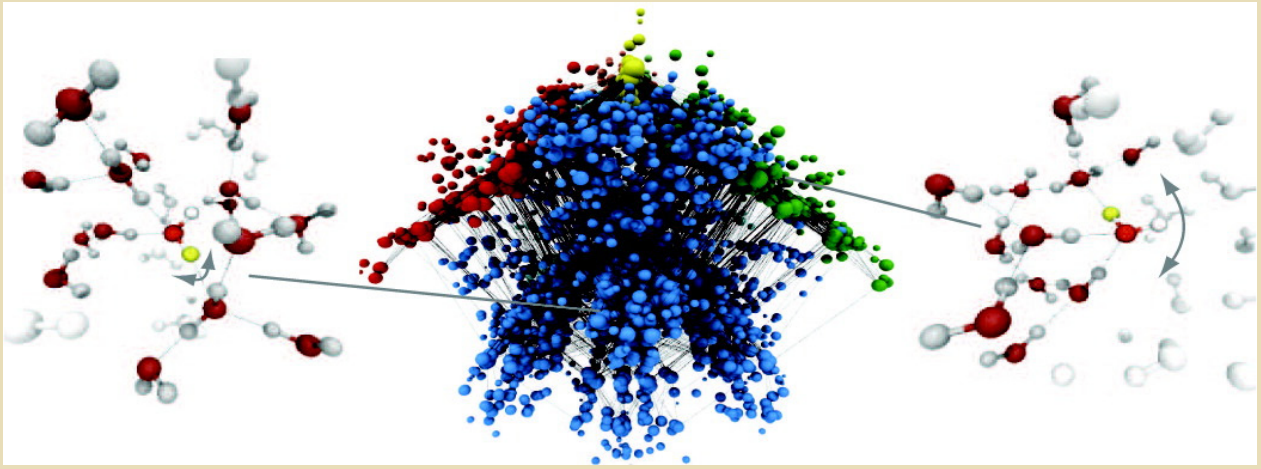
Water is the foundation of life on Earth and recent research by Francesco Rao, Sean Garrett-Roe, and Peter Hamm of the University of Freiburg, the University of Strasbourg, and University of Zurich shows that we are just beginning to understand the complexity of water's structure.

For decades, scientists have been trying to figure out why water, and not any other molecule, is the basis of life. This new study suggests several interesting possibilities.

The researchers describe the structure of water as having different sub-states, where a sub-state is defined as having kinetically homogeneous inter-compound interactions. Using computer models, the researchers were able to show that water has two main substates. The first of these is seen in bulk water, where a network of hydrogen bonding is created around a central water molecule. The water molecules exchange the molecules that they are bound to and, in fact, very rapid switching between this and the second state of water occurs in a fraction of a second. This leads to the dynamic nature of bulk water. In the second substate, the hydrogen atoms can also interact with organic molecules on the periphery of this chaotic and dynamic process, as water molecules constantly break apart and come together. This previously elusive property of water will allow researchers to understand water's influence on proteins and other organic molecules in complex biological systems.

—E. Atolia

Source: ""At the Smallest Scale, Water Is a Sloppy Liquid" -Science Magazine and "Structural Inhomogeneity of Water by Complex Network Analysis" -ACS Publication



Water's two main substates may underlie its essentiality for all living things.
Credit: American Chemical Society Publication

Enhancing Reality with the Flick of a Finger

Imagine picking the latest issue of the New York Times from your local newsstand, flipping to weather, and having updated forecasts projected onto the weather map in real time, right before your eyes.

Imagine walking past a breathtaking view while on vacation, framing the scene like a shot in a movie with your fingers, and having a digital picture captured automatically—a picture that you can later take home to project on a slide show, edit, and organize.

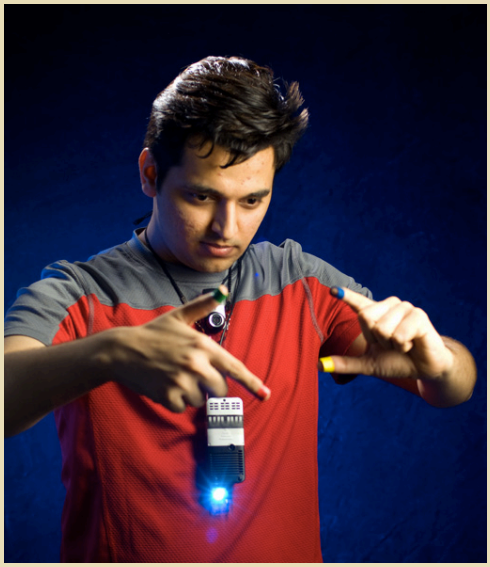
Or yet, imagine being able to play a video game, surf the internet, or make a phone call, all on a single sheet of white paper; no need to pull out your laptop or fumble with your cell phone.

Now imagine being able to do all this with a single homemade device that costs less than $350.

Thanks to the work of graduate student Pranav Mistry at the Fluid Interfaces Group at the MIT Media Lab, these science-fiction-like tasks have become possible with the use of his newest prototype, dubbed the 'SixthSense.'

According to Mistry's website, the SixthSense can most simply be described as "a wearable gestural interface that augments the physical world around us with digital information." The device "lets us use natural hand gestures to interact with that information," the site says.

The SixthSense is composed of simple elements like a webcam, a 3M pico projector, and a mirror that can be worn around the neck, and interacts wirelessly with a Bluetooth-enabled smartphone. The camera can recognize images or gestures that the wearer makes in the air, allowing the device to project relevant information onto a nearby surface, as in the case of a newspaper's weather map, or to complete certain tasks, as in the case of taking a photograph.



'SixthSense,' technology designed by the Fluid Interfaces group at MIT Media Lab, allows a user to transform the physical world into an interactive platform that can be navigated using natural hand gestures.
Credit: Sam Ogden

Recently, the SixthSense has gained numerous media attention from such sources as The New York Times and garnered awards from Popular Science and Technology Review.

In a TED (Technology, Entertainment and Design) talk given in 2009 and posted online, Mistry spoke of the importance of creating such a device: "I think integrating information to our everyday objects will not only help us to get rid of the digital divide—the gap between these two worlds—but will also help us, in some way, to stay human...to be more connected to our physical world."

"It will help us, actually, to be machines sitting in front of another machine, " he said.

More information on Mistry's latest version of the prototype can be found online at: http://www.pranavmistry.com/projects/sixthsense.

—A. Lyons

Redesigning the Airplane

Popular Mechanics magazine has awarded the 2010 Breakthrough Inventor Award to the MIT-led team of Mark Drela (MIT), Edward Greitzer (MIT), Jeremy Hollman (Aurora Flight Sciences), and Wesley Lord (Pratt & Whitney) for their dramatically efficient aircraft redesign. The fundamental design of the commercial airliner has not changed much over the last several decades, and while technology has improved aircraft efficiency by leaps and bounds, the shape of the airplane has remained generally static. The team proposed a new aircraft design that marks a significant improvement over the traditional design.

The team's design fuses two cylinders into a wider and shorter fuselage. This novel shape generates lift. In addition, long wings and a joint tail reduce drag. Other design features to improve efficiency include three fuselage-mounted engines, which consume less fuel, and the use of lightweight modern composite materials. The redesign was in response to a NASA research program to design the aircraft of 2035, roughly the date at which the volume of air travel is expected to be double that of today. The new aircraft the team designed may allow for 70% less fuel consumption than current planes, while simultaneously reducing noise pollution. Models of the aircraft are currently undergoing wind tunnel testing in MIT's Department of Aeronautics and Astronautics' Wright Brothers Wind Tunnel.

—A. Sinha

The 2010 Breakthrough Inventor Award awarded by Popular Mechanics magazine was given to an MIT team that proposed a radically novel and dynamic airplane design that more efficiently generates lift and reduces drag.
Credit: Courtesy of Popular Mechanics

Source: http://web.mit.edu/newsoffice/2010/pop-mech-award.html

http://www.popularmechanics.com/technology/aviation/news/radically-redefining-the-airplane-breakthrough

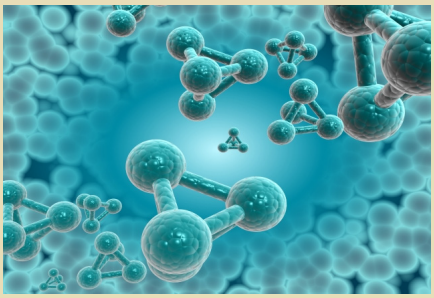
2010 Nobel Prize in Chemistry: Palladium-Catalyzed Carbon-Carbon Single Bonds

The 2010 Nobel Prize in Chemistry was awarded to Professor Richard F. Heck of the University of Delaware, Professor Ei-ichi Negishi of Purdue University, and Professor Akira Suzuki of Hokkaido University for palladium-catalyzed cross couplings in organic synthesis.

Carbon-carbon bonds are essential for the formation of complex biomolecules like sugars and fats – basic compounds that are key to life. Although large carbon molecules are stable, it can be difficult to synthesize them in the laboratory. Developments in the 20th century led to the use of transition metals as catalysts to synthesize organic molecules. In 2005, the Nobel Prize in Chemistry was awarded for the formation of metal-catalyzed carbon-carbon double bonds. Heck, Negishi, and Suzuki took this to the next step: the catalysis of carbon-carbon single bonds. The mechanism behind palladium-catalyzation is that the metal (palladium) forms metal-carbon bonds, holding molecules close together, allowing for the formation of carbon-carbon single bonds.

The palladium-catalyzing mechanism has been effectively used on an industrial scale, appearing in familiar household items such as Naproxen, an anti-inflammatory drug, and Singulair, a drug to treat asthma. The Laureates' discovery holds great potential in academic research and in industrial chemical production, having already deepened our understanding of organic chemical synthesis and changed the industrial chemical production of biologically active compounds.

—A. Sinha



This year's Nobel Prize in Chemistry was awarded for the discovery of a novel method to synthesize carbon-carbon single bonds.
Credit: FreeDigitalPhotos.net (jscreationzs)

Source:http://www.sciencedaily.com/releases/2010/10/101006085720.htm

http://static.nobelprize.org/nobel_prizes/chemistry/laureates/2010/Sciback_2010.pdf

MIT Dean Subra Suresh Named Director of the National Science Foundation

Subra Suresh ScD '81, the former Dean of Engineering and Vannevar Bush Professor of Engineering, was officially sworn in as the Director of the National Science Foundation this past October 18.

Suresh was sworn in by President Obama's Science Advisor and director of the Office of Science and Technology Policy, John Joldren '65, SM '66. He was officially nominated to the position on June 8 by President Barack Obama and was confirmed by the U.S. Senate on September 29.

As the new director of the National Science Foundation (NSF), Suresh will oversee the $7.4 billion federal agency that supports basic science and engineering research and education as part of his six-year term.

NSF funds more than half of all non-medical science and engineering research at American academic institutions, like MIT. Staff members review research proposals, which are awarded through a merit-based process.

Unlike many previous NSF directors who were no longer active scientists when they took over the position, Suresh has maintained an active lab during his time at MIT. His research spans a diverse array of engineering and bioengineering-orientated topics, ranging from the study of the mechanics of amorphous materials to the nanoscale properities of biomaterials as well as the mechanobiology of malaria.

Subra Suresh is the Director of National Science Foundation. Suresh holds appointments in Materials Science and Engineering, Biological Engineering, and the Health Sciences and Technology program. He served as Dean of Engineering since July 2007.
Credit: Justin Knight, MIT News

According to the MIT News Office, Suresh has authored over 220 research articles in international journals, co-invented over 12 U.S. and international patents, and has co-edited five books, including the prominent textbook Fatigue of Materials and Thin Film Materials throughout his research career.

Cynthia Barnhart, Ford Professor of Engineering, is acting as the interim dean of the School of Engineering while Suresh is serving in his new role.

—A. Lyons

Viability of Non-Food Biofuels

As the demand for crude oil and other fuels increases, it is important to find new sources of energy. The Ion Cyclotron Resonance (ICR) Department at the National High Magnetic Field Laboratory is using non-food materials to generate viable biofuels. There are two main types of sources that are being investigated – algae and other non-food sources such as peanut hull, pine pellets, and pine chips.

Algae store their energy in lipids, which can therefore be extracted and used for the production of biofuels. The ICR lab is using a new method of using polar lipids to characterize algae samples. This study is targeting a particular type of lipid called triacylglycerides (TAGs). TAGs are neutral lipids; however, they are formed by a series of polar lipids. Thus by studying the polar lipids, the lab is determining which algae can yield the fatty acid ester that will give the greatest biofuel efficiency when attached to lipids. The ICR lab has determined that the algae Nannochloropsis Oculata contains a high concentration of oleic acid – which, in addition to linoleic and linolenic acids, is known to have the most desirable characteristics

continued on p. 10

"Viability of Non-food Biofuels" continued from p. 9

for fuels – and is thus a potential candidate for biofuel production.

In addition to using algae as a source of fuel, the ICR team has been investigating the use of peanut hull, pine pellets and pine chips to make biofuels. After these feedstocks undergo pyrolysis, a two-phase (oily and aqueous) product is formed. The aqueous phase is generally not useful and is discarded when the biofuel is being produced. At present, the lab is characterizing these two phases in order to better understand the oily phase and to determine a use for the unused aqueous byproduct.

Dr. Ryan Rodgers, the Director of Environmental and Petrochemical Applications in the ICR lab, indicated that currently biofuels is not economically viable because of high costs of production. However, the lab has found that Nannochloropsis Oculata also produces Eicosapentaenoic Acid (EPA), which has beneficial medical properties, such as "cancer prevention, increased immune function, added brain health, and anti-inflammatory activity." Therefore, it is possible that the biotechnology-related manufacture of the EPA byproduct can assist in making algae-based biofuel production economically viable. Furthermore, finding a use for the unused byproduct in the production of peanut hull, pine pellets and pine chips may also make this process more economically favorable.

—E. Atolia

Biofuels derived from algae and other available crops produce byproducts that have beneficial therapeutic and medical side effects.

Source: Interview- Dr. Ryan Rodgers; "ICR team exploring alternative, non-food biofuels" -Mag Lab Reports; "Preferred Polar Lipids of Marine Eustigmatophyte, Nannochloropsis Oculata: Assessing Potential as Biofuel Feedstock and Eicosapentaenoic Acid Producer"

Possible Link Between Mitochondria and Parkinson's

Parkinson's disease is a neurological disorder that typically develops slowly over time, characterized by an increasing loss of motor control. There is currently neither a universally accepted cause, nor a cure. However, recent studies have suggested that broken mitochondria – cell powerhouses that convert food into fuel – may be the cause of Parkinson's. As a result, companies are beginning to turn their attention toward investigating whether repairing mitochondria could be a cure to this terrifying disease.

Neurologist Clemens Scherzer from Harvard Medical School an his team compared gene expression patterns between diseased and healthy brains on a scale over ten times that of what has ever been examined in a single study. Ten gene sets emerged that had never before been associated with Parkinson's disease, and all ten are related to mitochondrial function.

Researches at Harvard Medical School have discovered ten new gene sets the regulate mitochondrial function and that are also associated with Parkinson's disease.
Credit: Wikimedia Commons

In addition, it has been found that mutations in mitochondria-regulating genes are associated with a rare, genetically inherited form of Parkinson's. Another possibility emerging from this study concerns a single protein called PGC-1α. In the data, it emerged that Parkinsonian brain samples contained abnormally low levels of PGC-1α, which suggests (according to Sherzer) that raising levels of PGC-1α could be a way to treat the disease.

If mitochondria is indeed the cause of certain aspects of Parkinson's, the scientific community may well be on its way to providing a cure to the many who suffer from this disease.

—A. Ho

MURJ Features

Interview: *Director Tyler Jacks Speaks on the Opening of the David H. Koch Institute for Integrative Cancer Research*

By: Ana Lyons

View of the newly completed David H. Koch Insitute for Integrative Cancer Research from the corner of Ames and Main Street.
Credit: Cynthia Chen

"It is both exciting and challenging to envision bringing together biologists and engineers under one roof to work together."

As movers deliver the final crates of lab supplies from across campus and researchers settle into their newly acquired 180,000 square feet of lab space, construction of the David H. Koch Institute for Integrative Cancer research is finally coming to a close. The building's sleek glass facade reflects the nearby Whitehead and Broad Institute, and the iconic MIT dome peeks through the cluster of high-powered research buildings at the corner of Ames and Main Street.

The newly erected cancer institute houses 26 faculty from both the School of Engineering and the School of Science—bringing together a unique combination of biologist and engineers to solve some of the most pressing issues in cancer treatment. Among the roughly 120 postdoctoral fellows and associates, 50 principal research scientists, 70 technical assistants, 175 graduate students and 70 undergraduate relocate their research into the new building, five current and former faculty are Nobel laureates and 20 faculty members are members of either the National Academy of Sciences or Engineering.

Construction of the $24.5 million cancer center began in 2007 when MIT received a $100 million from alum David H. Koch '62 to expand cancer research facilities. The new institute will replace the MIT Center for Cancer Research, and it is one of eight National Cancer Institute-designated basic research centers in the United States. The new building will celebrate its official opening ceremony later this winter—although researchers will begin conducting work starting this November.

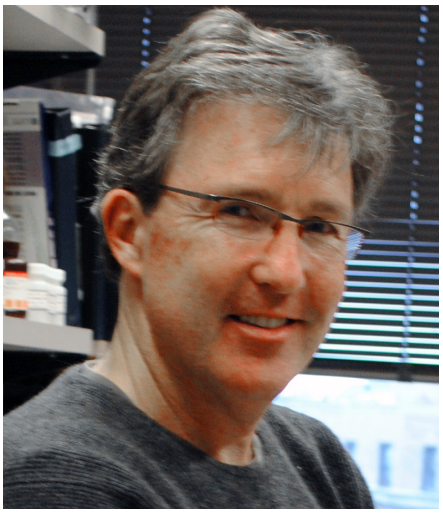
Here, director of the David H. Koch Institute for Integrative Cancer Research Tyler Jacks explains what the opening of the new institute will mean for cancer research at MIT. Jacks is also a Koch Professor of Biology and HHMI investigator at MIT, and his research pioneered new technologies to model genetic events that contribute to the development of cancer in mice. He serves on the Board of Scientific Advisors at the National Cancer Institute and was elected to the National Academy of Science in 2009. He received his PhD from the University of California, San Francisco and was first named an assistant professor at MIT in 1992.

In this interview, Jacks speaks with MURJ about the many benefits that will arise from moving scientists and engineers into one centralized location. He speaks of how the facilities of this new building help enable cancer research at MIT—allowing researchers to tackle problems in cancer detection and treatment that require both a deep understanding of biological processes and the application of cutting-edge engineering techniques.

MURJ: How do you think that having one centralized building will change the progress and culture of cancer research at MIT?

Tyler Jacks: I think that culture is the key word there. It is both exciting and challenging to envision bringing together biologists and engineers under one roof to work together. I say challenging because we are from different backgrounds and, at some level, we speak different languages and approach problems in different ways. But although there are challenges, that's also the excitement. In other words, by approaching problems from different perspectives, we think of new ways of solving them.

And indeed we've seen lots of examples already even before we moved of the advantage of thinking about a problem from both the perspective of a biologist and engineer. When we move in, we will increase the opportunity for those interactions to take place. That is the big advantage of having a shared building—it will enhancing the likelihood that engineering faculty, students, and post doctoral fellows will bump into and begin interacting and collaborating with their counterparts in the biology side and vice versa.



Tyler Jacks, a biology professor and HHMI investigator, is the first director for the David H. Koch Institute for Integrative Cancer.
Credit: Cynthia Chen

We've created a lot of opportunities in the new building to increase the chance encounters of these two groups through the design of the building and some of the building's elements. So we're excited about the potential for such cross-disciplinary research. We've seen examples of how it can be employed already before we've even moved in, and I'm quite confident that once we move in, there will be many more examples.

MURJ: Could you elaborate more on how the design of the building encourages cross-disciplinary interaction?

TJ: The layout of the floors was designed with a sense of interaction in mind. Each floor has both biology labs and engineering labs, as opposed to having them separated by floor or

Lab bench in the Koch Institute.
Credit: Cynthia Chen

in different wings. In addition, we set aside the central portion of each floor—it's a zone in the middle—where we have all the other things that a student or post doc would need outside of their research lab. All the faculty offices are ganged together in the middle. All of the meeting rooms are there as well. The tearooms, where people have their lunches, are in the middle. This is also where the elevators and the bathrooms are. And we actually have a separate "interactive space" with comfortable seating in this same part of the building to encourage people to come and hang out. So this is a way to kind of draw people together and encourage those kind of chance encounters.

What I've just described addresses horizontal interaction along the floor, but we also want to promote vertical interaction. So in the middle of the building, there's a central connecting stairway which we think will encourage people to use the stairs and interact with people on the floor above or below.

A third one is we have a café in the new building on the first floor, which again has the goal of bringing everyone together. A large auditorium: similar idea. Thsse are building-wide elements to foster interaction.

We also set up in this building a fair number of so-called "core facilities" or shared research facilities. We have them now actually, in the current building, but we have a lot more space dedicated to core facilities in the new building—about 22,000 ft., which will house about a dozen of these core facilities. These too promote interactions among the labs and people outside the building, because students and post docs from different labs end up meeting and interacting through their use of these core facilities.

MURJ: What new technologies will the new building provide? Have the current technologies been expanded, and which new technologies will be available to researchers?

TJ: They've all been expanded in terms of size. The expansions were necessary because they were way too small here. If you were going to our flow cytometry, for example, it is totally jammed. So the expansion is in part to deal with that. But also the expansion is also to deal with what we know to be an increasing demand on campus. We expect our core facilities to get a lot of business from outside the building, and we encourage that. We want them to be available for use by the community.

In terms of expansion of new technologies, I would highlight two. One is a facility that hasn't existed before, and it is coming into existence by virtue of the fact that we're bringing engineers in the new building. It's the "nanomaterials characterization facility," which will have all sorts of machinery designed to characterize nanomaterials: size, structure, composition, physical properties. We are purchasing a lot of new equipment for this purpose, including a transmission electron microscope, and an atomic force microscopy, for example. These are things that we have not had before but are needed for that group of investigators to support their research.

A second example would be a greatly expanded facility for in vivo imaging and preclinical therapeutics. This facility will be embedded within the animal facility, which is on the top floor of the new building. Within the animal facility is an imaging suite with multiple types of imaging equipment: Micro CT imaging, bioluminescence imaging, and MRI, which is new and only purchased for the new building. The purpose of all of this is to allow for longitudinal studies in, for example mouse models of cancer, in order to examine tumor development over time and to look at response to new treatments as well. This is a very important aspect to cancer research that we do here. We want to understand not only how tumors develop, but also how to treat them more effectively.

These new imaging facilities will be helpful not only to the biologist, who one might think of first in these kinds of applications, but also the engineers who build materials and new devices for cancer treatments, including nanomaterials for drug delivery and treatment. They're very interested in using sophisticated animal models to test the efficacy of their new materials or methods, and this facility is very helpful in that way.

MURJ: You've been emphasizing the importance of the collaboration between biology and engineering. Could you name or briefly describe a few examples?

TJ: There are several that I could mention, but here I'll highlight just two.

First, you might be familiar with RNAi and the value of RNAi as a therapeutic modality? We use RNAi in the laboratory as you probably know for lots of different types of experiments. We hope that RNAi can be used therapeutically, because if it can be used therapeutically, it opens up a huge new range of targets for cancer treatment. Genes or, more specifically, proteins that are difficult to drug today could be inhibited through the use of RNAi. This opportunity could be addressed if we could get RNAi to work in the body. As a first step to doing this would to be to demonstrate that RNAi could work well in an animal. At its roots, this is a delivery problem: how to get these RNAi molecules to the cells you care about efficiently? And as a delivery problem, it's really an engineering problem. Can you engineer new solutions for RNAi delivery?

And there are quite a few labs in the Koch Institute that are interested in solving this problem, including Bob Langer's lab, Sangeeta Bhatia's lab, Paula Hammond's lab, Dan Anderson's lab—there are a number of them. All of them are approaching the problem slightly differently, and all of them are interacting with their biology colleagues to understand what are the right RNAi molecules to use? What is the right model to test them? How do we ensure that we're distinguishing the target of interest? The therapeutic agent here, the RNAi molecule, is biological in nature, but what the major limitation—the delivery of the RNAi molecules—needs an engineering solution.

"Each floor has both biology labs and engineering labs, as opposed to having them separated by floor or in different wings."

A second example I would give you is in the area of the immunology/cancer interface. We are interested in what ways does the immune system recognize or fail to recognize a developing tumor. Can we understand that sufficiently well to be able to use the immune system to help treat cancer? We have a program in the area of engineering the immune system for improved cancer treatment. This is a project that involves biologists—experts in the immune system like Jianzhu Chen—working with engineers who are thinking of new ways to modify cells in the immune system to improve their survival or activity in the context of cancer treatment.

One of the Koch institute faculty Darrell Irvine, for example, has figured out ways to attach nanoparticles covalently to the plasma membrane of tumor-reactive T-cells. He can load up those nanoparticles with materials that stimulate the survival and activity of the T-cells. So in that way, when they're returned that animal, the T-cells are more active at killing a

Tyler Jacks' New Office.
Credit: Cynthia Chen

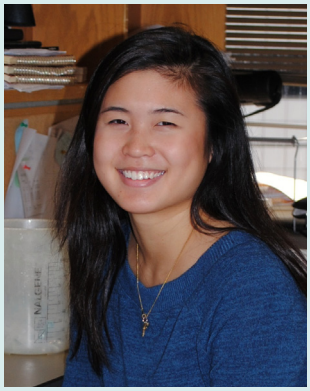
Left: Painters make final touches to the Koch interior the week before the first lab arrives.
Credit: Cynthia Chen

Above: Rows of empty lab benching fill the new Koch Institute prior to the move-in. **Left:** Biologists Tyler Jacks and Nancy Hopkins share a lab office space with their engineering colleague Sangeeta Bhatia.
Credit: Cynthia Chen

Above Left: Long corridors connect biology and engineering labs throughout the building. **Above Right:** The reverse side of the new cancer center faces a green courtyard between Building 68 and the Stata Center. **Below:** Bright yellow staircases connect several floors on either ends of the building-**.**

Credit: Cynthia Chen

UROP Spotlight



MIT junior, Diana Wang, is a current UROP student in the Jacks lab.
Credit: Cynthia Chen

Diana Wang, a junior in Course 7, is one of the many undergraduate researchers who will be moving her lab work to the new cancer institute this November. As a UROP in the Jacks lab, her work focuses on unraveling the specific mechanisms that regulate NKX2-1—a transcription factor found in lungs that appears to be related to cancer. By using a variety of tissue culture experiments, gene knockdowns, and mutations of phosphorolation sites, she hopes to better characterize the pathways which regulate the transcription factor's expression.

"[NKX2-1] is a novel transcription factor, and we don't really know much about it. It seems to be a tumor suppressor, but it actually might act as an oncogene." Wang said. "If we know NKX's specific role in lung tumorigenesis, it's a possible target for drugs."

Like her supervisor and director of the new Koch Institute Tyler Jacks, Wang looks forward to moving to the new facility for the increased synergy between biologists and engineers. "I definitely think collaborations will increase," she said. "Molecular biologist can't solve the issues of cancer by themselves. I'm excited to see what the engineers can bring in collaboration with the biologists."

Wang currently is a pre-medical student, but said she is also considering MD-PhD programs. She has been working in the Jacks lab for over a year, and has been published as co-author in an article appearing in the journal Cell from a previous UROP completed her freshman year.

cancer that expresses that specific antigen. This is a good example of an engineering solution to a long-standing problem in cancer biology, in this case cancer immunology.

MURJ: Are there any key differences between the new cancer initiatives and the old initiatives outlined by the national cancer institute?

TJ: We expanded the new program's initiatives to include one that is very focused on the incorporation of engineering and new technologies. This program is headed by Sangeeta Bhatia and Bob Langer and represents an example of our commitment for trans-disciplinary or inter-disciplinary cancer research.

On a related note, when we were received by last fall by a group of external site visitors organized by the National Cancer Institute, they were blown away by what we were planning to do. They commented in very explicit terms that this was a paradigm shifting initiative that was really at the forefront of the nation's approach to cancer and cancer research. The review group was absolutely, completely convinced that this new direction that we're taking is important and that we were on our way to organizing ourselves appropriately. And the building of course was a big portion of that—seeing a nice 350,000 square foot building rise up is pretty good evidence of that.

MURJ: Does the increase of space correlate to the labs' ability to increase its number of members?

TJ: The increase of space the new building is a reflection of the expanded program involving engineers. We also included a very large number of new faculty. We had thirteen faculty in the current Koch Institute and we'll have 25 or more in the new building. We're basically doubling.

In addition to that, we set aside space beyond the core facilities for a new class of investigators called the "Koch Clinical Investigators." These are individuals that will have small research labs. They will not have faculty positions at MIT, and in fact, they'll likely have faculty positions at a different institution—a local clinical institution, for example. We have one such clinical investigator now, who is a practicing oncologist at the MGH, and she will have a small lab here in the Koch Institute. We haven't had space for such an individual before, and we hope to have 7-8 such individuals in the future.

MURJ: How is the composition of the new faculty selected, and what were you looking for?

TJ: On one hand, half of the faculty for the new cancer center were preordained, because they were the occupants of the old Center for Cancer Research. The engineering faculty were chosen largely from the existing MIT faculty in engineering, in a process that involved input from the existing faculty. There was a lot of help from Bob Langer and Doug Lauffenburger, thinking about people who would be a good fit for this new enterprise. This was based on a lot of pre-existing relationships, past and current collaborations, demonstrating an interest in this kind of cross-disciplinary research, as well as thinking about the kind of people and the kinds of approaches that we wanted to focus on in the new building.

We were very interested in translational cancer research. The engineers whom we selected had demonstrated interest in building new devices for cancer monitoring or new materials for cancer therapy, and that was particularly attractive.

MURJ: I've read that there were a lot of collaborations planned for the new cancer institute between various hospitals and biotech companies. Where most of collaborations pre-existent or are any of them new?

TJ: There have been preexisting collaborations with individual labs, and that's how things have been organized in the past: one-off, ad hoc, as-needed interactions on a lab-by-lab basis. But in the new building with the new structure, we are hoping to and are formalizing some of those interactions. We created a new program, which we called "Bridge Project", to bridge between the Koch institute and others at MIT to the local hospitals, like MGH and the Dana-Farber. It is to foster collaborations between clinical investigators there and scientists and engineers here.

This project is still taking shape. There are some ongoing collaborations that have been setup under the Bridge Project, but now we're seeking funding, largely philanthropic funding, to support more such interactions.

MURJ: What were some of the main stages that went through the planning process.

TJ: The genesis of a new building dates back to before my time as director. Pre-2001, there were already discussions of replacing the existing building because it had really run it's course; it was not a adequate facility for modern-day, cutting-edge research. So there was already a discussion ongoing about replacing this building, and I took on the role and responsibility of director and convincing the administration that we needed to replace the building in 2001. That was during the presidency of Chuck Vest, and Bob Brown was the provost at the time. As [Chuck Vest's and Bob Brown's] tenure here at MIT was coming to a close, they agreed that there should be a replacement facility for the Center for Cancer Research at the time, around 2003. However, because Chuck knew that he was leaving, he didn't commit the Institute to building the new building. He left that decision to his successor, Susan Hockfield.

After Susan's arrival in 2004, she and I started to talk about the future of cancer research at MIT and what was happening already in terms of our research programs. I began to talk to her about some of these nascent interactions with the engineering faculty, such as in the case of nanotechnology, for example, and she got very excited by this. When we started to talk about building a new building, it was actually she that suggested maybe we build a building that is twice the size as the one that we were envisioning, and that we house the scientists and engineers under one roof. So I give her the

Movers deliver rows of refridgerators to fill the new lab spaces.
Credit: Cynthia Chen

credit for purposing that rather bold initiative, because the building was going to be twice as big and cost twice as much to build. But it seemed like the right thing to do, both to her and to me. That was the 2004-2005 timeframe.

Along the way David Koch became very interested in and very supportive of our activities, and encouraged MIT through his donations to accelerate the process. He was very clear actually, saying that he would contribute his money—at that stage, it was about 50 million dollars—if MIT would commit to begin the process of design and construction ASAP. MIT agreed to do so. That led to the planning process with architects, and so on.

About two years later in 2007, David doubled his gift to 100 million dollars, and at the time in recognition of this enormous contribution, we changed the name of the institute to the David H. Koch Institute for Integrative Cancer Research. In the spring of 2008, we broke ground with a very aggressive construction schedule to meet our needs and David's expectations about getting it done quickly. We planned a thirty-month building schedule, and they completed just about thirty months. It went extremely smoothly.

MURJ: What are the biggest challenges of moving so many labs to the new building?

TJ: One has to do with the very sensitive equipment that we have. It's tuned to work and has to be carefully maintained to work properly. Every time you move sensitive equipment, you have to worry that it's not going to work as well when you relocated it. We have a lot of such sensitive equipment.

The other is the simple fact that some of these faculty have been here 35 years. If you take Phil Sharp, for example, or Nancy Hopkins, or Richard Hynes, these are individuals that started their career here at MIT in the early 1970s. So it's really been more than 35 years, and they've accumulated a lot of stuff. I've been since 1992, and I've accumulated a lot of stuff myself.

Just cleaning up and clearing out all that research material and other material, so it won't be taken to the new facility and just take up space over there—that's an overwhelming process. It's sort of a liberating process at the same time. It's nice to clean it out, but it's certainly challenging.

"The increase of space the new building is a reflection of the expanded program involving engineers."

MURJ: How do you actually move all the various lab contents? I have a vision of grad students and postdocs just grabbing materials and moving it across the street…

TJ: [Laughs] It's not illegal, but [that approach] is probably not advised.

Over the course of about 6 weeks, movers will come to the current Koch Institute building as well as the labs of the engineers, and in a really accelerated time frame, they will pack a lab, put it into crates, put it onto trucks, drive it about a minute, and unpack inside new building. Remarkably, a lab will be moved within three days: they'll come, pack and move within three days. Our hope is that each lab will be up and running with pretty minimal interruption.

Of course, there are a lot of labs, so not every lab will be moved within three days. Over the course of six weeks, however, all 25 labs should be moved. The very first people to move will be November 5. My lab moves two weeks after that, the third week of November. We should all be in by mid-December.

MURJ: Could you summarize the type of work that your lab does and how you think that it will be propelled by moving to the new cancer institute?

TJ: My lab works on genetically-engineering mouse models of cancer. We are very interested in the molecules that contribute to cancer initiation and progression—that is the basic science understanding of those molecular changes. And we're very interested on how to treat cancers based on that knowledge. In that regard, our program will benefit significantly from the new institute.

I mentioned earlier, interactions with engineers in relation to RNAi therapeutics. We have many systems in which we can test well the ability to get RNAi therapeutics to work. We've already begun collaborations with some of our colleagues like Sangeeta Bhatia, and I'm really looking forward to that. Her lab and my lab are on the same floor and we have neighboring offices in the new building.

I'm also looking forward to the preclinical therapeutics and imaging facility upstairs, because we do a lot of tumor development and tumor treatment studies. That

facility which is miles better than what we currently have; it will be a huge help to us.

But I'm really looking forward to the cross-disciplinary interactions across the board. We dabble in tumor immunology, and I'm really looking forward to working with immo-engineers…but really everybody.

MURJ: Where do you envision the new cancer institute ten years from now?

TJ: I personally think that's a really important question and one that should be asked. How would we measure success ten years in? I would measure success in three ways.

Firstly, it would be that we can demonstrate that there was value added to bringing together scientists and engineers: to show that in this new environment, we are doing something differently and better than we could have done before, with tangible examples. We now have several [examples], but I would like to see many, many more. And that is something that I'm sure will be an obtainable measure of success in the ten-year timeframe.

The second is our training of students and post docs who themselves are highly cross-disciplinary. The faculty are the head of the labs, but as you know the research is done by the students and post docs—and I would include the undergraduate students in that. In the new building we will give students and post docs an exposure to lots of different things that they could not have seen in a traditional biology or a traditional engineering lab. I think we're going to generate a class of cancer researchers that look nothing like their mentors. They'll be much broader and speak many different languages; they'll be fluent in many different languages. To see how exactly that was successful is harder to know, but I'm sure we will see it.

And the third is to see examples of our discoveries and new technologies impacting cancer patients…the translational aspect to go beyond the discovery and participate in the implementation of these new insights and new technologies into better diagnostics, better treatments, better ways to monitor the disease. This is something that I've been emphasizing strongly, and the clinical interactions that I alluded to before are key parts of that. I know that our individual labs will be successful in terms of their research, but I think that in order for the institute to be successful as a whole, we need to demonstrate clinical impact. I'm really looking forward to that.

A Special Look into the Humanities at MIT

By: Anna Ho

Here at MIT, we are continually exposed to an impressive breadth and depth of cutting-edge scientific research, which makes it easy to forget that a less technical branch of research is conducted alongside it.

The MIT School of Humanities, Arts, and Social Sciences is spread over twelve departments, ranging from Music and Theatre Arts to Anthropology to Literature. Around 150 undergraduates have chosen to major underneath this umbrella, and over 400 have chosen to minor. The departments employ 174 full-time faculty members, including three Nobel Prize winners, nine MacArthur Prize winners, and four Pulitzer Prize winners. They conduct research and contribute to their respective fields, like their science and engineering colleagues, but doing so at an Institute of Technology presents challenges. "When you tell people that you teach literature at MIT," said Professor Noel Jackson of the Literature Department, "the reaction is often 'oh, they do that there?' To some degree, you get that within MIT as well – students don't know that this kind of research is bona fide."

A department that is, in Jackson's words, "often called upon to justify its existence," must be able to address the question: what are the humanities doing at MIT?

Dora Gao is a senior double-majoring in physics and history. When asked how the two fit together, she laughed and said that she is "still trying to figure this out". "How can I be simultaneously interested in physics and history? I guess that history is somewhat similar to science, because in the end you have the same goal: you're trying to determine something and you're using… your evidence, and you have to use [it] to make a compelling argument…There's a stereotype that science is objective and the humanities is not, and that's not true." She has participated in research in both departments, and has found fundamental similarities between the methods of inquiry. "When you start doing research at a higher level, you're subject to the same standards…You pose a question, and the driving process is the same. You're subject to the same standards [in history as you are in physics]."

"The humanities teach us how people think about their own lives, and therefore how to more effectively address real-world problems."

She is not the only one with this perspective.

Professor Jackson believes that advanced work in both fields - the "protocol" - is carried out in a "fundamentally similar" way. "MIT prides itself on having high expectations [of its research]," he said, "and this is no different for the humanists."

Jackson's research interests lie primarily in 18th and 19th century British literature, but more recently he has been branching into Modernist topics of the 20th century. He seeks to "understand the nature of human imagination" by examining art and art experience, particularly in poetry. His work addresses how human beings perceive beauty: the "how" of aesthetic experience, which, to him, "cuts to the heart of the purpose that I find in the humanities". He wants to know "why certain texts are acknowledged as powerful, beautiful, or sublime," and about the details of "our contact as human beings with these art objects." Research of this kind consists primarily of reading, particularly of colleagues' work, writing, and sometimes "stepping away" from the work to think. Professor Jackson has had UROPs perform research on, for example, the sonnet and its revival in 18th century England, but says that there is an "endless possibility" of research projects that students can get involved with.

Professor Jackson's work, and his research process, may seem far

removed from that being performed by his colleagues in the science and engineering departments. Professor Anne McCants of the History Department, by contrast, conducts research that tends to be "very technical." One of her projects is examining how Gothic Cathedrals in the 12th and 13th centuries were financed: how resources were allocated in order to put together a work of monumental architecture. She also studies the 17th-18th century Dutch Republic, which involves large data collection and analysis. Her UROPs have traveled to Amsterdam to work in real archives, but she concedes that the department probably does not use UROPs "as intensely as in the lab disciplines." Her day often involves sitting at a computer running statistical software, but the "key" is actually reading. "I spend a disproportionate amount of time reading," Professor McCants said.

There seems to be a reading-intensive theme carrying through humanities research. Professor Margery Resnick of the Foreign Languages and Literatures department believes that this process of critical evaluation—of reading critically—is essential to any scientist. "I can't imagine living in the 21st century and not being able to critically evaluate data of all kinds," she said.

Professor Resnick's interests lie in Women's Studies and Hispanic Culture, and in the intersection between the two. She is currently involved in AMITA, which is a memorial oral history project about MIT women that heavily involves UROP students. These students interview female alumni (starting from the Class of 1922) about their lives, and the transcriptions of the interviews are stored in archives. The project looks at how these women were involved within the school and what they are currently doing with their science and engineering degrees. The UROP students choose to interview women whom they are particularly interested in; for example, one student was able to interview the first African American woman at MIT. Students learn the ins and outs of the entire formal oral interview process. At this point, enough interviews have been amassed that the project team is now working on digitizing them to allow for more public access. In addition to this project, Professor Resnick runs the International Institute in Spain, which was founded in the 19th century by an American suffragist woman, to educate Spanish women. Additional research interests include Spanish female writers and examining, through their literature, what liberation from dictatorship meant to them. She considers her area of research in the humanities "very fulfilling", as it enables her to measure what she sees through a "variety of lenses". In her opinion, this is crucial for any engineer, as they need to understand both practical reality and how people imagine reality to be. "Engineers need to realize that you can't just tell someone that your new machine will give them a better life", she said. "You need to get into a person's mindset. We are both our lives and our imagination."

She pulled down the book *Chaos and All That* by Liu Sola from the shelf, as an example of literature that highlights the variety of cultural perspectives. "People tend to think of [globalization] as an economic and political reality, but it's also cultural," Professor Resnick said. "It's important to realize that in life you're always coming across people who have been brought up to think differently – without this kind of understanding, you can't be a good citizen." To Professor Resnick, the humanities teach us how people think about their own lives, and therefore how to more effectively address real-world problems.

Noel Jackson, Associate Professor of Literature.
Credit: lit.mit.edu

Along similar lines, Professor David Jones of the Science, Technology, and Society (STS) Department believes that by studying both the humanities and the sciences, which are "distinct, with common features", we can better understand and therefore improve human experience by applying insight and perspective about the past and present to building the future. He points out that through literature and history we see that there is "something trans-historical about the human experience," such as competition over resources. Another example is how people have struggled with mental illness over the centuries in a way that resonates with modern experience.

His own research concerns the history of medicine. "In a simple world," he said, "we would use treatments because they work, but the data shows that often people who need the medication don't get it, and people who get it often don't need it." His work addresses why this is the case: how it is that certain treatments don't work the way patients (and even doctors) commonly believe that they work, and why patients are keen to

"The humanities do have a place here at this Institute of Technology, regardless of how they are defined."

take medication in some cases but reluctant in others. Professor Jones has found that ineffective decision-making is sometimes a result of a poor common understanding of physiology – the model of the angioplasty procedure, for example, is grounded in knowledge of plumbing, not coronary arteries. He has also found that when we look into the history of efforts to understand cardiac surgery, we see a lot of effort put into how it works and whether it works, but not much focus on the study of complications. "For example," he said, "imagine that you cut off someone's fingers. The question you ask is: does the hand still work? If the procedure affects the foot, we might not notice. If it makes someone depressed, an orthopedic surgeon might not notice, because scientists often fail to notice effects that lie outside their area of professional expertise. They often have an excellent knowledge of the benefits, but not of the unforeseen negative effects that are, on the surface, unrelated." He listed the Vioxx case as another example of this phenomenon: the painkiller was advertised for a long time as a safer version of aspirin, but was recently taken off the market because it turned out to be dangerous for health. His research investigates why this happens: why it is easier to create knowledge of benefits, than of complications.

The nature of his research had led him to develop a more "pragmatic" attitude to the humanities: that through examining historical cases, we are able to "distance ourselves from the material, which makes it easier to find and identify reasons for why things happen" and to "identify forces that might still be relevant today." The humanities can inform the way that scientists typically think about issues, such as blood-letting. Were doctors just dumb? "Probably not," Professor Jones points out, as the procedure was performed for over 2000 years. George Washington harangued his doctors to bleed him more; patients clearly wanted the treatment, and his research looks at why this might be so. He explains that "blood-letting could treat fever, fear, and agitation, without being an actual cure for the disease. In many cases, it provided treatment for conditions that had no cure. Medication today, by contrast, often has no discernible impact, and so should it really be a surprise that patients are often unwilling to take it?"

Professor Jones believes that scientists who do not understand history are "susceptible to repeating it", as history can bring about awareness of patterns, and raise questions that scientists will need to address. In addition, history challenges the way that we conventionally think about issues, which in turn challenges science: "for example, most geneticists will tell you that obesity is 95% genetics. But obesity has doubled in the past six years – how could genes be the leading cause of this? There is clearly much more to obesity than just genetics. Phenomena are harder to explain than many scientists think, and history and the humanities continually raise this challenge."

To Professor Jones, then, the sciences and the humanities need each other in the problem-solving process. Gao has also found this to be the case. "One thing I've realized [in physics] is that learning how to do the problems is as important as learning how to think, and how to view the world – physics helped me view history in a scientific way, and I know that…I'm a much better writer because of [history], which definitely helps me in science." When asked whether she feels that she has a distinct humanities way of thinking and a distinct physics way of thinking, she paused, and said: "I'd like to think that I've internalized my training from both majors so much that I can apply the two without thinking 'I'm using my history brain now' or 'I'm using my physics brain now'. I do think that it's really important to have the two…I think that physics is the base of all the sciences – whereas history is the hub that everything's connected to. Economics has history, literature has history, and as I study history more and more, I'm starting to realize that the nature of science in a certain period explains the views that people had, which affects their actions, and affects politics…and I see that the role of history in the humanities reminds me of the role of physics in the sciences."

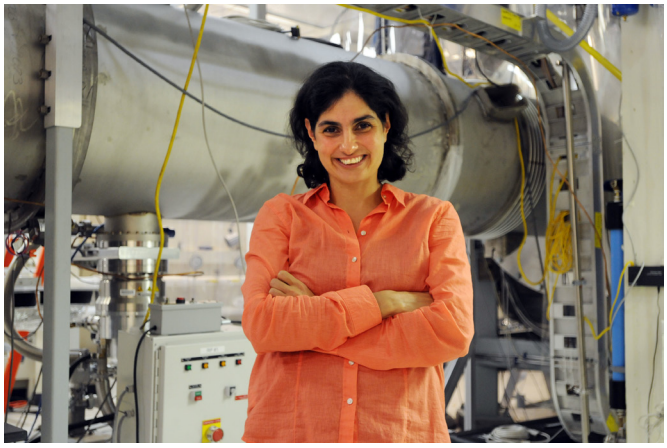
Her words echo those of the humanities departments' professors, who suggest that the humanities do have a place here at this Institute of Technology, regardless of how they are defined. The discipline needs a scientific approach to remain pragmatic, but equally the sciences need a humanities perspective to remain relevant and effective.

MacArthur Fellow Nergis Mavalvala's Gravitational Waves

By: Pratiksha Thaker

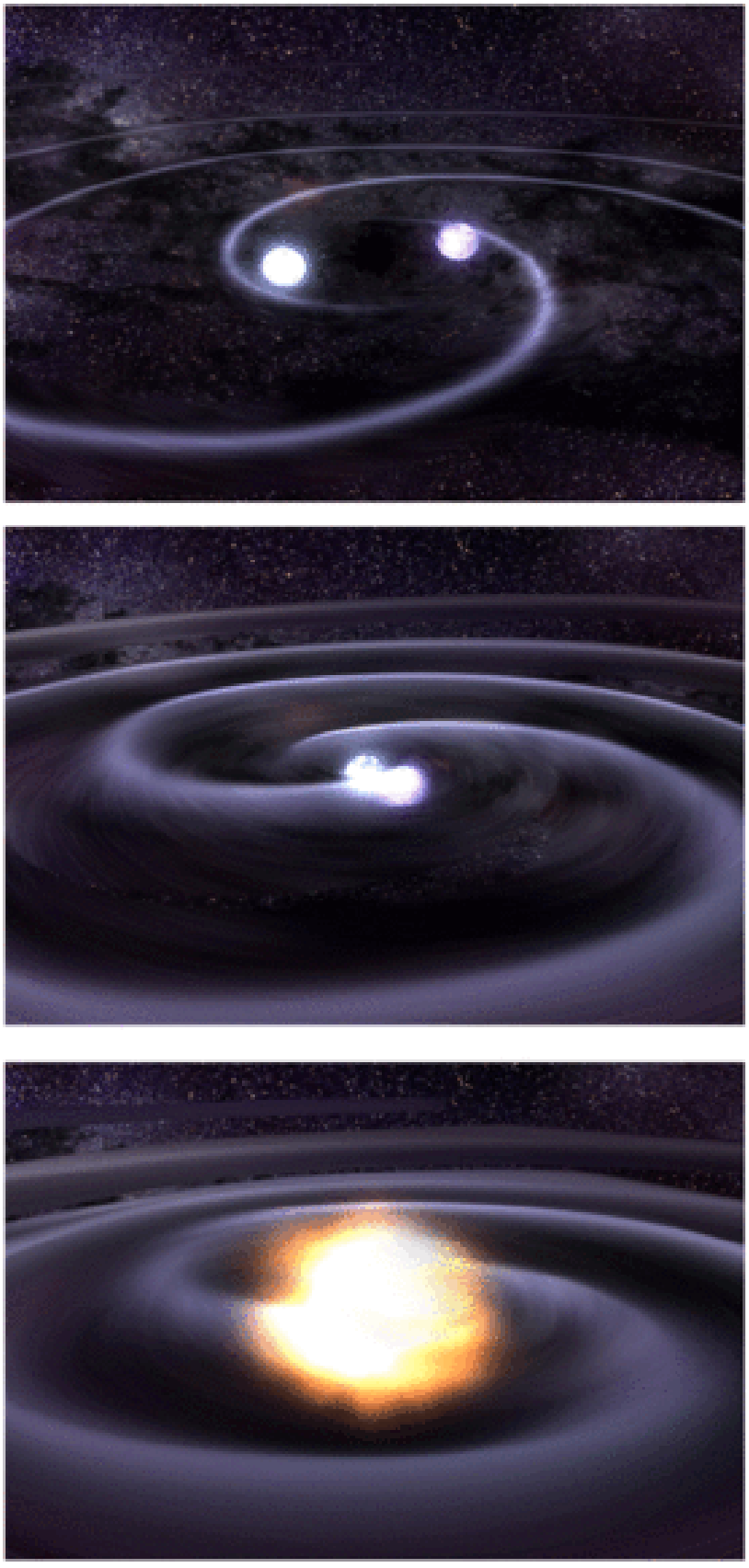
The earliest astronomers mapped the night sky by relying on their eyes: the brightest stars and planets were reference points for scientists and explorers alike. The invention of the telescope in the early 17th century brought the sky into focus; each improvement revealed more heavenly bodies farther away. Increasingly accurate techniques have allowed astronomers to create nearly complete maps of our entire universe. To date, most of the observed universe has been mapped using various forms of electromagnetic radiation: visible light, microwaves, infrared radiation, and other forms which have provided us with the capacity to look deep into the beginnings of the universe. As the universe expanded, slowly-forming planets and stars emitted radiation which, as it reaches us from light years away, now allows us to "see" the universe as it was billions of years ago. However, our ability to look into our own past is still limited by the objects that are perhaps obscured by dense clouds of stellar dust, or are dark—like the black holes that absorb radiation—and are invisible to our telescopes. It may not be long before scientists break past this limitation, however: with the advent of research in the field of gravitational waves, researchers may be able to look even further into the depths of time, to a point where the universe was just beginning to form.

Professor of Physics Nergis Mavalvala works with gravitational waves at the Laser Interferometer Gravitational Wave Observatory (LIGO) in addition to her teaching and research at MIT. Her work with gravitational waves began when she was a graduate student at MIT, continued through her post-doctorate work at Caltech, and is still in progress as her current research focus. Professor Mavalvala recently received the MacArthur 'genius' grant for her developments in the field of gravitational waves. The grant, which consists of $500,000 in research funding with "no strings attached," will allow her to push the boundaries of investigation into these new sources of information.



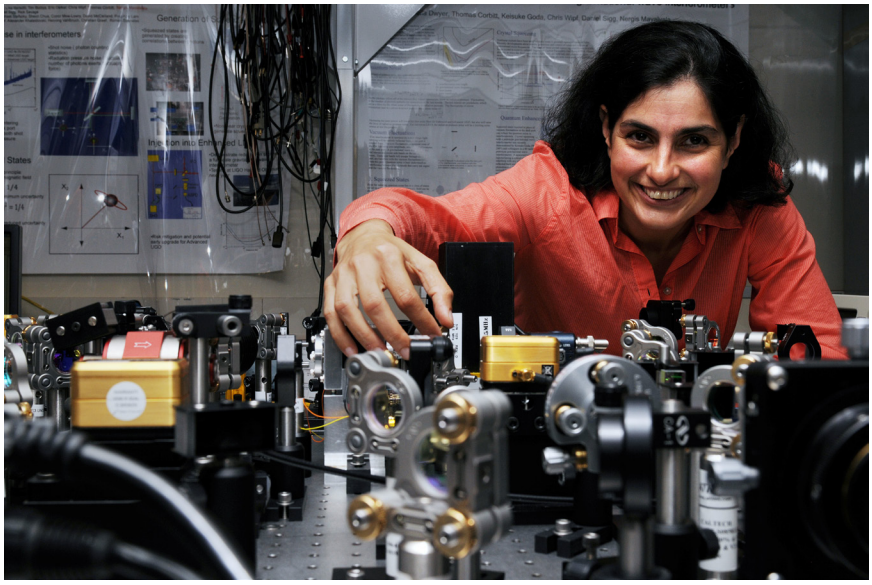
Professor Nergis Mavalvala (Ph.D '97) has been studying gravitational wave astrophysics since 1990. She is currently part of the LIGO Group at MIT and received the MacArthur Fellowship in 2010.

Gravitational waves differ from electromagnetic radiation in the fact that the former is generated from accelerating massive bodies, while the latter is generated from accelerating charges. In everything from radio antennae to x-ray machines, moving charges generate radiation that can then be detected and measured. Gravitational waves work in a similar manner, which can be visualized through the rubber-sheet analogy. If a rubber sheet is stretched out, a ball placed in the center would cause a small dip in the sheet—much like how a planet pinches spacetime. If the ball suddenly moves, it causes ripples in the sheet. Similarly, when a planet accelerates, the result is a propagated disturbance in spacetime, which can in theory be measured using extremely sensitive detectors—interferometers—on Earth. The expected result of such measurements is that scientists will be able to 'see' supermassive objects that otherwise emit little to no light. If the experiments are successful, says Professor



Gravitational waves are created as two white dwarf stars orbit each other, spiraling towards eventual collision.
Credit: NASA/Dana Berry, Sky Works Digital

Mavalvala, "it would be the first time that we'd be able to see what the spacetime next to a black hole looks like.[…] We'll be able to look at highly curved spacetime in the regime of small gravity. Another place where I think this will play out," she adds, "is—we don't really know what neutron stars, some of our favorite gravitational wave sources, are. We don't really understand the state of the material—the nuclear matter that neutron stars are made of. And there's an expectation that measuring the gravitational radiation from neutron stars and combining it with measurements from electromagnetic radiation will allow us to start to get a handle on exactly the composition of the neutron star."



Professor Mavalvala presents one of the LIGO Group instruments. Most prominently, Mavalvala works on designing different parts of the LIGO interferometers, huge detectors that can detect gravitational waves.

Although the prospect of detecting gravitational waves is enticing, the challenges that must first be overcome seem daunting. Interferometers, the major tools used to study gravitational waves, consist of very large mirrors—about 40 kilograms each—that must be isolated from vibrations to the extent that even their internal vibrations could affect measurements. Professor Mavalvala explains that "an interferometer splits a light beam into two halves at right angles to each other, and the beams go out to distant mirrors. In our case, at LIGO, the mirrors are four kilometers away. Then the beams reflect off of the mirrors, come back, and we measure the phase difference in the light. We measure the interference pattern between two beams of light, and by seeing how light or dark the patterns are we can tell the relative displacements of the mirrors." The major obstacle to such interference observations is that the fluctuations caused by gravitational waves are generally on scales as small as 1/1000 the diameter of a nucleus of an atom; researchers are constantly seeking ways to make the interferometers more sensitive and less vulnerable to the much stronger external forces on Earth—everything from seismic vibrations to tumbleweeds colliding with the LIGO facilities, which prompted one researcher to recommend a 30-meter tumbleweed barrier.

In the process of finding such isolation from external forces, Professor Mavalvala entered a related and equally intriguing field of work: macroscopic quantum measurements. "Three years ago now, I was working with another graduate student, an MIT graduate student now on the faculty here—we discovered something in our system, something nobody had thought of before—how to optically trap a mirror. Since the '80s, people have been trapping atoms to cool atoms down so that they can get them into quantum states. And we figured out how to do this with large objects, and that's become the building block for what we do now." By cooling atoms to extremely low kinetic energies, researchers are able to render them so still that their motion was at miniscule, quantum scales. The macroscopic quantum measurements are a result of efforts to improve the mirrors' precision to a degree that would allow for gravitational wave measurements; the mirrors have to be rendered nearly motionless in the same way that atoms can be. As the experiments demanded more precision, researchers had to take into account effects on the quantum scale—the only scale at which the mirrors were still moving. For the first time, quantum mechanics played a role in the physics of large-scale objects. And, notes Professor Mavalvala, the research has led to mirrors so precise that "the 40-kilogram mirrors of the LIGO detector will be held in place by laser beams such that the motion of that mirror along the interferometer axis will be one quantum of energy," the smallest unit that energy can take. In fact, the development of this field has become the main focus of her current work, due in part, she says, to her students' fascination with the topic.

A great part of Professor Mavalvala's work, both in macroscopic quantum

"With the advent of research in the field of gravitational waves, researchers may be able to look even further into the depths of time, to a point where the universe was just beginning to form."

measurements and otherwise, is influenced by the students she works with. Out of the aspects of her work that she enjoys the most, she states without hesitation that "the time I spend working with my students, my grad students and my undergrad students, that's the most valuable." Even in the scope of her cutting-edge research, Professor Mavalvala finds niches for students ranging from undergraduate freshmen to senior thesis students to graduate students. She emphasizes that she is "continuously learning" from her graduate students as well as her undergraduates, noting that "there comes a time when the graduate students have been working on a project for long enough that actually know way more about their projects than [she does]."

And what would Professor Mavalvala say to the undergraduates at MIT?

"Get involved in research, and just don't be afraid to try stuff. Because you will make mistakes and you'll feel goofy and your group will take a while to recover from things you've done. But the only way to not make mistakes is to not try things at all. And that's the way I learned in lab—mainly from doing work and taking risks. As an undergraduate I'd receive some instruction, and then I'd say oh yes, I get it, and I thought I did, and as soon as they leave I'd realize I didn't know what I was supposed to do—and then I'd try something, and a good fraction of the time it didn't work out, and I'd have to try something else, and eventually you just learn. And so my advice would be to get involved in research, don't be afraid to make mistakes and don't get frustrated."

Quantum Unification

By: Mobolaji Williams

Introduction

Imagine that you're a physicist. It's the late 19th century and you're sitting at your desk late at night pondering the state of your discipline. You recall the great theories of Newton, Maxwell, and Boltzmann. You think, from Newton we were able to understand the motion of the heavens, the earth, and everything in between. From Maxwell, we realized that the invisible forces of electromagnetism are connected to the visible light that illuminates the world. From Boltzmann, we learned to describe systems containing large amounts of particles in terms of probabilities and distributions. But, you say, the reliance on probability is too vast, albeit not the physicist's fault: In statistical mechanics the laws are definite; our information is not. Remembering the implications of these theories you look upon our world as the inevitable conclusion of a set of precise, although unknown, initial conditions. "The world is one big differential equation," you think to yourself, "and we have solved it."

We all know how this story ends. Quantum theory comes along and destroys any previous perception of a deterministic universe evolving to some inevitable end from the initial conditions of today. Instead we are given a picture of the world in which its microscopic origins obey laws which, although definite, lead to probabilistic results. All this was shown during the first three decades of the twentieth century. The theory was initiated by Planck who first postulated the existence of "quanta", extended by Bohr who applied the theory to the hydrogen atom, and completed by Dirac, Heisenberg, and Schrödinger. More importantly, the extension of this quantum paradigm past the mechanical motion of particles led to a unification of physical ideas which has culminated in the perception of physical quanta as a pervasive concept in physics. Nowadays, no physical formulation past currently accepted theories can claim to be complete without some incorporation of quantum motivations or concepts. The steps leading to the mandatory place of quantum ideas in our physical theories has an intriguing back story which follows the path that man took to discover the inherently quantized nature of the world.

Essence of Quantum Mechanics

Our Probabilistic World

The earliest experiments concerning quantum mechanics revealed the probabilistic nature of the world. The famous double-slit experiment showed that electrons, although presumably particle-like, do not have definite trajectories and instead can extend in a continuum like waves. The Stern-Gerlach experiment revealed that electrons have an intrinsic angular momentum called spin and used this new discovery to show that all quantum experiments have an aspect of chance. From these experiments we learned that the essence of quantum mechanics is indeterminacy. Unlike its classical counterpart, quantum physics cannot make definite predictions about the outcome of an experiment. All of the conceptual framework and calculational machinery goes towards understanding the computation of probabilities and the various technicalities associated with the energy spectrum that they predict. There are various formulations of quantum mechanics but each shares this inherent property of indeterminacy.

The wave mechanics invented by Erwin Schrödinger describes quantum systems in terms of a quantity called the wavefunction which describes the probability of a particle existing in a certain vicinity of space. This wavefunction obeys a wave equation, the famous Schrödinger Equation, which determines how the wavefunction looks for a given potential. The common analogy is that Schrödinger's equation does for quantum mechanics what Newton's Second Law does for classical mechanics; the equation provides a differential equation which can be solved in specific contexts to obtain information about a system. From this wavefunction one can calculate measurable quantities such as the line spectrum of hydrogen and the relative shapes of the electron orbitals which act as the basis of bonding theory in chemistry.

Matrix mechanics developed by Werner Heisenberg describes quantum phenomena in terms of matrices. When Heisenberg first formulated his theory he represented physical observables such as momentum and energy in terms of

linear arrays of numbers and only later discovered that these arrays could be written as the then abstract matrices of linear algebra. In general, any physical observable can be written as a matrix and the eigenvalues of these matrices represent the possible values obtained after a measurement of said observable.

It is worth mentioning that in spite of each physicist's respective eminence in his field, Heisenberg and Schrödinger had little professional respect for each other's formulation of quantum mechanics. Schrödinger, a classical physicist at heart, had trouble reconciling Heisenberg's discontinuous mechanics with his own physical intuition. His opinion of Heisenberg's theory is briefly stated in the following quote. "I knew of his theory of course, but felt discouraged, not to say repelled, by the methods of transcendental algebra, which appeared difficult to me, and by the lack of visualizability." Heisenberg's comments regarding Schrödinger's theory are even more scathing. In a letter to Pauli, he writes, "the more I think about the physical portion of the Schrödinger theory, the more repulsive I find it. What Schrödinger writes about the visualizability of his theory is probably not right." This battle between two distinguished scientists is reminiscent of the entitlement feud between Newton and Leibniz.

The path integral formulation of quantum mechanics, although initiated by Dirac, was extended to it its full power in Feynman's PhD thesis and his subsequent explorations into quantum electrodynamics. The basic idea is that the time development of Schrödinger's wave function can be interpreted as a quantum particle traveling through space. The main point, however, is that this particle has a completely indefinite trajectory because it travels through all of space as it moves from its initial point to its end point. In this way, Feynman expressed quantum time evolution as being the result of the contribution of every possible way a particle can go from one state to the next.

Although each of these formulations is quite distinct, modern quantum mechanics can be considered to be some combination of the three, with each theory having its advantages in certain problems. Wave mechanics is useful in studying the properties of atoms and molecules. Matrix mechanics allows us to study spin systems and simple quantum models such as neutrino oscillations. Path integrals are currently the most efficient method by which to study quantum field theory. Still, each of these theories, as summarized in Figure 1, detail the indeterminacy which is essential to any quantum theory.

Claustrophobic Quanta

The Heisenberg Uncertainty Principle is the most informative single equation of the many which describe quantum physics. Stated precisely, it asserts that upon measurement of the momentum p and position x of many identical quantum systems, the statistical results must satisfy

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

where σx is the standard deviation of x and σp is defined similarly. We can loosen this exact statistical statement if we consider σx to simply be the uncertainty, say Δx in a particle's position. Then, Heisenberg's uncertainty principle is the statement that unlike a classical particle, a quantum particle is never at rest because it always has some finite momentum Δp spread over a region of Δx. Since a quantum particle can find no rest, it always has some finite energy. For the case of a harmonic oscillator system we can calculate the minimum of this energy using elementary calculus and the Heisenberg Uncertainty Principle (try it!). The result is simple: we find that the minimum energy for a quantum harmonics oscillator system is

$$E_{min} = \frac{1}{2}\hbar\omega$$

The claustrophobia of quantum nature is given a precise definition in Heisenberg's uncertainty principle. We can give a physical description to this principle via the common thought experiment of a microscope shining light on a single particle. Suppose we use a microscope to view an electron. To view an electron most precisely requires the use of light of a very high wavelength. But, for light, high wavelength means high energy. So, to pinpoint the position of an electron we need to bombard it with very high energy light. However, bombarding the electron with high energy light makes the electron more energetic and hence makes its momentum less certain. Hence, higher certainty in position implies less certainty in momentum. This thought experiment, called the Heisenberg microscope, can be reversed to imply the other extreme of the uncertainty principle.

The story leading to the so called "Heisenberg microscope" is an interesting one. As a doctoral student Heisenberg nearly failed his final examination in experimental physics due to lack of preparation. His examiner, cognizant of Heisenberg's lack of preparation, asked many specific questions about experimental

technique most of which Heisenberg did not answer correctly. The examiner consequently declared Heisenberg should not obtain his Ph.D. The decision was only reversed by an appeal made by Heisenberg's adviser who recognized his student's scientific ability. One of the questions that Heisenberg failed to answer had to do with the resolving power of a microscope. After his disastrous examination, Heisenberg researched all the questions he missed including this one. This research led to the physical picture of the microscope analogy which now bears his name.

Quanta of Fields

A great leap in quantum physics occurred when Dirac applied the formalism of quantum mechanics to fields. Suddenly physicists realized that particles are not the only entities which have seemingly anomalous behaviors at the microscopic scale. In particular quantizing the classical field initiated the discipline of quantum field theory which concerns itself with the many strange properties of these quantized fields. One of these strange properties is the existence of a quantum vacuum from which particles can emerge and into which particles can disappear. This particle birth and death is called creation and annihilation in the standard physics literature and represents a significant departure from the classical physics perspective of an eternally stable particle.

One of the benefits to this new formalism of field theory was the unification of the particle field concept. Our entire conception of physics in the 19th century was built upon the distinction between fields and particles. Fields (gravitational and electromagnetic) are generated by particles and act on particles, while particles governed by Newton's Second Law act as source of fields and are acted upon by fields. This dichotomy disappeared when quantum field theory was developed and particles were now interpreted as mere excitations of a quantum field. For example, the photon, commonly known as the particle for light, is also the excitation of the electromagnetic field. Similarly, the electron is an excitation of a corresponding electron field which is not observable in everyday life because of its extremely minute effects. In general, all quantum particles are associated with a quantum field, and all quantum fields have particles as their excitations.

"Unlike its classical counterpart, quantum physics cannot make definite predictions about the outcome of an experiment."

Our conception of the vacuum changed with the development of quantum field theory. We soon realized that not only do quantum mechanical systems of particles have a residual vacuum energy due to the Heisenberg uncertainty principle but similar systems of fields have the same property. This vacuum energy which pervades both space and time acts as an endless repository of energy which allows particles to be created and annihilated in particle interaction processes.

Field Theory and the Life Giving Vacuum

An illustrative analogy of the quantum vacuum is the ocean. Imagine that we are exclusive land dwellers who cannot observe the ocean. All we can see are the sea creatures who spontaneously pop out of the ocean only to pop back into the ocean at another time. In this way the ocean, which we cannot directly observe, is the quantum vacuum and the sea creatures that spontaneously jump in and out of the invisible ocean are our observable particles. Although we can't see this ocean, we still know that it is there because the creatures must come from somewhere. Similarly we can infer the existence of a dynamic quantum vacuum from the birth and death of particles that we readily observe.

The Feynman Computer

Essential to the understanding of quantum field theory was the development of a computational framework which made calculations simpler. When quantum field theory was first developed by Dirac and Pauli, it used an extension of time dependent quantum mechanics which produced unambiguous results (to certain approximations) but was extremely difficult to implement. Even when quantum electrodynamics was reformulated in the late 1940s, most of the reformulations bore much of the complicated formalism of the previous theory. The only formulation which departed from these original complicated formalisms was the particle interaction model based on Feynman diagrams, which essentially abandoned the quantum mechanics of Heisenberg and Schrödinger. Feynman diagrams are pictures which model particle interaction processes in quantum physics. They are particularly useful because they provide an organized and relatively transparent method of keeping track of the complicated expressions in quantum field theory.

Schrodinger

$$i\hbar\frac{\partial}{\partial t}\psi = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi$$

Indeterminacy in position encoded by the probability density $|\psi|^2$.

Heisenberg

$$\hat{H}|\psi\rangle = E_n|\psi\rangle$$

Indeterminacy in measurement encoded by the large spectrum of eigenvalues.

Feynman

$$K = \int_{x_0}^{x_f} \mathcal{D}x\; e^{iS[x]}$$

Indeterminacy in trajectory encoded by adding the contributions of every path with the same final and initial conditions

More specifically these Feynman diagrams consist of different types of lines connected in various ways to one another. Each particular line or connection we use corresponds to a mathematical quantity. When all of these quantities for a particular diagram are multiplied together, we obtain a result which represents the probability of occurrence for the process. This computational framework is named after Richard Feynman who developed it while studying processes in quantum electrodynamics. As he developed these methods he soon realized that he had found a method of "doing" quantum field theory which was significantly easier than other contemporary methods.

The classic story demonstrating the efficiency of Feynman's method concerns the physicist, Murray Slotnick, who spent half a year completing a difficult calculation pertaining to the interaction between electrons and neutrons. He presented his calculations at a conference, but there was some discrepancy as to the veracity of his result. After Slotnick's presentation, Feynman returned home and spent the night calculating the relevant electron neutron processes using his diagrammatic approach. The next day he went to Slotnick and showed him the results he worked out the previous night. Slotnick's response, from constant retelling, has been all but etched in stone: "What do you mean you worked it out last night, it took me six months!" After inspecting Feynman's results in the appropriate limit, Slotnick found that they did indeed match his own. In quantum field theory, the indeterminacy central to any formulation of quantum physics is expressed through the multitude of ways a final set of particles can result from an initial set of particles. Feynman diagrams provide a remarkably simple way to understand this concept. Say we were studying the process of electron scattering by light. The initial particles are an electron and photon, and the final particles are also an electron and photon. The relevant Feynman diagram, representing the simplest way this scattering can occur is shown in Figure 2. We can make this process more complicated by including additional interactions, in the central connecting line.

In general, we can draw several diagrams of arbitrary complexity and from Feynman's path integral we know that each diagram contributes to the total probability amplitude for going from an initial to a final state. The catch is that for quantum electrodynamics, more complicated processes do not contribute as much as simple process because interactions are associated with a small constant. More interactions mean more of these small constants are multiplied and hence we have a smaller probability.

Symmetry and Unification

Symmetry is Beautiful

Classical symmetry has led to a unification of conservation laws. Quantum symmetry has led to a unification of interactions.

In 1918 Emmy Noether proved a theorem which soon immortalized her as one of the greatest mathematicians in history. Loosely stated, she proved that the continuous transformations which leave dynamical equations of a system unchanged correspond to physical conservation laws. In proving this connection between symmetry and conservation, Emmy Noether revealed mathematics' role not only as a tool to define physics but as something representative of physics itself. Conservation laws no longer had to be understood in terms of the existence of forces and special types of vector fields; for once, a physical principle was derivable not from a mathematical technique but from the properties of the mathematics alone.

Noether's Theorem states the now taken for granted requirement of conserved quantities for every symmetry of a quantity called the

Lagrangian (the Lagrangian, akin to Newton's Law, is a way of obtaining the Equations of Motion of a System). From Noether's Theorem, we find that the dynamical properties of mechanical systems are independent of the exact time we observe them. This time translation symmetry is related to energy conservation. You have seen a manifestation of this result whenever you solved a conservation of energy problem: By equating the energy at two different times, we get to ignore the specific time dependence of kinematic quantities such as velocity and position. Similarly, we find that if a system is independent of the initial position of its constituents then momentum is conserved. This is why initial position is never supplied (or needed) in collision and momentum conservation problems.

Jeffrey and Peter: Beautiful but Broken

While Emmy proved that the existence of mathematical symmetry has a conceptually beautiful association with physical conservation laws, Jeffrey Goldstone and Peter Higgs showed that the disappearance of an existing symmetry has even more profound physical consequences. MIT's own Jeffrey Goldstone—in a series of papers coauthored with other luminaries such as Yochiro Nambu, Steven Weinberg, and Abdus Salam—proved that for every broken continuous symmetry of a quantum field there is a particle with no mass. The specific meaning of broken is derivable from its literal English meaning; something which was, but is no longer, in working condition. But if a symmetry is perfectly fine and then somehow broken, what exactly breaks the symmetry? The cause has to do with the lowest energy states of a theory. More specifically if the lowest energy states are not invariant under the symmetry transformation, then the theory has a broken symmetry.

A simple analogy of this phenomenon involves having a penny represent the lowest energy state of a quantum field and the various points on the surface of a dish plate represent the range of possible vacuum states. We would like our "field theory" to be unchanged under certain transformation which we represent as rotations about the center of the plate. If we place the penny at the center of the plate then (barring the rotation of Abraham Lincoln's head) our vacuum is symmetric, it looks the same regardless of how we rotate it. However, if we place the penny at any point besides the center, then its position is no longer invariant under rotations and we say that this symmetry is broken. In the same way, a non symmetric ground state in a real quantum field theory represents a broken symmetry. The massless particles, so called Goldstone Bosons, have implications when we consider more complex theories.

If, in particular, we make the more stringent requirement that this broken symmetry occurs in a theory with particles such as the photon, then these photon-like particles acquire mass and the massless particle disappears. This effect named after Peter Higgs, one of its many independent discoverers, is called the Higgs Mechanism and defines our modern conception of nature. The builders of the Standard Model (physics' most widely accepted theory of particle interactions) revealed that this concept of broken symmetry was not a mere technicality but was a necessary consequence of nature. In our current understanding of particle physics, the interactions which define how particles decay and emit electromagnetic radiation are derivable from the same set of fields. These fields, before the symmetry is broken, are all related and are essentially equivalent in the same sense that a particular choice of axes for a three-dimensional coordinate system is essentially equivalent to

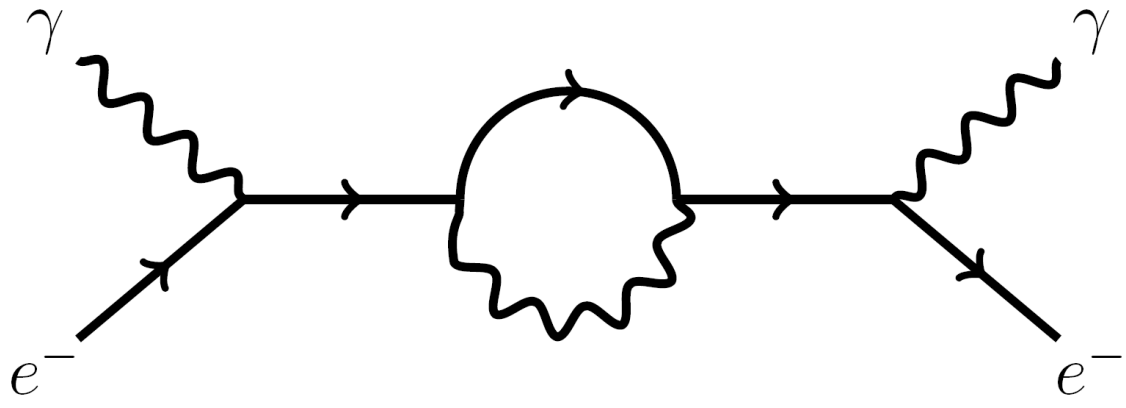


Figure 2: $e^- \gamma \to e^- \gamma$ with additional interactions

some permutation of those axes. When we break this symmetry by changing the properties of a few of these fields, then we obtain two sets of fields: those with mass and those without mass. The fields without mass correspond to photons and define electromagnetism. The fields with mass correspond to the W± and Z0 bosons and define how particles decay.

What's Next?

Modern theoretical physics has witnessed many revolutions, the most profound of which have been initiated by an appeal to the quantum nature of our world. The first unification occurred in our understanding of the relationship between fields and particles. The second unification occurred during the development of our understanding of symmetry and quantum physical phenomena. The last unification is occurring today and represents the most ambitious attempts to obtain the holy grail of quantum unification. After quantizing fields and particles physicists have appropriately asked, "What else is there to quantize?" The answer to this question has determined the direction of theoretical physics research past the accepted quantum field theory of the Standard Model.

Universe's Next Broken Symmetry

Supersymmetry: Give spacetime quantum dimensions

The standard coordinates of spacetime can be expanded by adding additional quantum dimensions. These dimensions are significantly distinct from our standard ones in that they do not multiply with each other in the traditional sense. Specifically these dimensions anticommute with one another, a fact that may seem like a technicality but actually provides a constraint and is an essential simplification of the crux of the theory. By requiring theories with these additional dimensions to be unchanged under certain symmetries (called supersymmetries), we obtain quantum field theories which allow fermions, which are spin particles like electrons, to transform into bosons, which are integral spin particles like photons, without affecting the theory. Consequently we require an equal number of bosons and fermions for this transformation to be self-consistent. But of course this is not what we observe in nature. The Standard Model, for example, has five types of bosons and 12 types of fermions, all of which do not exhibit supersymmetry. Therefore, we know that this supersymmetry is broken. The exact mechanism of supersymmetry breaking and the implications of various models that explain it have been the object of ongoing research in the past decades.

"One of these strange properties is the existence of a *quantum vacuum *from which particles can emerge and into which particles can disappear."

Einstein against the Quantum

Quantum Gravity: Quantize Spacetime Itself

Applying the now standard rules of quantization to Einstein's theory of space time has resulted in the field of quantum gravity. Once one has gone through the trouble to quantize every other known dynamical field in the universe, the idea to quantize the gravitational field of spacetime is not such a huge intuitive leap. However, the fact that quantum field theory, based solely on special relativity, considers spacetime to be flat meshes with the essential idea of curved spacetime in general relativity. So, at a basic level any incorporation of general relativity into the standard quantum field framework would require a modification of quantum field theory itself. Moreover, all of the accepted quantum field theories of today have a property called renormalizability which allows us to make consistent predictions of phenomena on our energy scale. All quantum versions of gravity do not have this property. These problems, together with the fact that gravity is much weaker than the other interactions and therefore difficult to test experimentally, gives quantum gravity the perception of a field from which progress will be slow and painful. Still, developments have been made in trying to understand a quantum version of gravity by engendering lofty models of nature.

The fact that Einstein's theory has been the most difficult classical theory to quantize is rather fitting considering how hard Einstein fought against the establishment of quantum mechanics as a field of physics. It's almost as though, even through the grave, Einstein is resisting accepting quantum physics as a legitimate discipline.

MURJ UROP Summaries

New Engineering of Beta Sheets[1]

Scott Jenkins[2]

1. Department of Chemical Engineering, MIT, 2010
2. Student Contributor, Class of 2013, Deparment of Chemical Engineering, MIT

Since the 1950s, biologists have known about proteins that form using beta sheets, a secondary structure. They are formed by the peptide binding of amino acids that create long strands which can form peptide sheets of either parallel or anti-parallel conformation. Studies have shown that these beta sheets will contort and change their shape and properties when exposed to different temperatures and pH levels [1]. However, the particular set of designer synthetic peptides examined here has not yet been tested to observe how the putative beta sheets will respond when exposed to different pressures. The influence of pressure on these peptides is the aim of this study, where we are using novel designer peptide strands and beta sheets (IKIK)x 4, (IEIK)x 4, (LELK)x 4, and (TVTK)x 4 to study the effects pressure has on peptide structure. The most important difference with the new beta sheets is not that they have never been used before, but the process we are trying to implement in creating and synthesizing them.

Currently the most commonly used method of beta sheet formation is to synthesize the beta sheets first and then perform tests to ensure that it has been synthesized correctly [3]. However, a civil engineer would never build a bridge over the river and test its strength afterwards. He would undergo a long design process to simulate stress at different points and ensure that the properties of the bridge are desirable. We are trying to achieve a similar process: design then creation. Using the molecular modeling program PyMol, in conjunction with a real life molecular simulator, we are modeling our beta sheets, testing if they have the desired requisite effects under pressure, and then bringing our results to be created to the laboratory.

I worked this summer to learn PyMol and used scripts written by others, such as the Build_Seq script written by Robert L. Campbell [4], to make the beta sheets. Using the Build_Seq script and my own knowledge of PyMol I was able to create strands of amino acids with the desired Phi (φ) and Psi (ψ) angles, and, using PyMol's GUI, I was able to align the strands in a beta sheet suitable for real life modeling. Figure 1 is (IKIK) x 4, which consists of 4 strands in parallel conformation. While the figure is very similar to a real model of a beta sheet, it lacks a side chain that attaches the strands on the ends. We plan to implement these side chains on an individual basis after further testing ensure the correct side chain is used for each sheet.

One of the next steps in our beta sheet formation process is to write a script that can be run in PyMol to create a beta sheet model, similar to Figure 2, through a series of prompts asking the user for the specifications of the desired beta sheet. These specifications will include the amino acid sequence, parallel or anti-parallel conformation, the Phi (φ) and Psi (ψ) angles, the number of strands, and the number of sheets. With this script, researchers working with beta sheets will have an easy way of modeling and visualizing at the molecular level.

While the creation of beta sheets is important, it is useless without real world applications. The main focus of our pressure dependent beta sheets is how they will help glaucoma patients. Glaucoma is caused by a discrepancy in the amount of fluid leaving the frontal sac of the eye and the larger inner sac which creates an imbalance causing pressure to rise in the frontal part of the eye eventually damaging it [1]. In glaucoma patients, these valves are not dysfunctional and a buildup of fluid causes higher pressure in the frontal sac of one's eye. What these peptides do is replace these schelm valves since the peptides will be regulated by pressure and act in the same way the valves do. The holes in the sheets change when pressure is applied, but further experiments still need to be conducted to calibrate the sheets such that someone with glaucoma will be able to regulate their own eye pressure. The holes I am referring to are the gaps between the peptide bonds and strands that vary according to pressure because the applied force contorts the beta sheet changing the size and shape of the gaps. Another potential benefit from this new method of protein engineering is helping neurodegenerative diseases, such as Alzheimer's, which are often caused by amyloids accumulating in certain regions of the brain and causing damage.

References:

1. Sommer, Alfred, James Tielsch, Joanne Katz, Harry Quigley, John Gottsch, Jonathan Javitt, and Kuldev Singh. "Relationship Between Intraocular Pressure and Primary Open Angle Glaucoma Among White and Black Americans: The Baltimore Eye Survey -- Sommer Et Al. 109 (8): 1090 -- Archives of Ophthalmology." Archives of Ophthalmology, a Monthly Peer-reviewed Medical Journal Published by AMA. Web. 13 Nov. 2010. <http://archopht.ama-assn.org/cgi/reprint/109/8/1090>.

2. Maren, T. H., Carbonic anhydrase: General perspective and advances in glaucoma research. Drug Development Research, 10: 255–276. doi: 10.1002/ddr.430100407

3. Liam J. McGuffin, Kevin Bryson, and David T. Jones: The PSIPRED protein structure prediction server. Bioinformatics (2000) 16(4): 404-405 doi:10.1093/bioinformatics/16.4.404

4. Campbell, Robert. "Build_seq." My Pymol Script Repsitory. Queens University, 17 May 2005. Web. 13 Nov. 2010. <http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/>

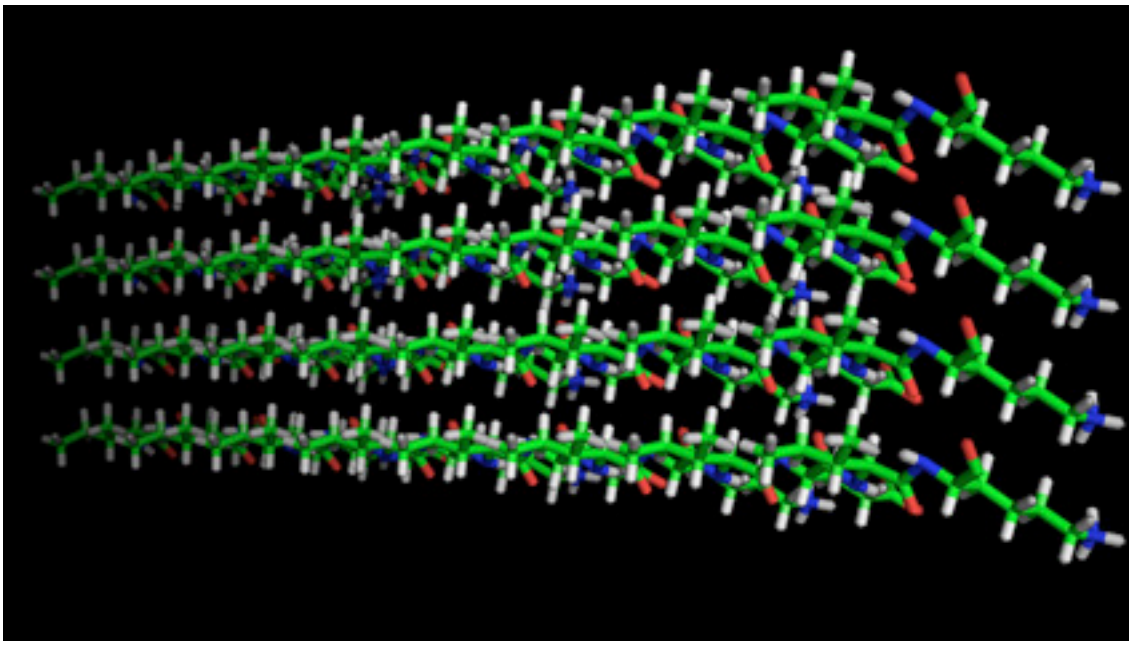


Figure 1. (IKK) x 4 beta sheet, consisting of 4 strands in parallel conformation.

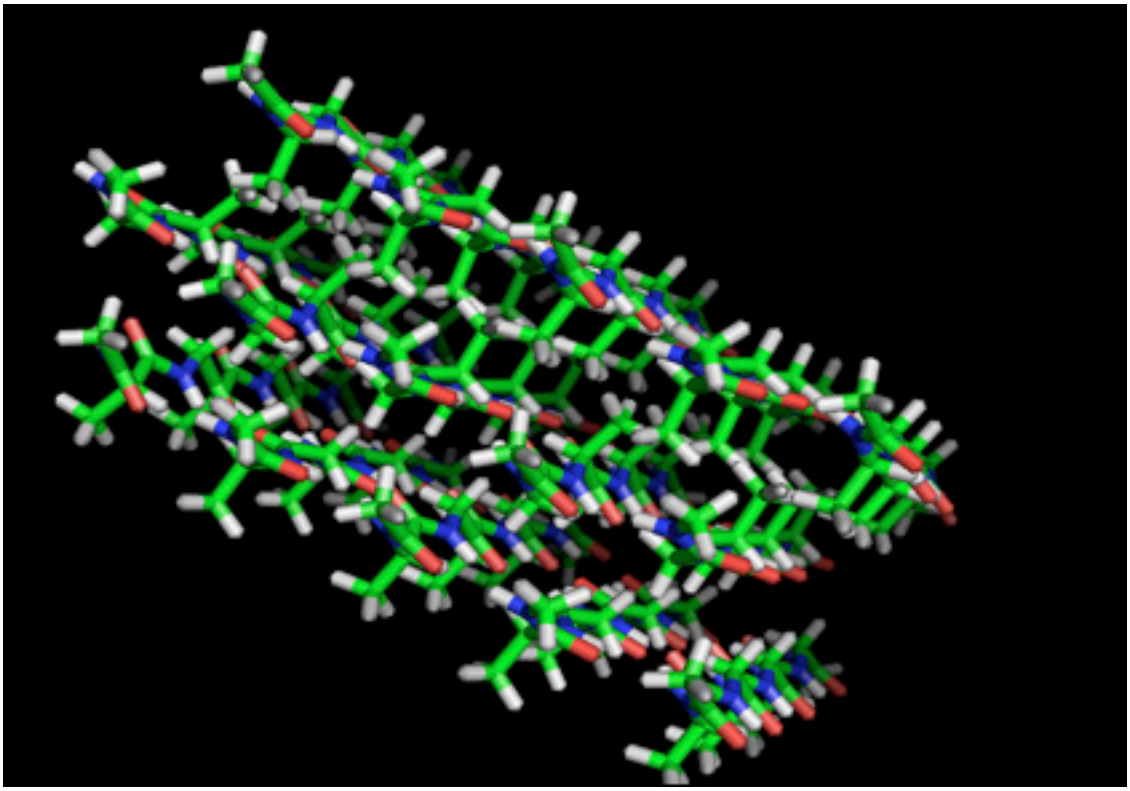


Figure 2. Beta sheet model created by script in PyMol.

Nonparametric Estimation of Gasoline Demand[1]

Jessie Li[2], Professor Whitney Newey[3]

1. Department of Economics, MIT, 2010
2. Student Contributor, Class of 2011, Deparment of Electrical Engineering and Computer Science, MIT
3. Principal Investigators, Department Economics, MIT

Nonparametric estimation is a class of techniques in econometrics for making inferences on data using flexible functional forms [1]. Examples of nonparametric estimation include estimating the probability distribution function for a set of observations and performing nonparametric regression. Nonparametric regression is a regression method in which the predicted dependent variable and the independent variables are related by an arbitrary functional form [3]. This is in contrast to ordinary least squares, in which the predicted value of the dependent variable is a linear combination of the independent variables. Equation 1 shows ordinary least squares model, and equation 2 shows nonparametric regression model where g(x) can be any function.

$$y = x\beta + \varepsilon \quad (1)$$

$$y = g(x) + \varepsilon \quad (2)$$

The current research applies different nonparametric regression models to estimate gasoline demand for US households. The first model used is a power series model where the quantity of gasoline demanded is a polynomial in price and income. A power series model is a good global approximation to smooth functions but it performs poorly on functions with jumps or kinks [3]. The second model used is the kernel estimator, which can be thought of as a weighted average of the observations, giving more weight to the values close to value to estimate and less weight to the values farther from the estimated value [3]. A drawback of the kernel estimator is that it produces biased estimates near the boundaries of the range of observations [3]. A third model used is the local polynomial regression model. This model can be thought of as fitting a polynomial on observations close to the value to estimate. Local polynomial regression does not suffer from the boundary bias problem of kernel regression because the polynomial automatically adapts to the boundary [3].

Once an estimate of the demand curve is obtained, it is used to estimate the welfare losses associated with a tax on gasoline. In economics, the deadweight loss is the loss in social welfare caused by deviations from competitive equilibrium price and quantity [5]. As one of the most widely used welfare loss measures in areas of economics such as public finance, deadweight loss estimates can help policy makers understand the ramifications of their tax policies on social welfare [2]. The deadweight loss is obtained from an estimate of the equivalent variation, which is the amount of money that a consumer would pay before a price increase in order to avoid the price increase [4]. Estimates of deadweight loss and equivalent variation depend on the income of the representative consumer. In general, individuals with higher income are willing to pay more to avoid a price increase and suffer greater welfare losses from price increases.

Estimates of deadweight loss and equivalent variation have already been obtained for the power series and kernel regression models. Currently, we are working on local polynomial regression. Future steps in the research including using new models for gasoline demand and applying the models on new datasets.

References:

1. Greene, William. Econometric Analysis. Upper Saddle River: Prentice Hall, 2008.

2. Hausman, Jerry, and Whitney Newey. "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss." Econometrica 63, no. 6 (November 1995): 1445-1476.

3. Newey, Whitney. "Nonparametric Estimation." November 2009.

4. Pathak, Parag. "Consumer Theory." September 2009.

5. Synder, Christopher, and Walter Nicholson. Microeconomic Theory: Basic Principles and Extensions. Thomson South-Western, 2008.

Student Achievement Benefits from Charter Schools[1]

Christian Perez[2], Parag Pathak[3], Christopher Walters[4], Joshua Angrist[4]

1. Lab, Department of Economics, MIT, 2010
2. Student Contributor, Class of 2012, Deparment of Economics, MIT
3. Principal Investigator, Department of Economics, MIT
4. Mentors, Department of Economics, MIT

Charter schools in the United States are publicly funded schools that are not subject to the strictures and regulations that are common to regular public schools. Charter school teachers' wages are not dictated by unions or state mandates; instead, these schools follow a charter drafted by the school's founders and ratified by the school's State. This allows charter schools to foster a unique culture of academic rigor, support, and discipline. Tuition is free for matriculating students.

Admission to charter schools is a controversial topic. Typically, charter schools conduct admissions using an unbiased lottery; each applicant, regardless of baseline test scores and demographics, receives one entry in the proverbial hat and waits to (hopefully) be chosen. Families whose children are rejected from a charter school must seek other alternatives. This can be especially devastating for families who've invested considerable resources to get their children into particular charter schools.

Such investment by all parties into charter schools leads to the question: do charter schools actually improve upon regular public schools? Defining improvement in terms of higher test scores, professors Angrist and Pathak have recently sought to answer this question. Throughout Spring, Summer, and Fall 2010, we've conducted several econometric studies using charter school data from charters in Boston and all of Massachusetts. Data on demographics, Massachusetts Comprehensive Assessment System (MCAS) test scores, and other student characteristics come from two state-wide data sets. Lottery records were collected from each charter school.

Principal among our econometric models are our Two-Stage Least Squares (2SLS) models, which we use to estimate the benefit (in terms of test scores) of a year in a charter school. We use the randomized nature of charter school admissions to create instruments for our 2SLS models. Estimates from the Boston study show that middle school students in charter schools gain .2 standard deviations on the ELA section and .4 standard deviations on the math section of the MCAS test per each year in a charter school; high school students gain .2 standard deviations on the ELA section and .3 standard deviations on the math section of the MCAS per each year in a charter school [2]. These positive test score results are consistent with other studies using similar designs in New York City and Chicago [3,4].

These findings suggest that Boston charter schools outperform their public school counterparts. For more robust results, we plan to expand our framework to include all charter schools in Massachusetts. We hope to contribute a quantitative analysis of the effectiveness of charters to the ongoing national charter schools debate and better inform education policy makers.

References:

1. Waiting for Superman. Dir. Davis Guggenheim. Writ. Davis Guggenheim, Billy Kimball. Paramount Pictures, 2010.

2. Abdulkadiroglu, Atila, Joshua Angrist, Susan Dynarski, Thomas J. Kane, and Parag Pathak (2010), "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots", Quarterly Journal of Economics, forthcoming.

3. Hoxby, C. M. and Murarka, S. (2009). "Charter Schools in New York City: Who Enrolls and How They Affect Student Achievement." NBER Working Paper 14852.

4. Hoxby, C. M. and Rockoff, J. (2005). "The Impact of Charter Schools on Student Achievement." Harvard Institute of Economic Research Working Paper Series.

MURJ Reports

Assessing the Fairness of the Primary Process[1]

Svetlana Chekmasova[2], Tatyana Deryugina[3], Professor Michael Greenstone[3], Emily Breza[3]

1. Department of Economics, MIT, May 2010
2. Student Contributor, Class of 2011, Deparment of Economics, MIT
3. Mentors, Department of Economics, MIT

Sequential voting has previously been studied theoretically and in a laboratory setting. According to the theory, early votes may be fundamental in shaping the cumulative outcome of the election as later voters may be influenced by the election outcomes which are revealed before they vote. I look for evidence of this in data from US Republican primaries in the 2000 election cycle; in particular, at the immediate effect the release of election results has on voters' preferences. This is separate from other sources which may influence voters' preferences, such as discussion and analysis of the revealed outcomes. I find that in general, voters do not immediately update their preferences following the release of election results, justifying the fairness of the sequential voting process. However, when the results for several states are revealed at once and are in favor of one candidate (Super Tuesday), voters do appear to respond immediately.

Introduction

Collective decision-making is of great importance to any democratically run community; even more so when it is a country. Every four years, the United States holds a general presidential election. Leading up to this election are a number of smaller elections – primaries – which help decide the presidential candidates for each political party. While the next president is chosen on Election Day in a country-wide election, these primaries are a sequential form of voting, as different states hold the primaries in a pre-determined order. Considering the importance of the outcomes of all these elections, it is of interest to gauge how effective these methods are at translating public opinion into the results of the election.

There has previously been some debate as to whether sequential voting is as "fair" to all involved as simultaneous voting. That is, do earlier votes count for more, since later voters already see part of the election results? Do early voters thus in some way shape the election? How do earlier voters' decisions affect later voters? Is it possible for an information cascade to develop, where later voters don't vote for their ideal candidate after observing earlier voters' choices, so their vote does not reveal any (or a limited amount of) additional information about the perceived quality of the candidates? Earlier voters may have more say in initially shaping the progress of the election, but later voters are better informed about the candidates when they do vote, which is also in some sense an advantage.

I am interested in looking for evidence of any of the above in the particular setting of United States primary elections. I would like to see if the sequential nature of the primaries affects the way people vote. There have been a number of theoretical papers written on the subject [1, 5]. There have also been laboratory experiments conducted which involved voting [4]. However, there doesn't appear to have been much application to natural experiments, so I would like to see if there is any evidence of the theory in a real world setting.

To simulate the results of the election had it been simultaneous, I would like to model the outcome based on initial preferences. Before any primaries take place, I assume people in different states have some initial impressions of the various candidates which dictate their preferences. Aggregating the individual preferences in each state, it would be possible to predict the outcome of the primary election in the state if we assume that individuals vote based on their preferences. This appears to be a reasonable assumption, since in a simultaneous election individuals only have their own preferences based on which to make a decision, and a rational human being would vote for the candidate he likes best.

It would be of interest to contrast the simple prediction above with the actual outcome of the sequential elections. Ostensibly, the outcome of the first state which votes should be predictable from the preferences its population expressed right before the election, since they receive no new information based on which to change their preferences. It is not as simple with the following states, as there are two sources of information which may affect people's preferences. The first is new information which is created between the time the result of the first primary comes out and the second primary takes place. This can be thought of as public appearances by candidates, political ads, any scandals which might come out, endorsements of a candidate by various politically significant figures, and more. These are all confounding variables to the second source of information, whose effect I am primarily interested in gauging – the outcome of the first primary. Once people in the second state see that candidate X won in the first state, does it affect their preferences for the different candidates and their voting patterns? Do opinions change with every new result that comes out? Does the population of the last state to vote update preferences following each prior election?

Since I am interested in separating the preference-changing effects of new information and the results of previous primary elections, I rely on the fact that new information is created over time. Thus, I should be able to isolate the effect of one state's

primary results by observing voters' preferences in other states immediately before and following the election.

One might say that preferences in the following states may not change instantaneously following the outcome of one state's election, but the results of an election are most widely publicized immediately following the election, so prospective voters are most likely to acquire this new information then. I can also safely assume that if the outcome of the previous election will affect preferences, it will do so immediately, as voters re-optimize. Voters may also continue to re-optimize following this initial influx of information, but this will be as a result of new information that enters the system (for example, conversation with others) and so is exactly the type of thing I wish to disregard here.

Literature Review

The original idea concerning herd behavior (information cascades) was proposed by Abhijit Banerjee [1]. He describes a simple model in which agents sequentially choose between two alternatives based on a private signal and the observed actions of previous agents. In this situation, one alternative is better than the other, but the agents do not know which. It is possible for an information cascade to develop, if enough agents select one of the alternatives. Then the rest of the agents would find it advantageous to follow suit, disregarding their private signal in the process, and create an information cascade. This is, of course, simpler than a voting setting, which has more than two choices and a private signal is less well-defined. However, it is useful to consider the idea in a qualitative sense to develop the notion of the "worth" of a vote, how much weight a vote has in the final outcome of the election.

The idea which inspired the current research project is a working paper by Knight and Schiff, Momentum and Social Learning in Presidential Primaries [3]. They develop a theoretical model in which agents act to maximize their utility as they update their beliefs concerning different candidates in a Bayesian way based on previous voters' actions. They use a dataset very similar to the one I am using (2004 rather than 2000), and, under some fairly strong assumptions, they find that earlier votes are much more influential than later votes.

My approach is different from theirs in that I do not attempt to quantitatively estimate the relative weight of individual voters at different times in the election cycle. Instead, I try to see whether the election itself has any effect – does the fact that previous voters indicated their opinion induce later voters to update their preferences?

Others have considered sequential voting as a theoretical topic (see the works cited page), but I am interested in looking specifically at real-world data and seeing to what conclusions it leads me.

Data

I am using data from the 2000 National Annenberg Election Survey from the Annenberg Public Policy Center at the University of Pennsylvania. This is an extensive opinion poll which surveyed individuals on a daily basis from December 1999, right before the primary season, until January 2001, the inauguration of President Bush. I am specifically focusing on several data sets which polled respondents nationally and in a number of key states during the primary season. Several hundred adults were interviewed daily by telephone and queried about their preferences for various candidates, their political involvement, their exposure to the political campaigns, and numerous other variables. I am particularly interested in looking at the Republican primary process. During that election cycle, there were not many contenders for the Democratic Party presidential nomination, and it was clear from very early on that Al Gore would be the Democratic nominee. The outcome of the Republican primaries was less clear, and the progression of candidates' popularity over time is much more interesting to observe. I choose to focus on George Bush and John McCain, the leading Republican candidates competing for the nomination.

Respondents in a specific state were contacted only prior to the occurrence of the primary in their state. Thus, if Iowa had its primary on January 24, then in my dataset Iowans were only contacted and interviewed until January 24. For every state, I drop any observations where the respondent was contacted on the day of a primary, since I cannot know if the respondent is or is not aware of the outcome of the primary. I assume that knowledge in any days following the primary. I also drop any interviews conducted in a state one or two days immediately preceding an election in that state to eliminate any potential bias stemming from the fact that the respondent's state is about to vote.

Contact phone numbers were selected randomly, following a certain procedure, and a special protocol was implemented to maximize the representativeness of the sample of respondents interviewed on any single day. More information concerning the data set can be found in its documentation, available upon request.

I. Effects of Win/Loss on Intent to Vote for a Candidate

	Bush (intended vote)	**McCain** (intended vote)
Effect of primary (win)	.0909048	.1692576
	(.0871577)	(.1612918)
	$z = 1.04$	$z = 1.05$
Effect of primary (loss)	-.0905237	-.1818898
	(.15884)	(.0904043)
	$z = -.57$	$z = -2.01$

Note: The top number is the coefficient and the number in parentheses is the robust standard error.

II. Effects of Win/Loss on Favorability of a Candidate

	Bush (favorability)	McCain (favorability)
Effect of primary (win)	1.116413	.1916861
	(1.103658)	(1.681811)
	z = 1.01	z = .11
Effect of primary (loss)	-1.524001	-1.217809
	(1.855784)	(1.022749)
	z = -.82	z = -1.19

The data sets which I use in my analysis are national and state-specific cross-section studies. The states on whose Republican primaries I focus are Iowa (Jan. 24), New Hampshire (Feb.1), South Carolina (Feb.19), Michigan (Feb.22), Super Tuesday (March 7; CA, CT, GA, ME, MD, MA, MO, NY, OH, RI, and VT), and Second Tuesday (March 14; FL, LA, MS, OK, TN, TX). No person was interviewed twice, so each interview is one observation. Each survey contains the same set of questions, so has the same variables. The ones of particular interest to me are individuals' ratings of various candidates' favorability and the candidate for whom the respondent intended to vote for GOP presidential nominee. The former is a rating from 0 to 100 of how favorable the respondent considers the candidate in question. The latter is the name of some particular candidate whom the respondent prefers over others. The time period which my data covers is December 14, 1999 to April 3, 2000. In total, I have approximately 23,000 observations.

For the Republican primary results and dates, I looked at numbers reported by George Washington University [6].

Econometric Model

My intent is to assess the effect that the occurrence of a primary has on public opinion. I attempted to qualify the effect by observing changes in intent to vote for a particular candidate and reported favorability. Since George Bush and John McCain were the main contenders (no other candidate won a primary), I looked at changes in their popularity over time.

I began simplistically by observing daily average favorability ratings around the time of each primary. This gave me approximately 100-200 observations per day, which is not enough to give more than a very rough idea of the situation.

To add more rigor to my analysis, I proceeded to look at the effect an election and its outcome has on respondents' intent to vote for a candidate. I began by creating dummies *intended_bush* and *intended_mccain* which equal 1 if the respondent indicated he plans to vote for Bush or McCain, and 0 otherwise. The motivation for this is to isolate the immediate effect of a win on a candidate's popularity. To tag a candidate's wins, I again applied dummy variables. When the respondent was interviewed within two days of an election in which Bush won, *bush_own_win* equals 1; otherwise, it is set to 0. Similarly, *mccain_own_win* equals 1 for the dates when McCain was the victor. There are two exceptions. Because the South Carolina and Michigan primaries are so close together and won by different candidates, I set *bush_own_win* = 1 for only one day following the South Carolina primary. Similarly, *mccain_own_win* only equals 1 for one day preceding the Michigan primary.

The assignment of purely a 0 or a 1 may be criticized as arbitrary, as the margin by which a candidate wins can vary. However, I assume that generally it is the winner of an election which is most memorable to the general population, not the margin, and consequently most relevant in shaping opinions. This approach is problematic in one respect – when several primaries occur in one day, such as for Super Tuesday or Second Tuesday, it is difficult to assign purely a 1 or a 0 to a candidate without making some assumptions. On Second Tuesday, Bush won all six states, so I assigned *bush_own_win* = 1 for that range of dates. On Super Tuesday, however, 11 states held primary elections, and Bush won 7 of these. I made the assumption that he "won" Super Tuesday and assigned him a 1 and McCain a 0 for that range of dates. This made for a crude analysis, but it allowed me to apply the model uniformly to each time period when there is an election. In this framework, there isn't a good way to distinguish between who won which state. In further analysis, I made an adjustment by including the margin by which one candidate won, which will be described below.

To isolate the respondents who were polled in the two days following an election, I assigned a dummy variable *post_primary* = 1 to those observations, 0 otherwise. To account for any differences among respondents which stem purely from their background, I included some demographics dummies in my analysis. Because my dependent variable is categorical, I used a probit regression with robust standard errors. The value of interest is β12, the coefficient of the interaction term between Bush winning an election and a person being polled right after. This coefficient shows the effect of the knowledge of the outcome of the election on respondents' intent to vote for him, taking out any effects of his overall level of popularity.

intended_bush =β0 + β1(sex) + β2(age) + β3(race) + β4(hispanic origin)+ β5(education) + β6(marital status) + β7(employment status) + β8(census division) + β9(urbanity) + β10(household income) + β11**(bush_own_win)** + β12**(bush_own_win * post_primary)**

I also ran a similar regression to see the immediate effect winning an election has on respondents' intent to vote for McCain, as well as two more regressions to see the effects of losing. For the latter two, I regressed *intended_mccain* on *bush_own_win* and *bush_own_win * post_primary*, and the opposite for *intended_bush*.

III. Effects of Win on Intent to Vote for a Candidate, by State

	Bush (intended vote)	McCain (intended vote)
Iowa (B)	-.1417407 (.0997019) z = -1.42	-
New Hampshire (Mc)	-	.0100199 (.1338652) z = .07
South Carolina (B)	-.201577 (.1569091) z = -1.28	-
Michigan (Mc)	-	.206544 (.1331112) z = 1.55
Super Tuesday (B)	.552961 (.1366952) z = **4.05**	-
Second Tuesday (B)	.9158533 (.5706559) z = 1.60	-

IV. Effects of Loss on Intent to Vote for a Candidate, by State

	Bush (intended vote)	McCain (intended vote)
Iowa (B)	-	-.1410688 (.102956) z = -1.37
New Hampshire (Mc)	-.0881798 (.132682) z = -.66	-
South Carolina (B)	-	.3431556 (.1546527) z = **2.22**
Michigan (Mc)	-.1449537 (.1339704) z = -1.08	-
Super Tuesday (B)	-	-.3960185 (.1423395) z = **-2.78**
Second Tuesday (B)	-	-.6242839 (.5887452) z = -1.06

V. Effects of Win/Loss on Favorability of a Candidate, by State

	Bush (favorability)	**McCain** (favorability)
Iowa (B)	.4716279 (1.298356) z = .36	1.748297 (1.206337) z = 1.45
New Hampshire (Mc)	1.5907 (1.72562) z = .92	2.133186 (1.616894) z = 1.32
South Carolina (B)	-1.689984 (2.055901) z = -.82	-.4127895 (1.967261) z = -.21
Michigan (Mc)	-2.764944 (1.766732) z = -1.57	-.3326053 (1.631713) z = -.20
Super Tuesday (B)	2.812115 (1.588571) z = 1.77	-3.762323 (1.433497) z = **-2.62**
Second Tuesday (B)	-2.143428 (3.252185) z = -.66	-1.795826 (2.786884) z = -.64

VI. Effect of Own Win/Loss and Other's Win/Loss on Own performance, by State

Intended Vote	**Bush** (as function of own performance)	**Bush** (as function of McCain's performance)	**McCain** (as function of own performance)	**McCain** (as function of Bush's performance)
Iowa (B)	Not sig.	Not sig.	Not sig.	-.4986482 (.2347376) z = -2.12
New Hampshire (Mc)	Not sig.	Not sig.	Not sig.	Not sig.
South Carolina (B)	Not sig.	Not sig.	Not sig.	Not sig.
Michigan (Mc)	Not sig.	Not sig.	Not sig.	.6722295 (.2688381) z = 2.50
Super Tuesday (B)	1.577578 (.3603004) z = 4.38	2.852478 (.6128403) z = 4.65	-2.063908 (.6384561) z = -3.23	-1.181744 (.3752937) z = -3.15
Second Tuesday (B)	Not sig.	Not sig.	Not sig.	Not sig.

Besides intent to vote for a candidate, I was also interested in the perceived favorability of a candidate. I ran a similar regression for the reported favorability ratings for each candidate, this time using ordinary least squares (with robust standard errors), as the favorability ratings are continuous and range from 0 to 100.

favorability_bush = β0 + β1(sex) + β2(age) + β3(race) + β4(hispanic origin)+ β5(education) + β6(marital status) + β7(employment status) + β8(census division) + β9(urbanity) + β10(household income) + β11**(bush_own_win)** + β12**(bush_own_win * post_primary)**

The above analysis groups all the primaries together. Since they occur sequentially, it is of interest to see the effects of each primary individually. I created a new set of dummy variables which equal 1 if the respondent was polled within two days following a specific primary. For instance, *IA12* = 1 if the respondent was interviewed on January 25 or 26, following the Iowa primary, and 0 otherwise. The only exception to this is the variable for South Carolina. Here, I only looked at one day following the primary, for reasons described above. Again, I interacted *bush_own_win* and *mccain_own_win* with this new set of dummy variables and ran a separate regression for each state or set of states. For intent to vote, I used a probit regression, and for favorability ratings, I used ordinary least squares. The coefficient of interest is once again that of the interaction term, which shows the effects of the knowledge of the outcome of the election. These regressions have the following form:

intended_bush =β0 + β1(sex) + β2(age) + β3(race) +β4(hispanic origin)+ β5(education) + β6(marital status) + β7(employment status) + β8(census division) +β9(urbanity) + β10(household income) + β11**(bush_own_win)** + β12**(bush_own_win * IA12)**

As mentioned above, I was not entirely satisfied with my binary approach to winning and losing primaries, especially when there were several occurring in the same time period. Instead of simply assigning a 1 or a 0 for a win or a loss in a state, I instead reassigned the values as the percent of the votes each candidate received. On dates with multiple primaries occurring simultaneously, I used the average percent of the vote received in all the states. This seemed logical, as each candidate's popularity was roughly equivalent across states at a particular time. I went on to rerun the regressions with the new values for *bush_own_win* and *mccain_own_win*.

Results

I had been hoping to see a noticeable change between before and after a primary in candidates' favorability ratings or in percent of people who intended to vote for them. When I looked at daily averages, for both candidates the favorability ratings hovered around upper 50s and lower 60s, with a standard deviation between 20 and 30. Since only a few hundred people were polled daily, the large standard deviation is to be expected, unless there had been a very large effect. I encountered a similar problem looking at how many people stated they intended to vote for each candidate. The simplest approach thus did not provide any insight.

When I regressed the intent to vote for either Bush or McCain on their winning primaries and being polled immediately afterwards, I found no significant effects for either candidate.

This led me to conclude that the primary itself had no effect on peoples' intent to vote for a candidate when the candidate won the election. Of course, it is possible (and highly probable) that with time, the fact that a candidate won a given election changed people's opinions of the candidate one way or another. However, this would be the effect of other influences, not purely the fact that an election occurred and a particular candidate won. For instance, in the weeks following an election, it is discussed on the news, on various sites on the Internet, among friends and family, and the political campaign issues more propaganda. All these factors are new information entering the system, as described earlier. I am interested in observing whether the fact that some segment of the population has voted prior to another segment has any influence on the latter's votes, and the first set of regressions indicates that there is no effect for the winning candidate.

For McCain, it appears that an election in which he lost had a statistically significant negative effect on people's intent to vote for him. For Bush, however, it does not seem that the election had any effect when he lost.

Looking at favorability ratings for both candidates, the election itself again does not seem to produce any effect, whether a candidate won or lost.

For the individual primaries, Bush winning Iowa, South Carolina, and the states on Second Tuesday had no immediate significant effect on intent to vote for him. However, winning on Super Tuesday did have an immediate statistically significant positive effect. McCain's wins in New Hampshire and in Michigan had no immediate significant effect on people's intent to vote for him. The winner of the primary is listed in parentheses next to the state.

Looking at the losing candidate in each election, Bush had no effect on intent to vote for him as a result of losing New Hampshire and Michigan.

Losing Iowa and Second Tuesday was not significant for McCain. However, surprisingly, losing South Carolina did have a significant *positive* effect on intent to vote for McCain. This may have contributed to McCain winning in Michigan a few days later, since there wasn't much time for new information to enter the system between the two elections. The model does not provide intuition as to why losing Carolina would have a positive effect; I attribute this to outside factors, such as the voters' knowledge and sentiments prior to the election.

Neither winning nor losing elections had any immediate statistically significant effect on Bush's reported favorability. For McCain, there was a statistically significant negative impact of the election on his favorability following his performance on Super Tuesday. This may in part be exaggerated by my treatment of wins and losses on Super Tuesday,

Looking at the regressions with percentages of votes gathered rather than a binary win or loss, I saw some significant results. For

VII. Effect of Own Win/Loss and Other's Win/Loss on Own performance, by State

Favorability	Bush (as function of own performance)	Bush (as function of McCain's performance)	McCain (as function of own performance)	McCain (as function of Bush's performance)
Iowa (B)	Not sig.	55.14175 (24.60238) z = 2.24	Not sig.	Not sig.
New Hampshire (Mc)	Not sig.	Not sig.	Not sig.	Not sig.
South Carolina (B)	Not sig.	Not sig.	Not sig.	Not sig.
Michigan (Mc)	-11.49562 (3.629274) z = -3.17	-7.931756 (3.299564) z = -2.40	Not sig.	Not sig.
Super Tuesday (B)	11.19014 (4.150877) z = 2.70	22.94545 (7.075031) z = 3.24	-18.71268 (6.358061) z = -2.94	-10.46883 (3.729855) z = -2.81
Second Tuesday (B)	Not sig.	Not sig.	Not sig.	Not sig.

the first two regressions, where all primaries are aggregated, I saw no significance in the occurrence of an election, i.e. the coefficient of the interaction term (winning margin * being polled right after a primary) was not significant. For individual states, I show the significant results for intended vote in theTable V1.

These results confirm the previous findings. The Super Tuesday election seems to have had an immediate significant effect both on the respondents' intent to vote for Bush and for McCain, in Bush's favor. It also appears that the Iowa election may have had a negative impact on voters' intent to vote for McCain, and the Michigan election may have had a positive impact. See Table VII for the candidates' reported favorability.

In this scenario, the importance of Super Tuesday for Bush's popularity is again underscored. These results indicate that the elections boosted Bush's favorability and diminished McCain's. It also seems that the Michigan election diminished Bush's favorability, and the win in Iowa may have increased it

Conclusion

The overall effect on other states' voting populations of a primary occurring in a specific state appears to be minimal. Finding out that a candidate won or lost a particular election or set of elections does not generally seem to affect either peoples' perceptions of the candidate's favorability or their intentions to vote for one candidate or another. According to this model, it then appears that if the system were in an information vacuum – that is, candidates were not allowed to campaign following the beginning of the primary season, and voters had no interaction with each other or any other possible access to new information aside from the outcomes of the elections – then agents would vote exactly the same way as they would if the primaries were held simultaneously. This is an interesting result which justifies the fairness of the sequential nature of the primaries. Of course, this result is not impervious to critique, and it cannot be generalized to the real world as such. In reality, candidates certainly continue to campaign throughout the primary season and, following that, the months leading up to the presidential election. People watch the news, read articles online, discuss the election process with each other, and otherwise introduce a whole host of convoluting variables which very likely affect the outcomes of the primary process.

It would be interesting to observe the effects of any new information which becomes available to voters between the time of one primary and another. We cannot assume that all voters will update in the same manner following this new information, as voters are not homogeneous. Different states may have different preferences inherent to that state. For instance, one of the candidates might be from state A, and thus voters in state A may automatically have a bias for that candidate.

While this access to new information is not inherently a critique of sequential voting, it is a consideration. The voting population does not have equal access to information regarding the candidates, which puts the early voters at a disadvantage. However, by the act of voting, the early voters indirectly become a source of new information for later voters, which gives them a sort of first-mover advantage. The indirect effect would be in that the primary would fuel society's reaction to the outcome of the election – news reports, candidate campaign strategies, etc. Thus, later voters may appear to have less of a say in the final outcome, but they are more informed, and their opinions and votes may become crucial if some controversial new information comes out concerning a particular candidate. Previous voters cannot change their decision, but later voters are still in a position to do so, and all votes are officially weighed the same (only the outcome matters, not the process which generated the votes).

In my analysis, I did observe some significance in the outcomes of primaries, particularly just following Super Tuesday. Based on my analysis, it appears that finding out the results of those primaries caused voters in states which had not voted to immediately update their intent to vote and perceived favorability in favor of Bush. This is evident both from the statistically significant positive effect for

Bush and the statistically significant negative effect for McCain. Although I expect the results to be exaggerated to some degree in the first case, since I essentially assumed Bush won all of the states during Super Tuesday and McCain lost all the states, it is still interesting to observe such an effect. The second step of the analysis which utilizes percents of votes gathered, mitigating the relative win or loss, confirms the initial results. This implies that when a win or loss is large enough (numerous states), voters react to its implications. I hypothesize that perhaps after hearing the result for one state, voters in another state are fairly indifferent as to how they should react, and wait for further information to come into the system before making up their minds. This is consistent with most of what we observe in the data for Iowa, New Hampshire, etc. Once the television or the internet offers an interpretation for the outcome of a primary, voters may update their preferences. However, when the outcome is overwhelmingly in favor of one candidate, such as Bush during the Super Tuesday primaries, voters do not require this additional information to update their intentions to vote for various candidates. Of course, additional information may cause them to update their preferences further.

An alternative hypothesis is that the effect is cumulative, and after Super Tuesday Bush had amassed the relevant number of wins, whatever that may be. In this case, wins or losses in the initial few states would not bias later voters one way or another, but once a candidate gains enough momentum, agents who have not yet voted may assume he won and be more likely to vote for him. This is reminiscent of the information cascade sentiment which Banerjee proposed. However, that would need to be tested further, as the only conclusions one can draw from this paper is that after a certain point, during a large election (if one generalizes Super Tuesday to be such), agents who have not yet voted do update their preferences essentially simultaneously, as a result of the election.

Although my approach is relatively simplistic, it does offer some interesting insights. Further research would need to be conducted before any definitive conclusions can be drawn, however. Some possible extensions could be to look at a sequential election in a scenario where information spreads less quickly, or to consider a model which handles a candidate winning or losing a state in a more concrete way.

References:

1. Banerjee, Abhijit. A Simple Model of Herd Behavior. The Quarterly Journal of Economics, Vol. 107, No. 3 (Aug., 1992), pp. 797-817

2. Bikhchandani, Sushil & Hirshleifer, David & Welch, Ivo, 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades," Journal of Political Economy, University of Chicago Press, vol. 100(5), pages 992-1026, October.

3. Knight, Brian G. and Schiff, Nathan, Momentum and Social Learning in Presidential Primaries (November 2007). NBER Working Paper No. W13637. Available at SSRN: http://ssrn.com/abstract=1033762

4. Rebecca B. Morton and Kenneth C. Williams, Information Asymmetries and Simultaneous versus Sequential Voting. The American Political Science Review, Vol. 93, No. 1 (Mar., 1999), pp. 51-67

5. S. Nageeb Ali and Navin Kartik, Social Learning in Elections. 2007. Availlable at Columbia University: http://www.columbia.edu/~ktr2102/papers/n_kartik1.pdf

6. Data on election outcomes from George Washington University:

http://www.gwu.edu/~action/primrepsum.html.

Effects of Age and Income on Individual Health Insurance Premiums[1]

Jessie Li[2], Professor Jerry Hausman[3], Alp Simsek[3]

1. Department of Economics, MIT, 2010
2. Student Contributor, Class of 2011, Deparment of Electrical Engineering and Computer Science, MIT
3. Mentors, Department of Economics, MIT

This paper examines the effect of an individual's age and income on the premiums that she pays in the individual (non-employer sponsored) health insurance market. After controlling for medical conditions, insurance plan type, and demographic characteristics, it was found using both OLS and 2SLS that older and wealthier people pay higher premiums than younger and less wealthy individuals, keeping demographics and medical conditions constant. Furthermore, results from nested logit regressions on people's choices of insurance plans indicate that wealthier people actually prefer the more expensive managed care plans over traditional (fee-for-service) plans. However, older people favored the less expensive traditional care plans over managed care plans. These results indicate the possible existence of adverse selection effects based on age in the individual insurance market. Such results may have important policy implications for greater regulation in the individual health insurance industry.

Introduction

Over the past decade, health insurance premiums have been on the rise, and the percentage of uninsured individuals has increased [3]. Furthermore, with the aging of the baby-boomer generation, the demand for health services has increased, which brings up the question of whether the higher average age of the US population would increase average health premiums charged by insurance companies [4]. Considering the recent debate over health care reform, it would also be interesting to examine whether there is any correlation between a person's income and the premiums that she pays. In an attempt to answer some of these questions, this paper examines the effects of age and income on out-of-pocket premiums in the individual (non-employer sponsored) health insurance market, while controlling for medical conditions, demographic characteristics, and the insurance plan type. The hypothesis is that older and wealthier people pay higher premiums, keeping other factors constant.

Background

Health care in America can be divided into roughly two categories. One category is called traditional care (or fee-for-service), in which prices for health care rendered are determined solely between the physician and the patient, and the role of the insurance company is simply to reimburse the physician for treatments of acute illnesses or injuries. On the contrary, in managed care, the individual pays a monthly premium to group of health providers in a managed care organization (MCO). In return, the MCO will negotiate deals with a network of physicians to provide care at a lower price to the members of the organization, but with certain restrictions on the number of times that the members can see the physicians. The MCO provides coverage for primary and preventive care in addition to acute illnesses and injuries [8].

The managed care model provides more cost control than the traditional model by limiting the number of unnecessary medical procedures prescribed by the physician. In the traditional care model, physicians have a financial incentive to prescribe unnecessary medical procedures, leading to rising health care costs for the patient [8]. In the managed care model, physicians need to maintain a good reputation in order to remain in the network, and they are held accountable for their actions. When they enter into an agreement with the MCO, they promise to reduce the per unit cost of services for members in the MCO and to limit the number of unnecessary medical procedures. In return, the physicians usually receive compensation in the form of capitation, in which the MCO prepays the physician for the treatment of the patient. This compensation reduces the physician's incentive to prescribe unnecessary procedures.

Another way that managed care provides cost control is by limiting the types of physicians that patients can visit at the cheaper price. In the traditional care model, patients can visit any physician they want, and as long as their illness qualifies as an "acute illness," the insurance company would have to provide coverage. On the contrary, the managed care model restricts the physicians that patients can visit to those physicians in the network, if the patients would like the cheaper price. Doing so allows the MCO to keep costs down since the MCO had already negotiated a lower price with the physicians in the network. Furthermore, in some managed care plans, members must select a primary care provider (PCP) and always see the PCP first before seeing any specialist. The rationale for selecting a PCP is that the PCP would first determine whether it is necessary for the patient

to see the more expensive specialist. If the PCP can handle the case himself, the MCO would save on medical costs.

Types of Managed Care Organizations

There are three main types of managed care organizations: Health Management Organization (HMO), Preferred Provider Organization (PPO), and Point of Service (POS). HMO is a type of MCO which offers health care services to its members for a fixed monthly fee. HMOs offer two types of managed care plans, closed-panel HMO and open-access HMO. Under the closed-panel HMO plan, patients are required to select a primary care physician who must authorize the patient before he can receive any specialist treatment. Patients pay a minimal copayment for office visits, and the HMO covers hospitalizations in full [8]. In the open-access HMO plan, patients have the option of seeing a specialist directly for a higher copayment. In both closed-panel and open-access HMO plans, the HMO does not provide coverage for visits to physicians outside the network.

PPOs are networks of doctors and health professionals that provide health services to individuals for a negotiated fee [8]. In contrast to HMOs, PPOs do not restrict patients to seeing their primary care provider first; patients can visit specialists without having to pay a higher copayment. Furthermore, PPOs provide coverage to individuals seeking care from physicians not enrolled in the network, although patients have to pay a high deductible and a high coinsurance.

A POS plan combines features of the HMO and the PPO plans. As in the HMO plan, the patient is required to select a PCP who must authorize any visits to specialists. Like the PPO plan, the POS plan provides out-of-network coverage at a high deductible and coinsurance rate.

Based on a study conducted by the Henry J. Kaiser Family Foundation and the Health Research and Educational Trust (HRET) in 2006, premiums for PPO plans are typically higher than premiums for POS and HMO plans. One reason for the higher premiums is that PPO plans have less cost control since patients can directly see specialists who usually charge more than primary care physicians [2]. POS plans usually have higher premiums than HMO plans because POS plans partially cover out-of-network visits, which are usually more expensive than in-network visits.

Compared to traditional care plans, managed care plans tend to have higher premiums because managed care plans provide better coverage. Based on a 2005 survey by the Henry J. Kaiser Family Foundation and the Health Research and Educational Trust (HRET), the average monthly premium for family coverage was 832 dollars for a traditional care plan, 871 for a HMO plan, 924 for a PPO plan, and 900 for a POS plan [6]. PPO has the highest premium because it offers the flexibility of seeing a specialist before seeing a primary care physician.

Studies have shown that managed care may have the problem of adverse selection, in which the managed care organizations only allow healthy individuals to become members of the organization [5]. When deciding what premiums to charge plan participants, managed care organizations have wide latitude in determining individuals' premiums. Factors that MCO's consider include age, income, and preexisting conditions. Although there are many different factors that contribute to a person's health, one possible factor is the person's age. Insurance companies may think that older people are more likely to have illnesses and therefore charge them higher premiums. To see if in fact older people do pay higher premiums, I examined data on out-of-pocket premium costs for people in the individual (non-employer sponsored) insurance market.

Data

The data was collected from the National Health Interview Survey, an annual survey conducted by the US Census Bureau on individual persons regarding their medical expenditures, personal health, and demographic characteristics. Cross-sectional individual data from the years 2004, 2005, 2006, 2007, and 2008 were used for this paper. Health insurance variables of interest include the annual out-of-pocket premium cost for the private insurance plan, the plan type such as HMO, PPO, POS, or fee-for-service, and whether the

Table 1: Explanatory Variables

PREMIUM	Annual out-of-pocket premium cost
AGE	Age in number of years
Income	Annual earnings
EDUCATION	Number of years of schooling
USCITIZEN	Dummy for whether the person is a US citizen
MALE	Dummy for whether the person is male
HISPANIC	Dummy for whether the person is hispanic
MEDICALCONDITION	principal component for medical conditions
HMO	Dummy for HMO plan
PPO	Dummy for PPO plan
POS	Dummy for POS plan
FEE	Dummy for fee-for-service plan
OTHER	Dummy for other plan

1

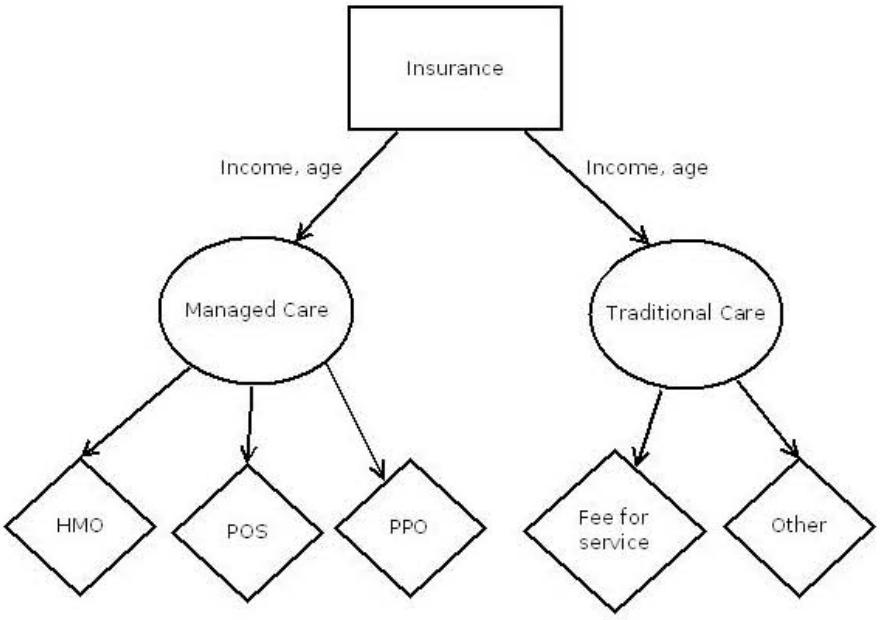


Figure 1: Nested Logit Tree for Choice of Different Insurance Plan Types

Table 2: Separate Year OLS Results, Dependent Variable: Log of Out-of-Pocket Premiums

Variable	2004	2005	2006	2007	2008
$\ln(AGE)$	.4204174	.4151197	.4538275	.3900424	.3581655
	(.0581931)	(.0622168)	(.0752566)	(.0664133)	(.06177)
$\ln(INCOME)$	.0415217	.0557738	.0305913	.0387983	.0535124
	(.0176526)	(.0166717)	(.0166773)	(.0171019)	(.0187026)
$EDUCATION$	.0304392	.0461535	.0466376	.0311253	.0235074
	(.0066975)	(.007141)	(.0086489)	(.0088529)	(.0065983)
$USCITIZEN$	.1221257	.2999292	.1211093	.2544095	.3497774
	(.0972408)	(.0954521)	(.0926383)	(.114078)	(.0834037)
$MALE$	.0771887	.0375758	.1011693	.0386174	.071505
	(.0388674)	(.0403688)	(.0453056)	(.0469243)	(.042062)
$HISPANIC$	−.2304746	−.117611	−.2036018	−.1141885	.0494393
	(.0631677)	(.0590579)	(.0824605)	(.0658442)	(.0599691)
$MEDICALCONDITION$	−.0123592	−.0046269	.0448208	.0251137	.0196089
	(.0275418)	(.0141291)	(.0206096)	(.0292753)	(.0248476)
$dummy\_for\_hmo$	.0879942	.0748514	.1457179	.3420263	.1525497
	(.0718251)	(.0694337)	(.0740579)	(.0914668)	(.071746)
$dummy\_for\_ppo$	.3129684	.1489021	.2489223	.3720535	.203056
	(.0659825)	(.063657)	(.0675239)	(.0837974)	(.0662049)
$dummy\_for\_pos$	.1790635	.234418	.139051	.2061048	−.0768062
	(.1591573)	(.1132297)	(.1766579)	(.168013)	(.1348936)
$dummy\_for\_fee\_for\_service$	−.0075153	−.0225754	−.3068197	−.1012478	−.1120505
	(.1151747)	(.1306332)	(.224515)	(.1565182)	(.1833302)
$dummy\_for\_other$	−.0245825	−.0242369	−.1191394	−.1173949	.1675383
	(.0868418)	(.0996339)	(.0973304)	(.1147397)	(.0993519)
$Numberof\ Observations$	3068	2804	2184	2420	2658
R^2	0.0662	0.0611	0.0740	0.0550	0.0425

plan was paid for individually or by the employer. Demographic characteristics include the individual's annual earnings, age, gender, race, education, and citizenship. Health status variables include whether the individual has trouble walking or remembering and whether he is limited in any other way.

Because the individual's income was reported as a number between 1 and 10, each corresponding to a specific income bracket, the values needed to be replaced with the national mean income in each bracket. Data from the Consumer Expenditure Survey was used to determine the national mean income values.

Since over twenty variables for various medical conditions existed, adding all of the variables into the regression would over-parameterize the model. Therefore, a principal component was created to provide a weighted average of some medical conditions. Because many medical condition variables had a non-response rate over 80%, using a principal component created out of those variables in the regressions would reduce the degrees of freedom significantly. Therefore, only those medical conditions with response rates greater than 50% were included in generating the principal component. The principal component used in this paper combines the dummy variables for the following questions: does the person have trouble walking, does the person have trouble remembering, and is the person limited in any other way?

Table 1 summarizes the variables used in analysis.

Empirical Methods

First, ordinary least squares was used to regress log of out-of-pocket premium costs on log age, log income, a variety of demographic characteristics, a principal component for medical conditions, and dummy variables for plan type.

The OLS specification is the following:

$$\ln(PREMIUM) = \beta_1 \ln(AGE) + \beta_2 \ln(INCOME) + \beta_3 EDUCAT\ ION + \beta_4 USCITIZEN + \beta_5 MALE + \beta_6 HISP\ ANIC + \beta_7 MEDICALCONDIT\ ION + \beta_8 HMO + \beta_9 PPO + \beta_{10} POS + \beta_{11} FEE + \beta_{12} OTHER + \varepsilon_i$$

The idea of this specification is to determine if older or wealthier people tend to pay higher premiums, keeping medical conditions, plan type, and demographic characteristics constant. If they do, then either the insurance companies could be charging them higher premiums, or the older or wealthier people could be demanding good coverage plans with higher premiums. If it turns out that older people do not demand plans with typically higher premiums, then it would be reasonable to assume that insurance companies are charging older people higher premiums, which could signify an adverse selection problem.

After OLS was performed, 2SLS was performed on separate years of data and pooled data. The instruments for income were a variable for whether the individual qualified for food stamps, a variable for whether the individual received income from pensions, and a variable for whether the individual received social security or Railroad Rewards. An overidentification test was performed to verify the exclusion restriction, and a Hausman test was performed to test for errors in variables [7].

Both OLS and 2SLS found that older and wealthier individuals pay higher premiums, keeping other variables constant. In order to test if older or wealthier people demand plans with higher premiums, a nested logit regression was performed on the five different insurance plans: HMO, PPO, POS, fee-for-service, and other. Figure 1 shows the nested logit tree structure.

HMO, PPO, and POS both fall under the category of managed care, while fee-for-service and other fall under the category of traditional care. The case specific variables are income and age, and unfortunately, there are no non-case specific variables due to lack of data on the individual plans. Nested logit regressions using income and age as the top level variables were performed separately on each year.

Results

Separate Year OLS

Table 2 shows the OLS regression results for each year. As we can see, the coefficients on log income and log age are positive and highly significant in all years, which suggests that wealthier and older people do in fact pay higher premiums. The large increase in premiums with age could be due to the fact that elderly people are more prone to illnesses so the insurance company charges them higher premiums in order to compensate for the risk that it is bearing.

Looking at the variable for education, we see that the coefficients for all years are positive and highly significant, which suggests that better educated people pay higher premiums. This result is plausible because better educated people may desire insurance of better quality so they choose the expensive health insurance with extensive coverage.

US citizens also tend to pay higher premiums than non-US citizens, conditional on income, age, and other characteristics being constant. A possible explanation for this finding is that US citizens have faster wage growth than non-US citizens, which means that US citizens can afford better health insurance [1].

The coefficient on the variable for male is positive but not significant in the years 2005 and 2007, which suggests that the effect of gender on premiums is unclear. Although it is true that men on average earn more than women, it is not necessarily the case that men prefer the insurance plans with the higher premiums.

The coefficient for the Hispanic dummy is negative and significant in all years except 2008. The negative sign can perhaps be explained by the fact that Hispanic people on average earn less than non-Hispanics, which suggests that Hispanics are less likely to pay for expensive health insurance [9].

For the medical condition principal component, the coefficient is positive in the years 2006, 2007, and 2008, but negative in the years 2004 and 2005. However, the t-statistics for the coefficients are less than one in all years except 2006, which suggests that health status does not seem to have a significant effect on insurance premiums.

Examining the coefficients on the dummies for the plan types yields some interesting trends. The dummy for the PPO plan has a positive coefficient that is significant at the 5% level for all years, which suggests that PPO plan subscribers pay higher premiums than non-PPO subscribers, keeping other variables constant. Also, it is interesting to note that the magnitudes of the PPO dummy coefficients are larger than the magnitudes of the coefficients for the other managed care plans in all years. This suggests that PPO plans charge higher premiums than HMO or POS plans to individuals with identical characteristics. This finding is consistent with the fact that PPO plans incur the greatest costs for the insurance company since patients have the flexibility to see a specialist directly before consulting with a primary care physician.

When we look at the coefficients for the POS plan, we see that the coefficients are not significant, which means we cannot conclude anything about whether POS plan subscribers pay higher premiums than non-POS plan subscribers. Similarly, the coefficients of the fee-for-service dummy are not significant and neither are the coefficients for the "other" plan type dummy. Since "other" encompasses different varieties of health insurance, it is not surprising that there is no clear trend in premiums for that group.

The separate year regression results have shown us that wealthier and older people pay higher premiums, conditional on other factors staying constant. Also, US citizens and non-Hispanics pay more for health insurance than non-US citizens and Hispanics. Moreover, PPO plans typically charge the highest premiums out of the three types of managed care plans, which is probably because the insurance company has the hardest time with cost control in the PPO plan.

Since the coefficients on income and age do not differ by much over the years, it might be reasonable to run a pooled regression and see if the same effects for income and age emerge. In order to determine if pooled regression would produce the same results on age and income as the separate year regressions, we need to determine if the coefficients on income and age stay the same across all years.

For this purpose, year dummies were created, and the data was pooled. During the pooling process, income and premiums were deflated to 2008 dollars using the CPI for all urban consumers using all items. Then, log income and log age were interacted against the year dummies for 2007, 2006, 2005, and 2004. The following OLS regression was performed:

$$\ln(PREMIUM) = \beta_1 \ln(AGE) + \beta_2 \ln(INCOME) + \beta_3 EDUCAT\,ION + \beta_4 USCITIZEN + \beta_5 MALE + \beta_6 HISPANIC + \beta_7 MEDICALCONDITION + \beta_8 HMO + \beta_9 PPO + \beta_{10} POS + \beta_{11} FEE + \beta_{12} OTHER + \textit{interaction terms between age and year} + \textit{interaction terms between income and year} + \varepsilon_i$$

Then, a Chow test was performed on the null hypothesis that the coefficients for the interaction terms are equal to zero. Jointly testing that the interaction terms are equal to zero is equivalent to testing that the coefficients on income and age are the same across all years. The Chow-statistic with 8 degrees of freedom in the numerator and 13109 degrees of freedom in the denominator from that test was equal to 0.28, which corresponds to a p-value of 0.9714. Therefore, we cannot reject our null hypothesis that the coefficients for income and age are the same across all years. Pooling the data would thus give the same OLS results on income and age as the separate year OLS regressions.

Pooled OLS Results

Table 3 shows the pooled OLS results.

As we can see, the pooled OLS coefficients for age and income are similar to the coefficients obtained in the separate year OLS regressions. The coefficient for age is 0.40, which lies within the 0.35 and 0.45 range for the separate year coefficients. Similarly, the coefficient for income (0.045) lies within the 0.03 to 0.06 range for the separate year coefficients. The coefficients on the other variables remain significant and retain the same sign as in a majority of the separate year regressions.

2SLS Results

The OLS results have given support to the fact that older and wealthier people pay more for health insurance, keeping other factors constant. However, the OLS specification may produce inconsistent estimates in the presence of errors in variables. Since we only observe an individual's income bracket instead of the actual income, we suspect measurement error in the income variable, which will cause OLS income estimates to be biased downward. In order to test for errors in variables, two stage least squares was performed on each year's of data using the following three instruments for income: FOODSTAMP which measures whether the individual qualified for foodstamp programs, SSRR which measures whether the individual qualified for social security or Railroad Rewards, and PENSION which measures whether the individual qualified for pension programs. An overidentification test was performed to test a joint exclusion restriction that all of the instruments are uncorrelated with the error term. Table 4 shows the 2SLS results for each year. All the overidentification statistics are less than the chi-squared statistic with 2 degrees of freedom at the 30% level, so therefore the exclusion restriction holds. Also, in the first stage regressions, the coefficients on the instruments are significant, which indicates that the relevance condition is satisfied. Intuitively, it makes sense for FOODSTAMP, PENSION, and SSRR to be correlated with income since whether or not one receives foodstamps is correlated with one's income. Similarly, whether or not people receive pensions or social security depends on whether or not one has retired, which is correlated with one's age and income, but not the premiums that one pays.

If we compare Table 4 to Table 2, we see that the coefficients on income are greater in the 2SLS specification than in the OLS specification, which indicates the presence of measurement error in the income variable. Interestingly, the age coefficient is smaller in the 2SLS specification than in the OLS specification, which suggests that some of the variation in the age coefficient was

Table 3: Pooled OLS Results, Dependent variable: log of out-of-pocket premiums

Variable (Standard Error)	Pooled OLS
$\ln(AGE)$	.405053 (.0287427)
$\ln(INCOME)$	.0446586 (.0078165)
$EDUCATION$	.0349317 (.003324)
$USCITIZEN$	.2353483 (.0433879)
$MALE$	.0608728 (.018999)
$HISPANIC$	−.1184009 (.0297783)
$MEDICALCONDITION$	.0125261 (.0107097)
$dummy\_for\_hmo$	.153905 (.0339851)
$dummy\_for\_ppo$	.2552552 (.0312121)
$dummy\_for\_pos$	.138467 (.0676477)
$dummy\_for\_fee\_for\_service$	−.0919594 (.071846)
$dummy\_for\_other$	−.0247504 (.0446242)
$year2004$	4.914516 (.1404129)
year2005	4.998325 (.1415235)
year2006	4.98287 (.1419267)
year2007	4.993465 (.1420729)
year2008	4.922989 (.141527)
$Numberof\ Observations$	13134
R^2	0.9828

Income and Premiums were deflated to 2008 dollars using CPI for all items.

absorbed by the instruments. This makes sense because PENSION and SSRR are positively correlated with age.

The coefficients on education lose their significance when we use 2SLS, which suggests that much of the variation in education could be explained by the instruments. Once we include the instruments in the model, the effect of education on premiums goes away.

The coefficients on citizenship remain positive and significant as in the OLS model, which confirms the fact that keeping other factors constant, US citizens pay more for health insurance than non-US citizens.

Interestingly, the coefficient on male switches sign when we use IV, which suggests that the effect of gender on premiums is indeterminate.

The coefficient on Hispanic remains negative and significant in all years except 2008, which confirms the OLS result that Hispanics pay lower premiums, keeping other variables fixed.

The medical condition coefficient becomes negative and significant in years 2004, 2005, and 2007, and it is insignificant in other years. A possible explanation for this effect is that the principal component variables of whether the person has difficulty walking or remembering are not indicative of one's health status. However, due to the high non-response rates for other medical conditions, a tradeoff was made in favor of preserving the number of degrees of freedom.

In order to verify that the OLS estimates are inconsistent, a Hausman test was performed on each year's OLS and 2SLS results [7]. The Hausman statistics with 1 degree of freedom are shown in

Table 4: 2SLS Results for each year

Variable	2004	2005	2006	2007	2008
ln(*AGE*)	.2270316	.2570671	.3073388	.2292546	.2269913
	(.0672485)	(.0690353)	(.0796716)	(.0788999)	(.087212)
ln(*INCOME*)	.4715432	.4334326	.4840748	.3339356	.2858897
	(.064938)	(.0609526)	(.0825295)	(.0698639)	(.1091581)
EDUCATION	−.0093745	.0115651	.003576	.003419	.002331
	(.0093035)	(.0097041)	(.011818)	(.0108912)	(.0123945)
USCITIZEN	.1115718	.2781028	.0972838	.1728524	.291594
	(.0900009)	(.0949033)	(.1037401)	(.0983807)	(.0928799)
MALE	−.176253	−.2056726	−.2053188	−.1571799	−.0743381
	(.0560732)	(.0578194)	(.0745904)	(.0657968)	(.080502)
HISPANIC	−.3015365	−.1999004	−.2694939	−.1990006	.0039647
	(.0677656)	(.0731439)	(.0795579)	(.0769117)	(.0697193)
MEDICALCONDITION	−.1031864	−.0480112	.0038743	−.0330214	−.0330578
	(.031579)	(.0222164)	(.0292037)	(.0329017)	(.036995)
Dummy HMO	−.0721288	−.0533092	−.0881288	.2397646	.0587676
	(.0783737)	(.0768556)	(.0970516)	(.0874482)	(.0892901)
Dummy PPO	.1737884	.0673431	.0135002	.2764946	.1265857
	(.0744768)	(.0702699)	(.0917101)	(.0802111)	(.0802169)
Dummy POS	−.0024859	.1127285	−.0994332	.2258198	−.1047753
	(.1385513)	(.1384202)	(.1678121)	(.1574213)	(.1392627)
Dummy FEE	.023998	.1168527	−.3527816	−.069674	−.0660702
	(.1317011)	(.1248004)	(.1506026)	(.1563709)	(.1599011)
Dummy OTHER	−.002216	.0274523	−.0831459	−.053187	.2611311
	(.0974525)	(.1057953)	(.1240865)	(.121285)	(.1240023)
Numberof Observations	3068	2804	2184	2420	2658
Overid Statistic	.466826	.411061	.467722	1.38813	.573292
p-value	0.7918	0.8142	0.7915	0.4995	0.7508

Table 5: Hausman Statistics for Test of Errors in Variables

	2004	2005	2006	2007	2008
Hausman statistic with 1 degree of freedom	46.74	41.32	31.71	19.17	4.65

Table 5. As we can see, all of the Hausman statistics are greater than the chi-squared statistics with one degree of freedom, which means we reject the null hypothesis that OLS is consistent. Next, 2SLS was performed on the pooled data using the same instruments. Table 6 compares the pooled OLS and 2SLS results.

Since the overidentification statistic of 1.00 is less than the chi-squared statistic with two degrees of freedom at the 20% significance level, we cannot reject the null hypothesis that the joint exclusion restriction holds, giving support to the validity of the instruments. The Hausman statistic of 6349 confirms the presence of an errors in variables problem.

The OLS and 2SLS coefficients on US Citizen differ by about 20% but retain their sign and significance, confirming the fact that US citizens pay more for health insurance than non-US citizens, keeping other factors constant. The coefficient on the PPO dummy remains positive and significant between the two specifications, confirming that PPO plans charge higher premiums than other plans.

By comparing the OLS and 2SLS results on year-by-year and pooled data, we have seen greater income coefficients and smaller age coefficients in the 2SLS specification than in the OLS specification. Despite the large differences in the magnitudes of the coefficients, they nevertheless remain positive no matter what specification is used. It would be good to provide some explanation for why after keeping demographics and indicators of potential health problems constant, older and wealthier people are likely to pay more for health insurance than younger and poorer people. Two possible explanations exist. First, it is possible that older and wealthier people are more risk averse and prefer the plans with

Table 6: Pooled OLS and 2SLS: Dependent Variable: log premiums

Variable (Standard Error)	Pooled OLS	Pooled 2SLS
$\ln(AGE)$	.405053	.2372954
	(.0287427)	(.033888)
$\ln(INCOME)$	.0446586	.407034
	(.0078165)	(.0367217)
$EDUCATION$	.0349317	.0014867
	(.003324)	(.0048946)
$USCITIZEN$	.2353483	.1884017
	(.0433879)	(.0423754)
$MALE$	.0608728	−.1654679
	(.018999)	(.0306198)
$HISPANIC$	−.1184009	−.1922166
	(.0297783)	(.0326707)
$MEDICALCONDITION$	.0125261	−.045
	(.0107097)	(.0131676)
$dummy\_for\_hmo$	.153905	.0114374
	(.0339851)	(.0381438)
$dummy\_for\_ppo$	.2552552	.132894
	(.0312121)	(.0352123)
$dummy\_for\_pos$	.138467	.0409361
	(.0676477)	(.0656601)
$dummy\_for\_fee\_for\_service$	−.0919594	−.039643
	(.071846)	(.0633929)
$dummy\_for\_other$	−.0247504	.0330499
	(.0446242)	(.0498907)
$year2004$	4.914516	−.077244
	(.1404129)	(.0316351)
year2005	4.998325	.0064345
	(.1415235)	(.0322337)
year2006	4.98287	5.145605
	(.1419267)	(.1411761)
year2007	4.993465	.0222087
	(.1420729)	(.0333382)
year2008	4.922989	−.063389
	(.141527)	(.0326214)
$Numberof\ Observations$	13134	13134
R^2	0.9828	0.9826
Overid Statistic		.999881
Hausman Statistic		6349.3504

Instruments: FOODSTAMP, PENSION, and SSRR

the higher premiums, which usually have better coverage than the plans with the lower premiums. Second, it is possible that adverse selection is present, in which the insurance companies charge older people higher premiums because they believe that older people are more prone to medical illnesses. In order to disentangle these two effects, a nested logit regression was performed to determine if older people prefer managed care plans over traditional care plans. If it turned out that older people prefer the more expensive managed care plans over traditional care plans, then it is plausible that older people are choosing the plans with the higher premiums (perhaps because they would like better coverage). But if it turned out that older people actually favored traditional care over managed care but are still paying higher premiums, then it could be that adverse selection based on age is present.

Nested Logit Results

In order to determine the plans favored by the elderly and wealthy individuals, nested logit regression was performed. Table 7 shows the nonnormalized nested logit results where income and age are case-specific variables at the top level of the nested logit tree.

As we can see, the coefficient on the interaction term between income and traditional care is negative and significant at the 1% level for all years. This strongly suggests that wealthier people prefer managed care over traditional care. A possible explanation is that wealthier people desire plans that provide primary and preventive care coverage in addition to coverage for severe illnesses since wealthier people may routinely visit a physician for a regular checkup. Or it could be that wealthier people place more trust in the physicians that are part of the network of a managed care organization and therefore enroll in the managed care plans.

The coefficients for the interaction term between age and traditional care are positive and highly significant across all years. This suggests that older people prefer traditional care over managed care. A possible explanation is that older people may be accustomed to the traditional care plans that were popular in the 1980s and consequently decided to stay in the same plans that they had in their youth. Or, it could mean that older people would like the traditional care plans with the better deals even if they do not provide as good of coverage as the managed care plans.

Based on the nested logit results, we cannot say that adverse selection based on income exists because wealthier people are choosing the more expensive managed care plans. However, it is possible that adverse selection based on age exists because the older people prefer the less expensive traditional care plans, and yet they are paying higher premiums than younger people, keeping medical conditions, income, and other demographics constant.

Answers to Possible Objections

Omitted Variable: Regional differences in health care premium

A possible argument is that regional differences in health care plans may contribute to the different premiums that individuals face and that leaving out the regional effects might introduce bias if one of the explanatory variables was correlated with the region variable.

Table 7: Nested Logit Results with Income and Age as Top level Variables

Variable (standard error)	2004	2005	2006	2007	2008
dummy for hmo	1.383061 (20.97903)	1.467541 (.3298407)	1.414274 (.3806032)	1.61868 (11.87173)	1.71384 (.1187688)
dummy for ppo	1.786377 (20.97902)	2.055911 (.3289393)	2.020572 (.3795774)	2.315035 (11.8717)	2.434116 (.115598)
dummy for pos	−.9544614 (20.97927)	−.7203107 (.3454905)	−.7950022 (.3984572)	−.6262243 (11.8723)	−.4548043 (N/A)
dummy for fee for service	−.9080833 (.1196628)	−.4347192 (.1212794)	−.5632885 (.1418441)	−.7149679 (.1479859)	−.8244304 (.1588591)
traditional care * ln(age)	1.282378 (.1807629)	2.153326 (.2245949)	2.277842 (.2697607)	2.075714 (.2526602)	1.807026 (.2479094)
traditional care * ln(income)	−.2573889 (.0429034)	−.2501298 (.0454457)	−.2516521 (.0509986)	−.2276727 (.0512258)	−.3247668 (.0511939)
Numberof Observations	3068	2804	2184	2420	2658

Table 8: Hausman statistics for test of IIA

	2004	2005	2006	2007	2008
Hausman statistic with 6 degrees of freedom	-964.80138	-102.52895	-238.15736	332.04984	586.91815

To test this argument, dummies for the four regions (Northeast, Midwest, South, and West) were added into the pooled least squares regression. After adding in the dummies, coefficients on all variables retained the same sign, and their magnitudes did not change by more than 15%. Therefore, omitting regional dummies from the specification should not be a problem.

Variables for the principal component

Some may argue that the variables used to determine the principal component are not indicative of serious medical conditions. It is true that the existence of any limitations in one's ability to walk may not necessarily indicate any serious medical problems. However, due to the high non-response rate for questions on medical conditions, there was a tradeoff between degrees of freedom and quality of indicators of medical conditions. In order to preserve a reasonably high number of degrees of freedom (2000 as opposed to 150), only the variables with the highest response rates were used.

Heteroskedasticity

Initially, there were concerns over whether heteroskedasticity would be a problem because the White Test performed on the least squares results rejected null hypothesis of homoskedasticity. However, when robust standard errors were used, the p-values did not change by more than 15%, which suggests that heteroskedasticity should not affect the results.

Conditional Logit

Some may wonder why the conditional logit was not used in determining the preferences of wealthy and elderly people for managed care versus traditional care. The reason is that conditional logit regression relies on the assumption of the Independence of Irrelevant Alternatives, which means that individuals' preferences for a particular insurance plan should not depend on the other insurance plans. In order to test whether this assumption was true, a Hausman-McFadden test was performed on an unrestricted and restricted conditional logit model. In the restricted logit model, the HMO plan was dropped from the choice set. Table 8 shows the Hausman statistics from that test.

As we can see, the Hausman statistics are either larger than their threshold chi-squared values, or they are negative, suggesting that the asymptotic assumptions of the Hausman-McFadden test are not met. Since none of the statistics are positive and less than their threshold chi-squared values, we cannot say that the Independence of Irrelevant Alternatives assumption is satisfied. Therefore, it is better to use the nested logit model which does not rely on the IIA assumption.

Conclusion

This paper has examined the effects of the purchaser's age and income on her health insurance premiums in the individual (non-employer sponsored) market. After OLS and 2SLS regressions were performed, highly significant positive effects were found for both age and income when controlling for medical conditions, plan type, and other demographic characteristics. These results confirm the original hypothesis that older and wealthier people tend to pay higher premiums, after controlling for medical condition and demographics.

In an attempt to determine if the higher premiums are a result of adverse selection by the insurance company or of demand for more expensive insurance plans by wealthier and older buyers, a nested logit regression was performed to determine which plans were favored by wealthier and older individuals. It was found that wealthier people favor the more expensive managed care plans, which provide better coverage; this finding suggests that greater demand for managed care plans by wealthy individuals may be driving up the premiums for these plans. It was also found that older people actually prefer the less expensive traditional care plans over managed care plans, which suggests that adverse selection based on age may exist because insurance companies are charging older people higher premiums even when they prefer the less expensive plans. A possible explanation for this adverse selection effect is that insurance companies believe older people are more prone to serious illnesses and therefore need to insure against the greater expected losses by charging higher premiums.

The result of this study on the effects of age and income on out-of-pocket health insurance premiums in the individual market has important implications in light of the aging baby-boomer generation. According to the results of this study, as the average age of the US population increases, average premiums in the individual market would also increase. Furthermore, the results showing the preferences of older individuals for less expensive traditional care indicate the possibility of adverse selection based on age. The possible existence of adverse selection may have important policy implications for deciding whether it is necessary to have more governmental regulation in the individual health insurance industry.

Acknowledgement

I would like to thank Professor Jerry Hausman and Alp Simsek for helping me with this paper.

References:

1. Bratsberg, B. et al. (2002). The effect of naturalization on wage growth: A panel study of young male immigrants. Journal of Labor Economics, 20(3):568–597.

2. Claxton, G. et al. (2006). Health benefits in 2006: Premium increases moderate, enroll¬ment in consumer-directed health plans remains modest. Health Affairs, 25:w476–w485.

3. Davis, K. (2007). Uninsured in America: problems and possible solutions. British Medical Journal, 334:346–348.

4. DAY, J. C. (2008). Population profile of the united states. http://www.census.gov/population/www/pop-profile/natproj.html.

5. Frank, R. et al. (2000). Measuring adverse selection in managed health care. Journal of Health Economics, 19(6):829–854.

6. Gabel, J. et al. (2005). Health benefits in 2005: Premium increases slow down, coverage continues to erode. Health Affairs, 24(5):1273–1280.

7. Hausman, J. (1978). Specification tests in econometrics. Econometrica, 46(6):1251–1271.

8. Knight, W. (1998). Managed Care: What it is and how it works. Aspen Publishers, Inc.

9. Reimers, C. (1983). Labor market discrimination against hispanic and black men. The Review of Economics and Statistics, 65(4):570–579..

The Structure of Stable Vector Fields on Surfaces[1]

John Berman[2], Sergei Bernstein[2]

1. Department of Mathematics, MIT, November 2010
2. Student Contributors, Class of 2013, Deparment of Mathematics, MIT

The Poincare-Hopf theorem tells us that given a smooth, structurally stable vector field on a surface of genus g, the number of saddles is 2−2g less than the number of sinks and sources. We generalize this result by introducing a more complex combinatorial invariant. Using this tool, we demonstrate that many such structurally stable vector fields are equivalent up to a set of basic operations. We show in particular that for the sphere, all such vector fields are equivalent.

Introduction

In order to study combinatorial aspects of vector fields, we restrict our attention to surfaces that are compact, connected, and orientable. The Poincare-Hopf Theorem yields a simple combinatorial invariant of a vector field on such a surface; in particular, it tells us that

$$U + I - A = 2 - 2g$$

where we write U; I;A for the numbers of sources, sinks, and saddles, respectively, and g for the number of holes in X, also known as the genus of X.

We seek to extend this simple result to a more powerful combinatorial invariant in the form of a graph embedded on X, with edges representing integral curves of the vector field. In order to accomplish this, we must restrict our attention to structurally stable vector fields, or vector fields whose geometric behavior is unchanged by small perturbations.

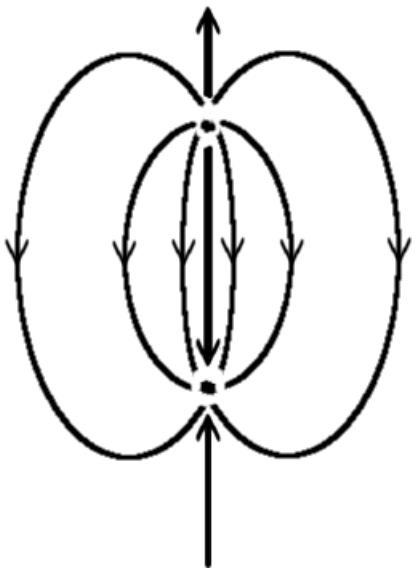


Figure 1. An example of a primitive vector field on a sphere.

We show that our graph invariant is a triangulation of X, and, using this machinery, we seek to describe classes of vector fields that may be deformed into each other via simple, local operations we term vector field cobordisms. In our main result, we show that all structurally stable vector fields on the sphere are cobordant in this sense, and we discuss the extent to which this holds true for surfaces of higher genus. Unfortunately, for surfaces of higher genus, the existence of periodic integral curves hampers the project, so we leave the readers with some open questions.

The Graph Invariant

Primitive and Generic Vector Fields

Consider a smooth, structurally stable vector field on a (compact, connected, orientable) surface. By Peixoto's theorem [2], structural stability is equivalent to a simple set of properties, as follows.

- The set of non-wandering points consists only of periodic orbits and fixed points. In other words, an integral curve becomes either a periodic orbit or a fixed point in the limit.
- The set of fixed points is finite and consists only of hyperbolic equilibrium points. In particular, there are saddles of degree 4, sinks, and sources.
- There are a finite number of attracting or repelling periodic orbits.
- There are no saddle-saddle connections. We shall call such a vector field *generic*, and in particular a generic vector field with no such periodic orbits *primitive*. The fixed points consist of sinks, sources, and saddles, the numbers of which we will denote throughout I, U, and A, respectively.

The Invariant

Suppose we have a primitive vector field v on a surface X. As it has a finite number of fixed points, it is natural to construct a graph $G(v)$ (or just G if v is known from context) with the vertices corresponding to the fixed points of v. We then say that there is a

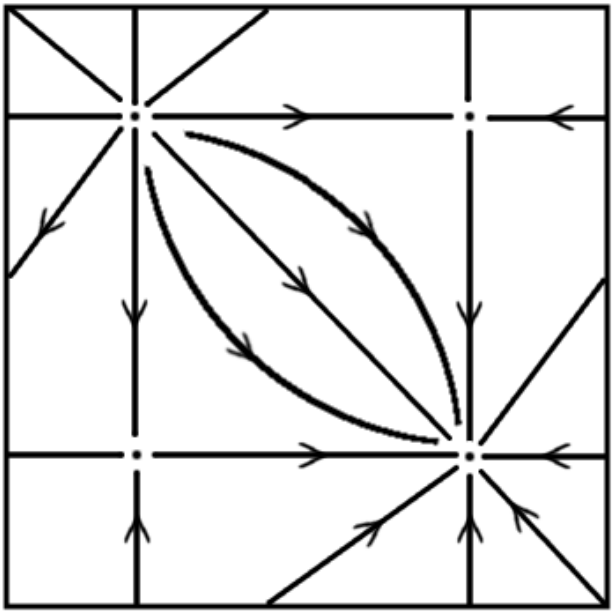


Figure 2. An example of a primitive vector field on a torus

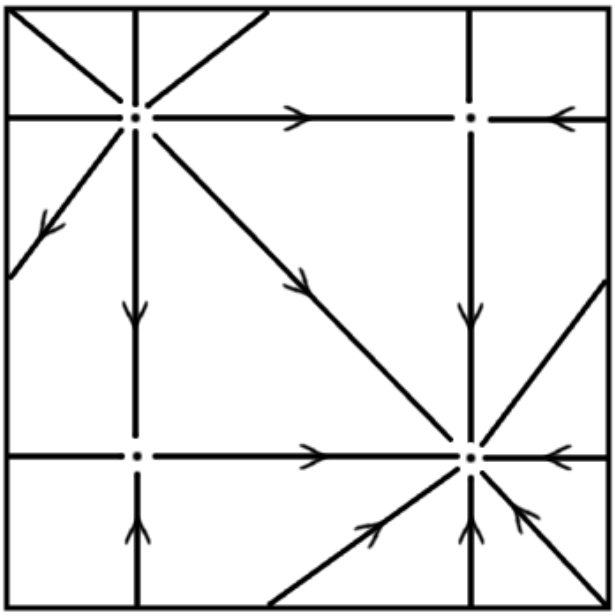


Figure 3. G in the previous example.

directed edge from A to B for each homotopy class of integral curves that start at A and end at B. We notice a few properties

- There is a natural way to pick an isotopy class of embeddings of G on X.
- By structural stability of v, G must be 3-colorable with the colors corresponding to sources, sinks, and saddles.
- Also by structural stability, each saddle point must have in-degree 2 and out-degree 2.

This graph $G(v)$ will be our primary tool for studying vector fields.

If we remove all the sources and associated edges from G, we get a new graph, which we denote $G_I(v)$ (or just G_I). We notice that in G_b, every saddle connects to exactly 2 sinks. Therefore, we can construct yet another (undirected) graph $\tilde{G}_I(v)$ where the vertices are the sinks of GI , and there is an edge between two (not necessarily distinct) sinks for each saddle that is connected to both of those sinks. We can construct similar graphs $G_U(v)$ and $\tilde{G}_U(v)$ by interchanging sinks and sources. It is immediately clear by definition that $\tilde{G}_I$ consists of I vertices and A edges, and that $\tilde{G}_U$ consists of U vertices and A edges. Furthermore, all of these graphs may be naturally embedded on X, as G itself may be embedded on X.

We now present some examples in Figures 1 through 7. Note that we represent vector fields by their flows. For the vector field in Figure 1, G consists of a source, a sink, and a directed edge from the source to the sink. The graphs G_U and $\tilde{G}_U$ are each just a single vertex.

In Figure 2, we present an example of a primitive vector field, v_T , on a torus. In this figure, we draw the torus as a square whose opposite edges are glued together. Figure 3 shows a natural embedding of $G(v_T)$ on the torus. Figures 4 and 5 are embeddings of $G_U(v_T)$ and $\tilde{G}_U(v_T)$, respectively.

In Figure 6, we give an example of a possible value of $G_I(v)$ on the sphere.

For the sake of clarity, we have stereographically projected the sphere onto the plane. Figure 7 then gives us a possible value of $\tilde{G}_I(v)$ on the sphere.

Conditions on G(v)

Given an embegging of a graph into X, we define a region to be a connected component of the complement of that graph in X.

Theorem 1. *Given an embedding of the graph $\tilde{G}_I(\tilde{G}_U)$ in X,*

- *In each region, there is exactly one source (sink) when the vertices of G are added.*
- *Each region (taken as an open set) is diffeomorphic to a open set of R^2.*
- *The graph $\tilde{G}_I$ is connected.*

Proof. Each region must contain at least one source, as each region has some sink on its boundary which must connect to a source in the region via an integral curve. Say that there are v_I , e_I , f_I vertices, edges, faces respectively in $\tilde{G}_I$. This means that $v_I = I, e_I = A,$ and

$f_I \leq U$, with equality if and only if there is exactly one source of G in each region. Letting g denote the genus of X, we can write the Poincare-Hopf result as

$$I + U = A + 2 - 2g$$

and Euler's result as

$$v_I + f_I \geq e_I + 2 - 2g$$

with equality only when each region is diffeomorphic to an open set of R^2 and $\tilde{G}_I$ is connected. Combining our observations results in

$$2 - 2g \leq v_I + f_I - e_I \leq U + I - A = 2 - 2g$$

Thus, both equalities cases hold, proving the theorem. Of course the same argument can be made for $\tilde{G}_U$.

From the first point of Theorem 1 and the observation that saddles have in degree 2 and out degree 2, it follows immediately that $\tilde{G}_I$ and $\tilde{G}_U$ are dual graphs on X.

Suppose that v has at least one saddle. Let us look at a single region in $G_I(v)$. As v contains a saddle, G_I contains an edge. Suppose that the region is bounded by the edges $e_1, \ldots, e_n$ in order (edges may repeat, but would then be oriented differently). Let us examine an edge e_i. Then it is not hard to see that both vertices of e_i must define a homotopy class of integral curves (with the correct orientation) connecting to the single source in the region, for they have nowhere else to connect. This gives us a triangle between the single source and these two vertices. Hence, when G is reconstructed, the particular region of G_I is partitioned into triangles. It follows that G is a triangulation of the entire manifold.

Corollary 1. *As long as v has at least one saddle, G is a triangulation of X. Note that by the word triangulation, we dictate not only that each region be triangular, but also that each region be diffeomorphic to an open set of R^2, as in Theorem 1.*

Of course, G will also satisfy some additional trivial properties. As v is structurally stable, with no saddle-saddle connections, G will be 3-colorable by the "colors": sink, source, and saddle. In fact, each sink will have out-degree 0, each source in-degree 0, and each saddle both in-degree and out-degree 2. Call a graph that satisfies all of these properties saddled. It happens that G is characterized by being a saddled triangulation.

Theorem 2. *Given a graph G on a surface X, there is a primitive vector field v satisfying G = G(v) if and only if G is a saddled triangulation of X.*

Proof. Corollary 1 proves the forward direction. We shall now show the other direction. Suppose G is a saddled triangulation of X. We can embed G into X in such a way that all the edges are smooth curves. This splits the surface into some triangular regions. We note that because G is a saddled triangulation, the vertices of each triangle must be a source, a sink, and a saddle. It is sufficient to specify the flow in each triangular region. We can just specify the ow to be diffeomorphic to the one in Figure 8.

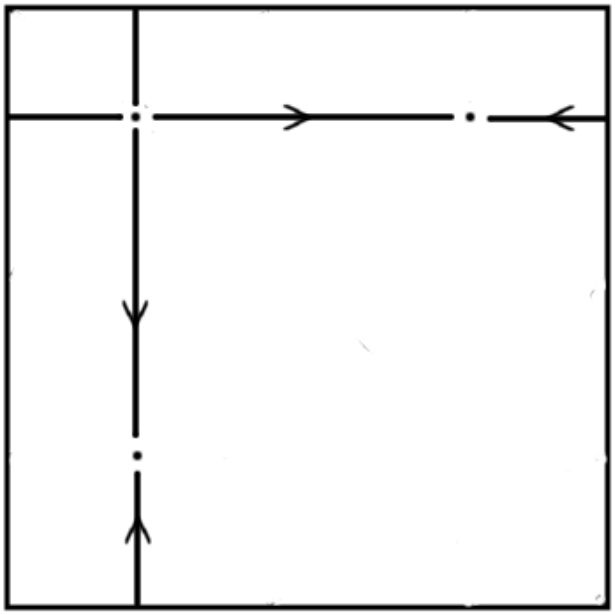


Figure 4. G_U in the previous example.

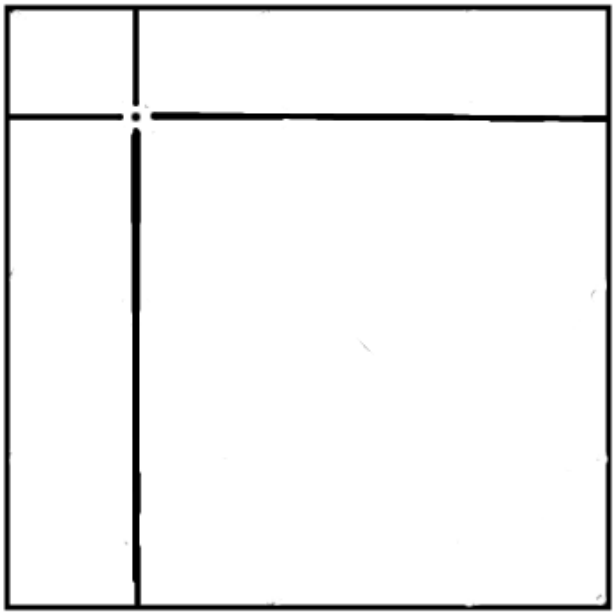


Figure 5. $\tilde{G}_U$ in the previous example.

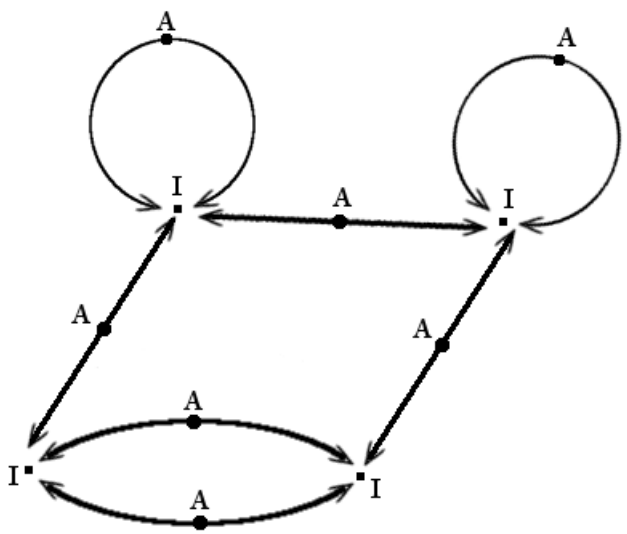


Figure 6. A non-trivial example of GI the sphere (stereographically projected to the plane).

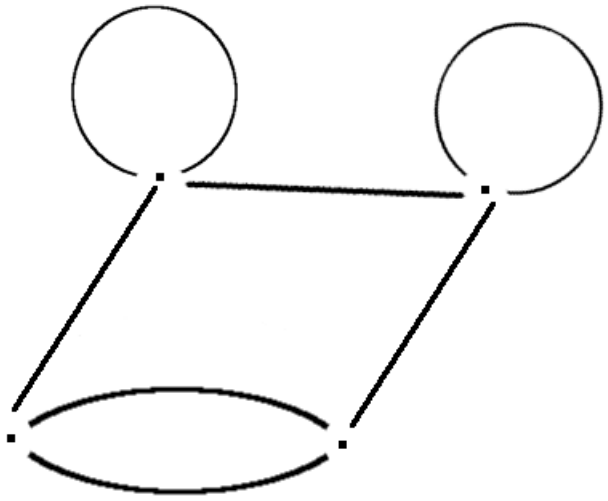


Figure 7. $\tilde{G}_1$ in the previous example.

Non-primitive vector fields

Suppose now that v does have periodic integral curves. We can think of each periodic curve as dividing a small open set around it into two regions. Consider its behavior in one of the regions. The curve acts either as a 'sink' or a 'source,' with integral curves either progressing away from it or towards it in its neighborhood. In fact, within this region, we can close of the integral curve (or associate it with a single point) in such a way that it truly does define either a sink or a source (and no periodic curve) (see Figure 9).

Stated more clearly, suppose there exist periodic integral curves. Ennumerate the regions $R_1, \ldots, R_n$ into which they divide X. In each region, the bounding integral curves may be associated with (closed of to yield) points, with resulting manifolds $X_1, \ldots, X_n$ and derived vector fields $v_1, \ldots, v_n$. Each of the derived vector fields exist on the corresponding manifold and are equivalent to v except that the original periodic curves are now associated with either sinks or sources. In particular, this means that each v_i is primitive, and therefore the results of the previous sections hold.

Vector Field Cobordisms

Definitions

The object of this section is to describe elementary local transformations of vector fields, which we call vector field cobordisms. We define two classes of cobordisms for primitive vector fields. Given two sources, each connecting to a given saddle, we can smoothly deform the vector field in such a way that the sources and the saddle are `pushed' together. Outside of a small neighborhood of the three points, the entire neighborhood acts as a source (see Figure 10), with the vector field emanating away from the region. Thus we define a cobordism to be the transformation from this vector field to another where the three points have been identified with a single source. Of course, the reverse transformation is also a cobordism. The second class of cobordisms we define takes two sinks and a saddle to a sink in an identical way. We call these source-cobordisms and sink-cobordisms, respectively.

We say that two vector fields are cobordant if there is a series of cobordisms and isotopies that takes one to the other. This gives us cobordism classes of vector fields.

Primitive Vector Fields

Consider the effect of cobordisms on the graph $\tilde{G}_1(v)$. Sink-cobordisms act on two adjacent (and distinct!) vertices and an edge between them by removing the edge and equating the two vertices, so that all the edges of each vertex are now assigned to the new vertex. Source-cobordisms take an edge between two (distinct!) regions and remove it, equating the regions. In this way, cobordisms are not only operations on vector fields, but also on connected graphs embedded on X.

Figure 9. A picture to clarify why a periodic curve can be viewed as a source (or a sink).

Figure 8. An example of a smooth flow within a triangular region.

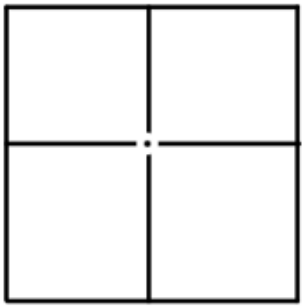


Figure 11. A simple embedding of $\tilde{G}_0$: Note, we represent the torus as a square with opposite edges set equal.

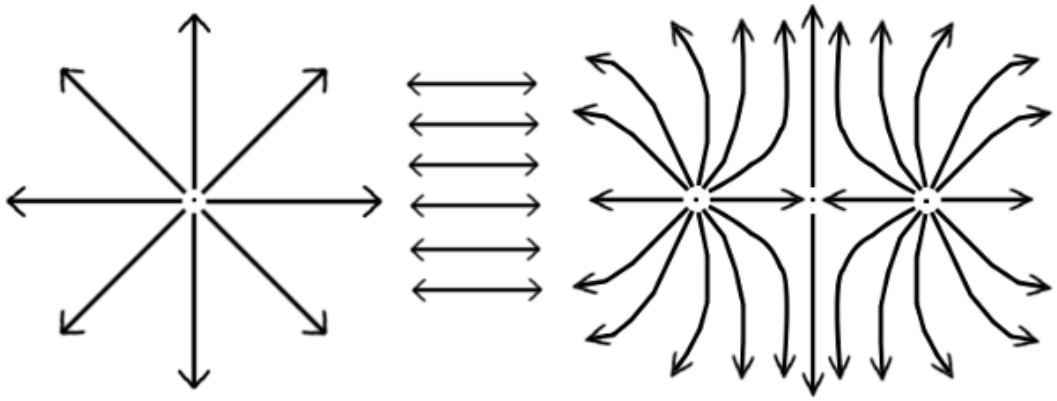


Figure 10. A source cobordism.

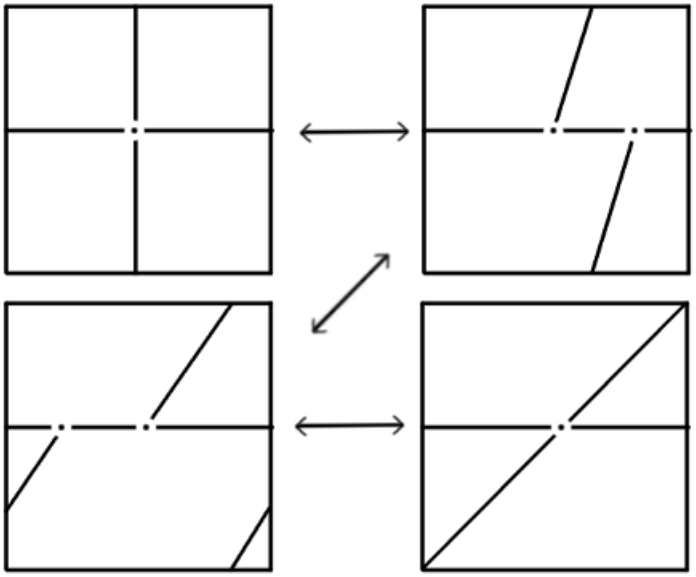


Figure 12. A series of cobordisms and isotopies.

Lemma 1. *Every primitive vector field can be reduced via cobordisms to a vector field with just one source and one sink.*

Proof. Suppose a vector field with more sinks (for example) could not be further reduced via cobordisms. But $\tilde{G}_v$ is connected by Theorem 1, so there are two adjacent distinct vertices with an edge between them, and we can reduce via a sink-cobordism. On the other hand, all primitive vector fields have infinitely many integral curves that need a place to start and a place to end, so every primitive vector field must contain at least one sink and one source.

Call such a primitive vector field with just one source and one sink reduced. To what extent can these differ? The Poincare-Hopf theorem tells us that the number of saddles is specified at $2g$. Each saddle must connect twice to the source and twice to the sink. However, we know that G is a triangulation of X, so we see that the number of homotopy classes of integral curves between the sink and the source is thus fixed at $4g$ (provided $g > 0$). Thus, given the manifold's genus g, there is only one possible graph $G(w)$ for w reduced. Care must be taken, however, as G may embed in multiple ways into X. In fact, if we assign a certain embedding to the identity element in the mapping class group of X, then the different embeddings of G correspond to elements of the mapping class group. There is not, therefore, necessarily only one cobordism class of primitive vector fields.

It is true, however, that if X is a sphere, all primitive vector fields are cobordant. As above, all primitive vector fields are certainly cobordant to a reduced vector field with just one sink, one source, and one homotopy class of integral curves between them. As the mapping class group of the sphere is trivial, we can deform all such vector fields into each other, so there is just one cobordism class.

Theorem 3. *There is just one cobordism class of primitive vector fields on the sphere.*

We now show that there is one cobordism class for the torus and we conjecture that there is also only one cobordism class for surfaces of higher genus.

Primitive Vector Fields on a Torus

Suppose we have two vector fields v, w on X such that the natural embeddings of $G(v)$ and $G(w)$ are isotopic. Because these graphs are triangulations, it is clear that there is an isotopy that takes v to w.

Next, we suppose that we have two vector fields v, w such that the natural embeddings of $\tilde{G}_v(v)$ and $\tilde{G}_v(w)$ are the same. By Corollary 1, each region made by these graphs is diffeomorphic to R^2, and there is exactly one sink in each of these regions. Because the mapping class group of R^2 is trivial, there is only one way to put one sink in each region up to isotopy. In other words, we conclude that $G(v)$ is isotopic to $G(w)$, which in turn implies that v is isotopic to w.

Call the complete tripartite graph with one source, one sink, and two saddles G. By lemma 1, we know that every vector field is cobordant to a vector field v such that $G(v) = G$ as a graph. $\tilde{G}_v$ is just the graph with one vertex and two edges. Therefore, to show that every vector field on the torus is cobordant, it is sufficient to show that any embedding of $\tilde{G}_v$ is cobordant to every other embedding. In particular it is sufficient to show that the simple embedding in Figure 11 is cobordant to every other embedding. Suppose we have an embedding E that is not isotopic to this one. There is some element of the mapping class group of the torus that takes our simple embedding to E. If two embeddings correspond to the same element of the mapping class group, there is an isotopy taking to the other. Therefore, we want to show that we can use cobordisms to reach an embedding corresponding to any element of the mapping class group.

However, we can perform the series of cobordisms shown in Figure 12. In this way, we have found two embeddings that are not isotopic but are cobordant. In fact, we notice that Figure 12 simply gives us a Dehn twist on the torus along one of the two edges of the graph. Therefore, if we can reach a certain embedding through Dehn twists along these edges, we can reach it through cobordisms. However, by the Lickorish twist theorem [1], we know that Dehn twists along these curves generate the mapping class group.

Non-Primitive Vector Fields

When allowing for periodic curves, the two cobordisms we have defined will certainly not be sufficient to alter or remove periodic curves, thus creating an infinite class of cobordism classes. To account for this, we propose an additional three classes of cobordisms, each of which is in spirit a local collapsing of two objects, just like our previous cobordisms.

Consider a periodic curve which borders a region and acts as a source in that region. Consider also a point source in the region and a saddle which connects to both the point source and the periodic curve. Once again, we can continuously deform the two points together and into a neighborhood of the periodic curve, such that the

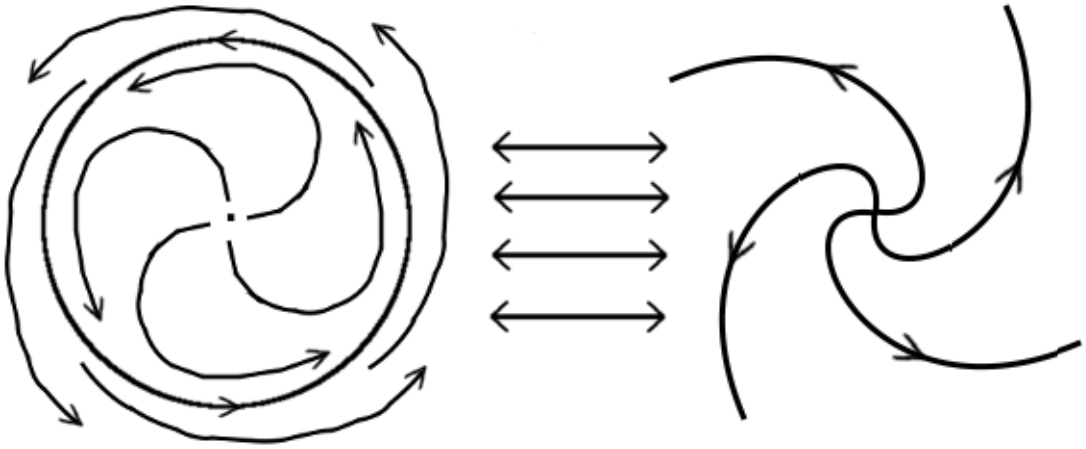


Figure 13. A basic periodic curve cobordism.

periodic curve, when deformed slightly to encompass the two points, continues to act as a source. As above, we can then delete the point source and saddle and treat their integral curves as integral curves of the periodic curve. There is of course a symmetric cobordism with a sink-like periodic curve, a point sink, and a saddle.

These cobordisms provide a mechanism for handling periodic curves, but they still do not allow us to remove periodic curves, which is essential to avoid large sets of cobordism classes. Thus we propose a final cobordism, not quite in the spirit of the others. Given a region of genus 0 bordered by exactly one periodic curve which acts as a source (sink), and containing exactly one zero point, we propose that we can shrink the periodic curve to a point, swallowing the zero point inside, and leaving the remains of the periodic curve to act either as a sink or source depending on the original behavior of the periodic curve outside this region. This allows us to 'swallow' periodic curves provided their interiors are as simple as possible (reduced, effectively). This operation is shown in Figure 13.

Theorem 4. *All structurally stable vector fields on the sphere are cobordant to each other.*

Proof. If there are no periodic curves, we are done as above. We will proceed by induction on the number of periodic curves, assuming there are n periodic curves in the smallest counterexample to the theorem. Certainly, we can always find a region bounded by just one periodic curve. Treating this region itself as a sphere, we note that all cobordisms of the new sphere are proper cobordisms of the original sphere. Therefore we can reduce this to the vector field with just one sink and one source, one of which corresponds to the original periodic curve. Then we can use our final cobordism to remove the periodic curve, thus contradicting the induction hypothesis.

Conclusion

Given an appropriate vector field v on a surface X, we have developed the combinatorial tool of a graph G on X composed of homotopy classes of integral curves of v. By reducing the graph G to just sinks or just sources and then rebuilding, we have shown that G is a triangulation of X. Using this fact, combined with the tools of the sink and source-only graphs, we have demonstrated that vector fields on the sphere may be transformed into each other via basic localized operations which we call cobordisms.

Unfortunately, the presence of periodic integral curves keeps us from generalizing to higher genus surfaces. This motivates a number of open questions:

- Are all primitive vector fields on surfaces of higher genus cobordant?
- How many cobordism classes of generic vector fields exist for surfaces of fixed genus with fixed periodic curve configurations?
- How many cobordism classes of generic vector fields exist for surfaces of fixed genus with fixed numbers of periodic curves?
- What can be said about vector fields which are not structurally stable?

Acknowledgements

The authors would like to acknowledge the generous support of Professor Victor Guillemin, who gave the initial idea for this research and has provided direction for us many times. We would also like to thank John Reed, chairman of the MIT corporation, for funding this research. Finally, we would like to thank the MIT Undergraduate Research Opportunities Program for making undergraduate research like this possible.

References

[1] Lickorish, W.B.R. A _nite set of generators for the homeotopy group of a 2-manifold. Proc. Cambridge Philos. Soc. 60 (1964) 769-778

[2] Peixoto, M. M. Structural Stability on two-dimensional manifolds. Topology 1 (1962) 101-120

Developing Molecular Markers for the Unfolded Protein Response[1]

Josepher Li[1,2], Hyung Don Ryoo[1], Min-Ji Kang[1]

1. Department of Cell Biology, NYU School of Medicine, New York, NY, USA
2. Student Contributor, Class of 2013, Deparment of Aeronautics and Astronautics, MIT
3. Mentor, Department of Aeronautics and Astronautics, MIT

The unfolded protein response (UPR) has many distinct branches. One such branch is mediated by IRE-1 and *xbp1*. The properties of this particular branch of the UPR have previously been used to develop the UPR marker, *xbp1*-EGFP. This marker was created through the use of artificial promoters. Here, it is shown that when *xbp1*-EGFP is fused to its endogenous promoter, it is a more physiological UPR marker. It is demonstrated that the UPR marker, *xbp1*-14-*xbp1*-EGFP, is in fact a viable UPR marker, capable of showing presence of the EGFP epitope when endoplasmic reticulum-stress is experimentally induced by DTT and by the expression of mutant proteins. To facilitate understanding of the UPR in disease, the fruit fly, *Drosophila melanogaster*, was used as a tool. A *Drosophila* model for the UPR was developed by misexpressing a membrane protein, Rhodopsin-1, in the developing eye tissue. These developments of UPR markers and genetic tools serve as a powerful means to examine the connection between the UPR and disease *in vivo*.

Introduction

The endoplasmic reticulum (ER) is a central organelle that is responsible for protein folding maturation. Proteins carry out many important functions, but in order to do so, proteins must be folded into complex three-dimensional structures. However, a general problem in the folding of proteins is that in the process of being folded, the proteins must pass through partially folded states. In these partially folded states, a protein can either fold correctly or misfold by entangling with other molecules [22]. Conditions interfering with the function of the ER that result in these misfolded proteins are collectively called ER-stress [13, 10, 14].

The cell's response to the misfolded proteins in the endoplasmic reticulum is known as the unfolded protein response (UPR) [15, 7, 19]. The UPR helps the ER recover from the accumulation of misfolded proteins by increasing its folding capacity [8] and increasing the degradation of misfolded proteins [18].

Although the UPR can reduce stress in the short term, chronic exposure of cells to ER-stress can result in apoptosis. The loss of vital cells through apoptosis causes many degenerative diseases, such as cystic fibrosis, Huntington's disease, Alzheimer's disease, Parkinson's disease, retinitis pigmentosa, and certain forms of diabetes [23, 4, 11]. Therefore, the ability to detect ER-stress and the UPR may allow for more diverse approaches to developing therapies for these diseases.

The method by which the UPR is activated is well understood in many organisms. In *Saccharomyces cerevisiae*, or yeast, unfolded proteins in the ER cause oligomerization of the ER transmembrane domain protein Inositol-requiring enzyme-1 (IRE-1). Upon oligomerization, activated IRE-1 splices *hac*-1 mRNA. Splicing of *hac*-1 mRNA allows for its translation [17] and its protein product acts as a transcription factor, by binding to DNA motifs collectively called unfolded protein response element (UPRE). By binding to UPRE, the active HAC-1 protein induces the expression of various ER-stress responsive genes like ER-chaperones and genes involved in misfolded protein degradation [12].

IRE-1 plays a similar role in *Drosophila* as it does in yeast. The *Drosophila* genome contains a single homolog of *Ire*-1 [8], which splices X-box–binding protein 1 (*xbp1*) mRNA, a homolog of hac-1 mRNA, upon detection of misfolded proteins [20, 24, 3]. *Xbp1* has two isoforms, RA and RB. The RA isoform contains an extra 23-base sequence compared to the RB isoform. The RA isoform also has the double stem-loop structure required for

IRE-1 splicing [20, 24, 3]. The UPR is initiated through the unconventional splicing of the 23-nucleotide sequence by IRE-1.

In response to ER-stress, IRE-1 undergoes oligomerization, activating its ability to cleave *xbp1* mRNA. Once *xbp1* mRNA undergoes splicing, a reading frame shift occurs and the mRNA is translated into an active transcription factor. The active XBP1 protein binds to UPRE and induces expression of various ER-stress responsive genes involved in misfolded protein degradation and production of ER-chaperones.

Previous research has shown that an *xbp1*-EGFP fusion construct driven at high levels through a heterologous gene expression system called Gal4/uas [2] can detect ER-stress and the activation of the UPR in vivo [18]. In this construct, EGFP was subcloned after the putative IRE-1 splice site, allowing EGFP to be expressed in frame with xbp1 only after the 23-base stem-loop of *xbp1* is eliminated by IRE-1-mediated splicing [18]. While this method worked, it had two potential weaknesses. First, since this system used an artificial promoter, it could not detect any UPR pathways input on the natural *xbp1* enhancer itself. Second, as other studies have shown that too much xbp1 transcripts inhibit the UPR, overexpression of *xbp1*-EGFP through the Gal4/uas system may have weakened the sensitivity of the *xbp1*-EGFP sensor [25]. To improve upon these aspects, the

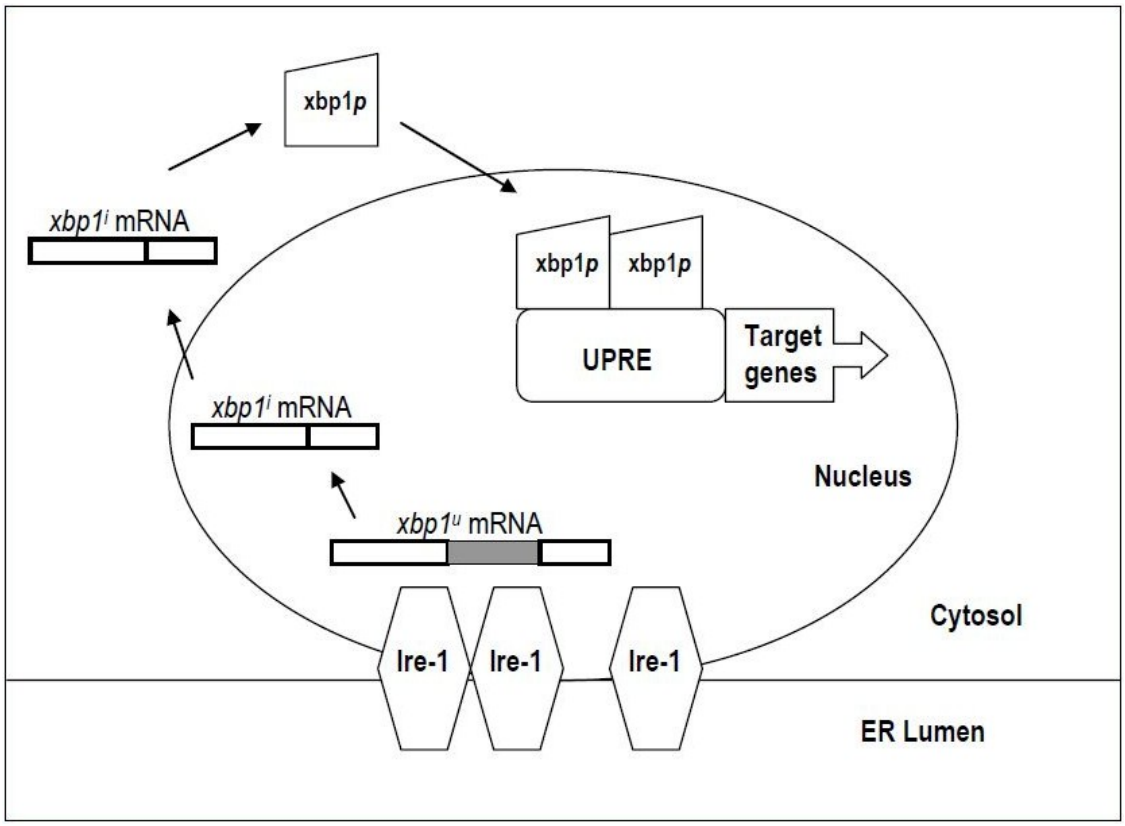


Figure 1. UPR signaling pathway mediated by *Ire-1/xbp1*. Upon detection of misfolded proteins, IRE-1 splices xbp1 mRNA, and though a series of steps, results in the induction of ER-stress responsive genes.

endogenous *xbp1* promoter was found and used to generate an *xbp1*-EGFP sensor that is under the control of its endogenous promoter.

Materials and Methods

Computational genomics

The VISTA set of computational tools [5] was used to locate the putative endogenous regulatory sequence in *xbp1*. The *xbp1* sequence was aligned with the *Drosophila melanogaster* base genome, and compared with sequences of other species of *Drosophila* in order to find regions that are conserved. Regions conserved are expected to have some genetic importance. The region that showed the most conservation was *xbp*1-4, a 0.85 kb region upstream of *xbp1*. Because the region included non-coding DNA that appeared to have much phylogenetic conservation, it is very likely that this region is in fact responsible for *xbp1* gene expression.

Luciferase assay

Upon determining the putative regulatory sequence, a test was done to verify its ability to function as a promoter. To do this, the putative regulatory sequence was inserted into the BglII site of the pGL3-basic vector, a Luciferase Reporter Vector from Promega, and luciferase activities were measured using the Duo-Glo Luciferase Assay System (Promega). To measure firefly and *Renilla* luminescence, the PerkinElmer 2103 EnVision Multilabel Plate Reader was used. As a control, *Drosophila* S2 cells were also transfected with tub-*xbp1*. Results were normalized from experimental samples to control samples repeated in each plate to minimize the impact of variables like temperature, plate order and timing on the ratio of experimental to control reporter activity.

Construct design

To generate the fusion construct in which *xbp1*'s endogenous regulatory sequence,

xbp1-4, is fused to *xbp1*-EGFP, the 0.85 kb promoter region was subcloned into the BglII site of the pattB vector. EGFP-polyA was then subcloned into the same pattB vector using the EcoRI and BamHI restriction sites, to create *xbp1*-4-EGFP-attB. Finally, xbp1 was subcloned into the EcoRI site of *xbp1*-4-EGFP-attB, to create *xbp1*-4-*xbp1*-EGFP. The proper construct was confirmed by using PCR and restriction enzyme digests.

Recombinant DNA and transformation methods

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. Standard recombinant DNA protocols were performed (ref: Molecular cloning: a laboratory manual). During ligation, the ratio of insert to vector used was 9.5 to 0.5 for a single

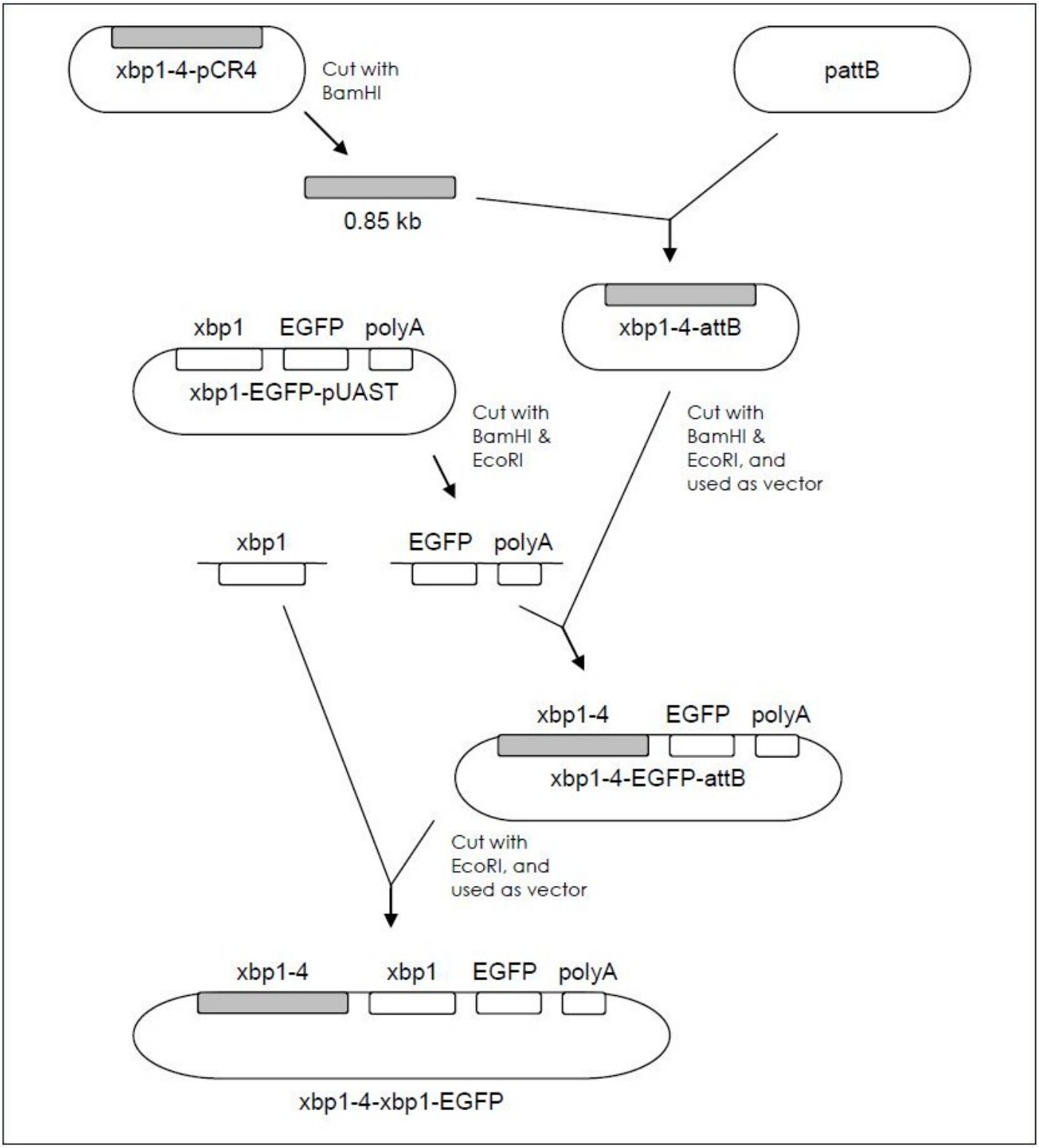


Figure 2. Cloning scheme for *xbp1*-4-*xbp1*-EGFP construct. The *xbp1* endogenous regulatory sequence, *xbp1*-4, was fused with *xbp1*-EGFP.

enzyme, and 7 to 3 for two enzymes. To prevent self-ligation and to improve the yield of properly ligated product when a single enzyme was used, 1 uL of Calf Intestinal Alkaline Phosphatase (CIP, New England Bio Labs, Inc.) was added after enzyme digestion and incubated at room temperature for 10 min. *E. coli* was transformed with plasmid DNA by the $CaCl_2$ method (ref: Molecular cloning: a laboratory manual).

DNA isolation and purification

During subcloning, plasmid DNA was prepared with the QIAprep Spin Miniprep kit. To produce a large volume of the final construct for cell culture and transfection, the QIAprep Midiprep kit was used, according to manufacturer's instructions (Qiagen, Inc.). DNA fragments from agarose gel was purified using the High Pure PCR Product Purification Kit, consistent with manufacturer's recommendations (Roche Diagnostics).

Cell cultures

Drosophila melanogaster Schneider line 2 (*Drosophila* S2) cells were used. The S2 cell line was derived from a primary culture of late stage (20-24 hours old) *Drosophila melanogaster* embryos. These cells were transfected with *xbp1*-4-*xbp1*-EGFP, using the Effectene Transfection Reagent (Qiagen), and analyzed for emergence of the EGFP epitope in the presence of Dithiothreitol (DTT), which blocks disulfide bond formation necessary for the folding of many ER proteins and experimentally causes ER stress. As a control, cells were also transfected with *xbp1*-EGFP and treated with DTT. *Drosophila* Schneider S2 cells were cultured under standard conditions (ref: invitrogen manual).

Western blotting analysis

Standard protocols were used (ref: Molecular cloning: a laboratory manual). Rabbit anti-horseradish peroxidase (HRP) antibodies were used for anti-GFP Westerns. Mouse anti-horseradish peroxidase (HRP) antibodies were used for anti-Profilin Westerns.

Fly mating

Fly stocks were grown at room temperature. All crosses were kept clean by starting with virgin females. To obtain virgin females, flies were collected several times a day, within 8 hours of their emergence as adults. In addition, the fly cultures were placed at 18 degrees Celsius overnight to maximize the number of virgins present in the morning, and virgin females were collected within 16 hours of their emergence as adults. Figure 3 shows a schematic of the crosses needed to establish the fly line exhibiting the eye ablation phenotype that was used in this study. To establish this fly line, which serves as a model for ER-stress-triggered apoptosis in the *Drosophila* eye, the Gal4/uas system was used for ectopic gene expression [2].

Immunostaining

Anti-Rhodopsin-1 monoclonal antibodies were used to detect Rhodopsin-1 and rabbit anti-GFP antibodies were used to detect GFP activation.

Diagrams

Diagrams and graphs were generated with the sole use of Microsoft Word or Excel and their component applications.

Results

xbp1-4 functions as promoter

To measure promoter activity of *xbp1*-4, a reporter gene assay was conducted using the pGL3-Basic luciferase reporter vector. Because the pGL3-Basic vector lacks eukaryotic promoter and enhancer sequences, the putative regulatory sequence *xbp1*-4, can

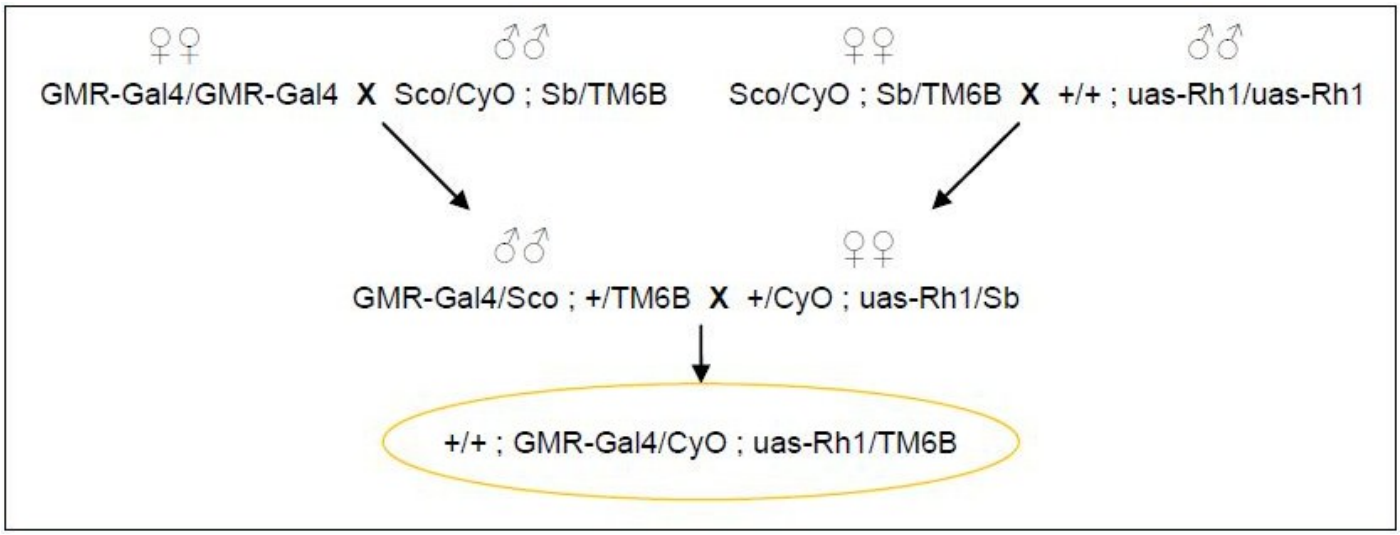


Figure 3. Map of fly cross. Through a series of crosses, a fly line was established in which the flies exhibit an eye ablation phenotype. To create this eye ablation phenotype, the *Gal4*/uas system was used. Flies containing uas-*Rh1* (the target) were crossed to flies expressing *Gal4*. In the progeny of this cross, it was possible to activate uas-*Rh1* in cells where *Gal4* was expressed, in this case the *Drosophila* eye.

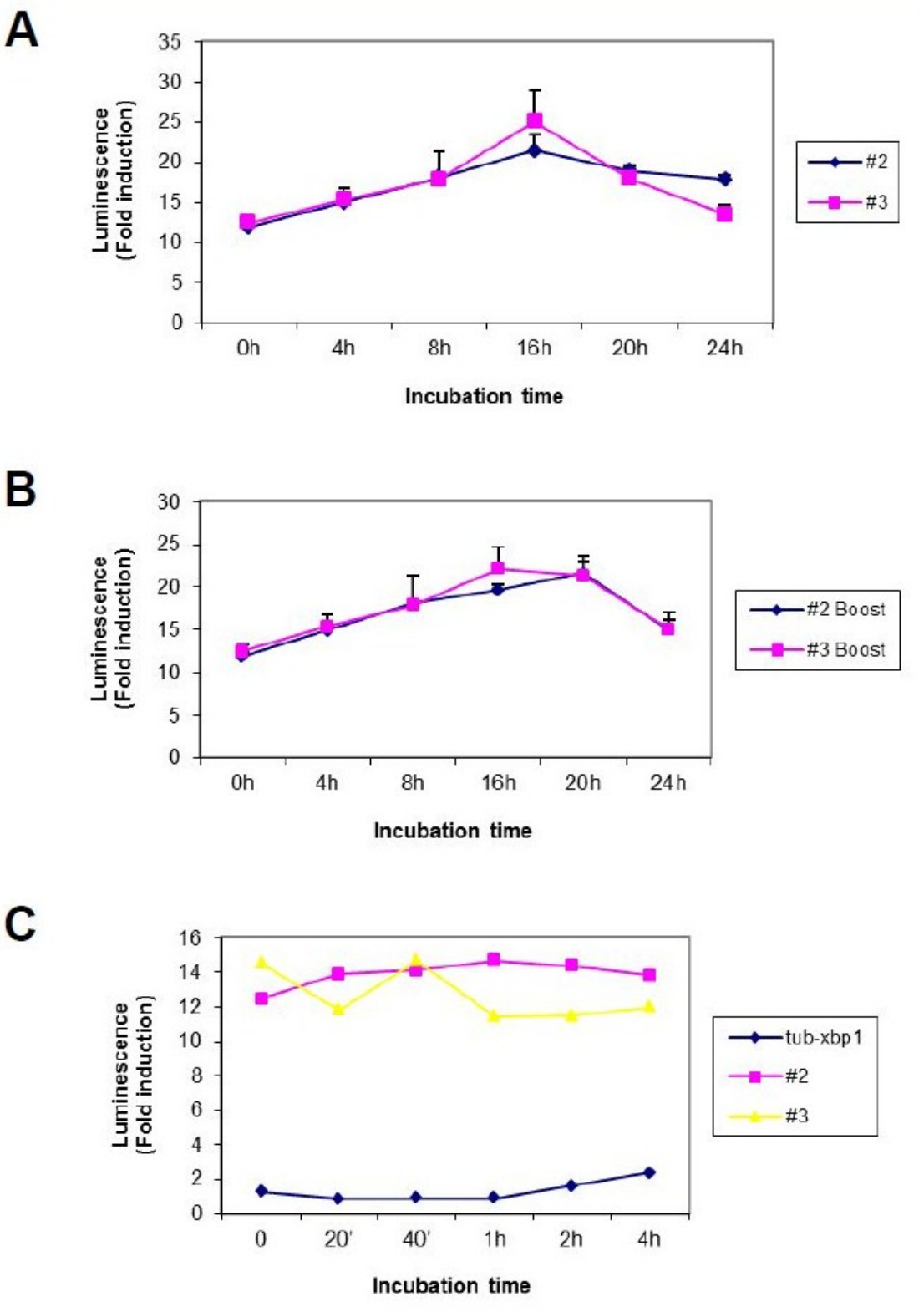


Figure 4. Relative promoter activity of the putative *xbp1* endogenous promoter in transiently transfected *Drosophila* S2 cells. For each transfection, firefly luciferase activity was normalized with *Renilla* luciferase activity to control for transfection efficiency. Fold induction for each transfection is defined as the ratio between induced firefly luciferase activity and basal firefly luciferase activity. **(A)** and **(B)** are results from one experiment. Up to the 8 h incubation time, results are the same in **(A)** and **(B)**. In **(B)**, the 16 h, 20 h, and 24 h incubation times were given an extra boost of 1 mM DTT treatment 16h after the start of the experiment, in order to account for the relatively short half-life of DTT. For *xbp1*-4-pGL3, results for two clones, #2 and #3, are shown. In **(C)**, tub-*xbp1* was used as a control.

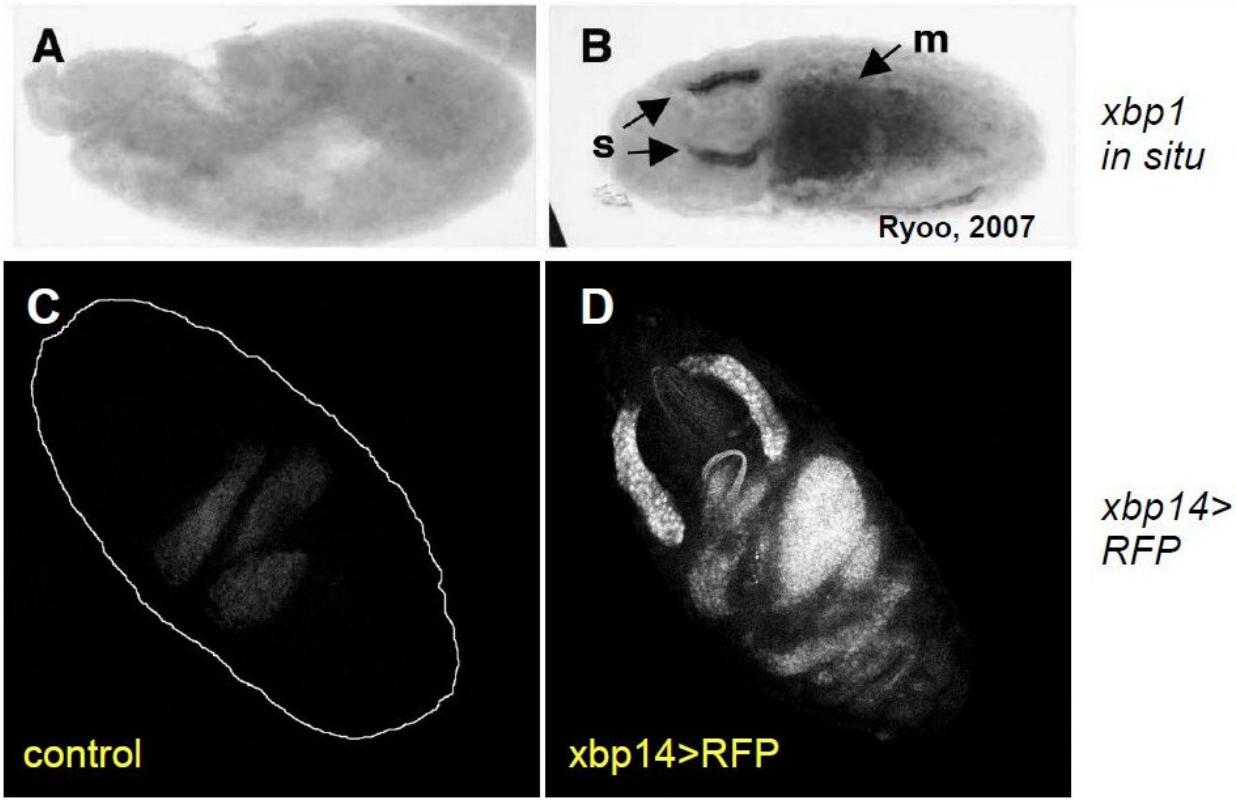


Figure 5. RFP expression similar to that of *Drosophila xbp1*. **(A,B)** In situ hybridization against *xbp1* mRNA in embryos. **(A)** A 6 h-old embryo shows little *xbp1* mRNA. **(B)** A 12 h embryo with arrows pointing to high *xbp1* mRNA in salivary glands (s) and the midgut (m). **(C)** No *xbp1*-4 promoter driving RFP expression. **(D)** *xbp1*-4 promoter driving RFP expression. *Xbp1*-4-rfp pattern is nearly identical to that of *xbp1* mRNA *in situ*, shown in **(B)**.

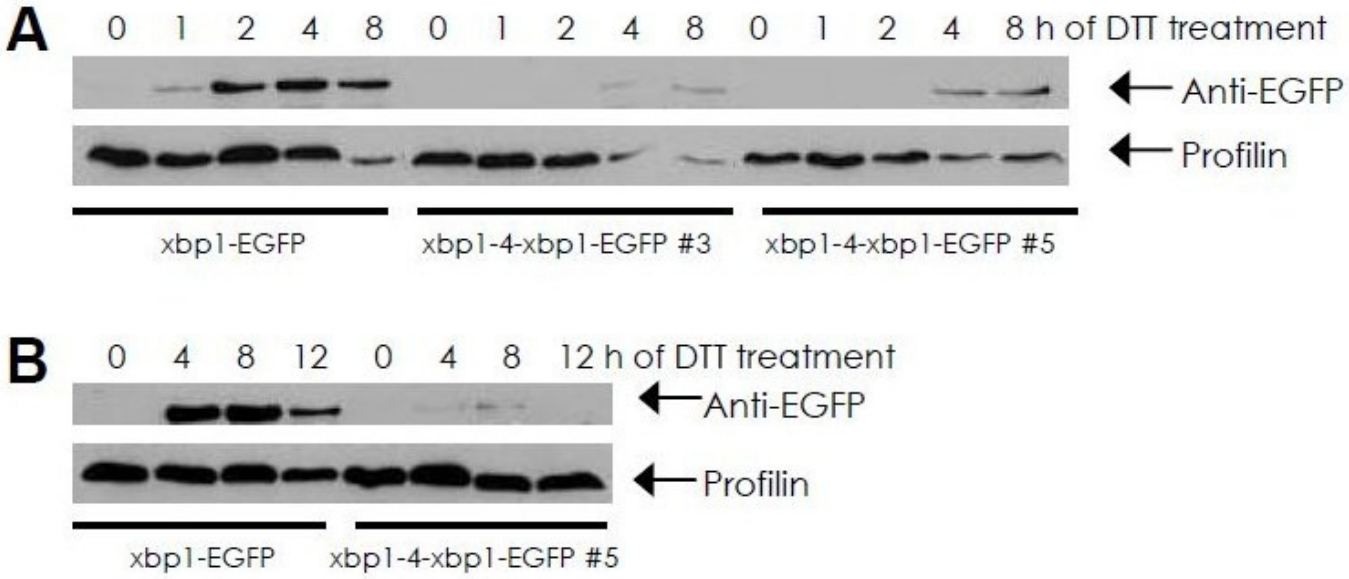


Figure 6. *xbp1*-EGFP marker is activated by IRE-1 in response to DTT treatment in both *xbp1*-EGFP and *xbp1*-4-*xbp1*-EGFP-transfected cells. **(A)** The top panel shows anti-GFP Westerns whereas the lower panel shows anti-Profilin as a control. S2 cells with *xbp1*-EGFP and *xbp1*-4-*xbp1*-EGFP transfection were incubated with or without 1 mM DTT for up to 8 h. In *xbp1*-EGFP-transfected cells, *xbp1*-EGFP activation is seen at 1 h and peaks at 4 h. In *xbp1*-4-*xbp1*-EGFP-transfected cells, *xbp1*-EGFP activation is seen to begin at 4 h. For *xbp1*-4-*xbp1*-EGFP, results for two clones, #3 and #5, are shown. **(B)** The top panel shows anti-GFP Westerns to detect *xbp1*-EGFP activation, whereas the lower panels show anti-Profilin blots as a loading control. S2 cells with *xbp1*-EGFP and xbp1-4-*xbp1*-EGFP transfection were incubated with and without 1 mM DTT for up to 12 h. In *xbp1*-EGFP-transfected cells, *xbp1*-EGFP activation is seen at 1 h and peaks at 4 h. In *xbp1*-4-*xbp1*-EGFP-transfected cells, *xbp1*-EGFP activation is seen at 4 h and 8 h.

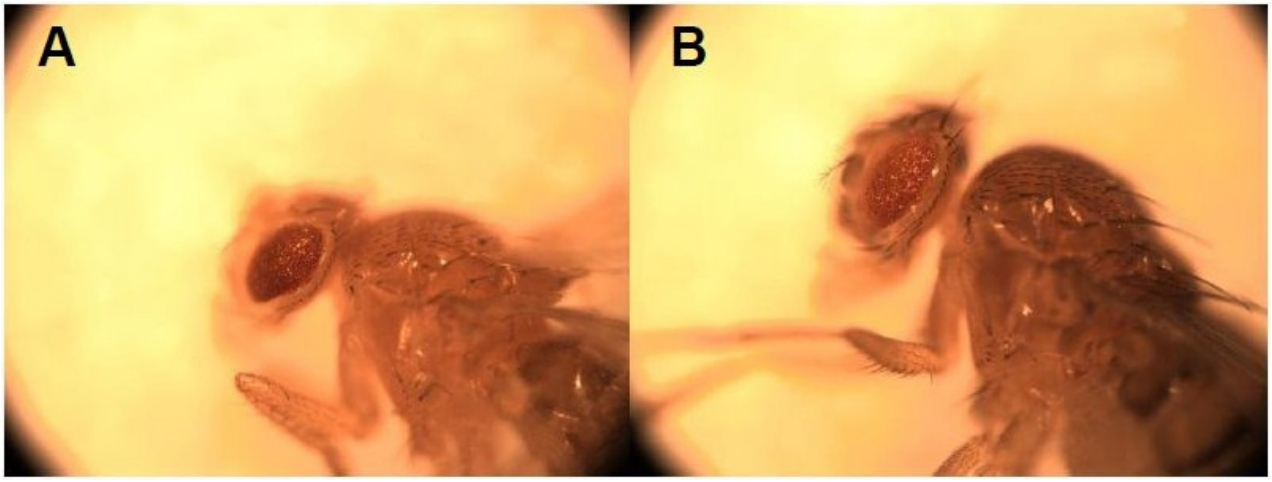


Figure 7. ER-stress triggered retinal cell death in *Drosophila*. **(A,B)** Flies exhibiting the eye ablation phenotype as a result of cell death. Flies such as these can be used to test the ER-stress construct. The fly on the left is male and the fly on the right is female.

be cloned. When *xbp1*-4 was inserted upstream of the *luc*+ gene, and *Drosophila* S2 cells were transfected this plasmid and treated with 1 mM DTT, expression of luciferase activity indicated that *xbp1*-4 is a viable promoter. Firefly luciferase activity was normalized for each transfection with *Renilla* luciferase activity to control for transfection efficiency. Figure 4 shows the relative promoter activity of the putative *xbp1* endogenous promoter in graphical form.

In situ hybridization

To further validate the identity of the xbp1 regulatory sequence, the *xbp1*-4 promoter was used to drive RFP expression. Because the *xbp1*-4-rfp pattern turned out to be nearly identical to that of the *xbp1* mRNA in situ that had previously been reported [18], *xbp1*-4 is in fact the true *xbp1* regulatory sequence. Figure 5 shows both *xbp1 in situ* and *xbp1*-4 driving RFP expression.

xbp1-4-xbp1-EGFP acts as a UPR marker

To test the *xbp1*-4-*xbp1*-EGFP construct and examine its properties in *Drosophila* cells, *Drosophila* S2 cells were transfected with the construct and cultured under standard conditions (ref: invitrogen manual). These cells were then analyzed for the emergence of the EGFP epitope in the presence of DTT. When DTT is added, *xbp1* undergoes IRE-1 mediated splicing, and only then can *xbp1*-EGFP be expressed in frame [18]. *xbp1*-EGFP was used as a control by cotransfecting Actin-*Gal4* and uas-*xbp1*-EGFP into cells. Through the uas-*Gal4* method, *xbp1*-EGFP expression can be driven. As shown in Figure 6 and in previous studies [18], when 1 mM DTT was added to the media of *xbp1*-EGFP transfected cells, *xbp1*-EGFP was activated within an hour, with the amount of this marker peaking by 4 hours (Figure 6). When 1 mM DTT was added to the media of *xbp1*-4-*xbp1*-EGFP-transfected cells, *xbp1*-EGFP was activated within 4 hours (Figure 5). Activation of *xbp1*-EGFP when *xbp1*-4-*xbp1*-EGFP-transfected cells were exposed to DTT confirmed that *xbp1*-4 does in fact function as a UPR marker. Shown in Figure 6 are results from two clones, #3 and #5. Unlike *xbp1*-EGFP, however, *xbp1*-4-*xbp1*-EGFP shows presence of the *xbp1*-EGFP marker only at 4 and 8 hours, with the amount of this marker greater at 8 hours. The *xbp1*-EGFP marker was not detectable at 12 hours (Figure 6B).

Establishing a model for ER-stress in Drosophila

Drosophila melanogaster was used to facilitate understanding of the UPR in disease. Flies were generated that misexpressed the membrane protein, Rhodopsin-1 (Rh-1), in the developing eye tissue. Because Rhodopsin-1 did not fold properly in the endoplasmic reticulum of the fly eye primordia, the UPR was activated. Stress caused by this misfolding in the endoplasmic reticulum led to an eye ablation phenotype. This eye ablation phenotype is in the process of being used to perform a large scale genetic screen for modifiers of this phenotype, including those genes that reduce stress in the endoplasmic reticulum. Currently, approximately ten genes have been found to be candidates.

In vivo analysis

Evidence that *xbp1*-4-*xbp1*-EGFP acts as a UPR marker was confirmed through cell culture studies. To further develop *xbp1*-4-*xbp1*-EGFP as an *in vivo* marker, this construct was used to develop a transgenic fly harboring this marker. Crossing the transgenic flies with the flies misexpressing Rhodopsin-1 (Figure 7) produced flies that were be used to test whether *xbp1*-4-*xbp1*-EGFP is functional *in vivo*. Because EGFP is activated in cases of ER-stress and the UPR, crossing these two flies resulted in the presence of green fluorescence in *Drosophila* larvae eye imaginal discs, as detected by immunostaining (Figure 8).

Discussion

Understanding the unfolded protein response (UPR) and its many branches is significant because of its implications in a variety of neurodegenerative diseases. In order to examine the role of the UPR in diseases, finding reliable methods with which to detect the UPR is necessary. Here, a UPR marker, *xbp1*-4-*xbp1*-EGFP, is developed, and its ability to detect ER-stress and the UPR is demonstrated.

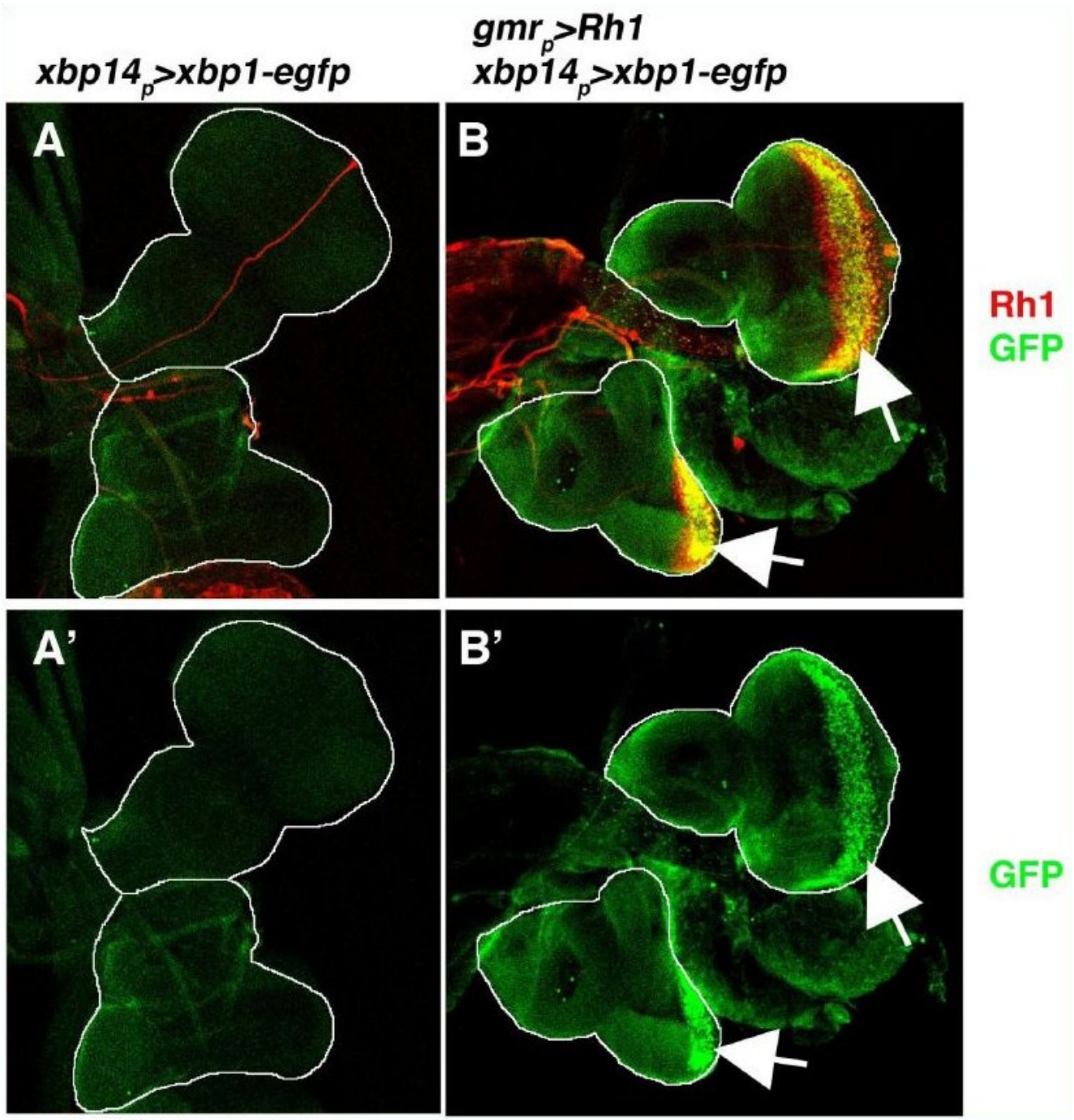


Figure 8. *In vivo* analysis of UPR marker. The *xbp1*-4-*xbp1*-EGFP marker is activated in response to mutant Rh-1 in Drosophila larvae eye imaginal discs, as indicated by arrows. In all panels, eye imaginal discs are outlined in white. **(A)** *xbp1*-4 promoter driving *xbp1*-EGFP expression in eye discs. As a control, there is no GMR promoter driving Rh-1 expression **(A')** The GFP channel only. **(B)** Rh-1-expressing eye discs with *xbp1*-4 promoter driving *xbp1*-EGFP expression shows strong *xbp1*-EGFP activation. **(B')** The GFP channel only.

A faithful UPR marker

UPR markers previously developed showed GFP fluorescence in response to ER stress. However, when inserted into *Drosophila* and mice, these constructs did not display significant reporter activity during development [18, 9], which was unexpected since UPR target genes are expressed at high levels in many tissues during development [18]. This suggests that the previous UPR markers might not have been sensitive enough to detect low but functionally relevant levels of xbp1 splicing. In particular, the *xbp1*-EGFP fusion construct previously tested may not be as sensitive since *xbp1*-EGFP is artificially overexpressed in the system.

Our in vivo analysis confirmed that the *xbp1*-4-*xbp1*-EGFP marker works not only in cell culture studies, but also *in vivo*. Using the *xbp1*-4-*xbp1*-EGFP construct, it is possible to detect *xbp1* splicing in Drosophila eye imaginal discs, as well as in cells treated with DTT, which artificially causes ER-stress. Because here *xbp1*-EGFP is driven by its endogenous promoter, as opposed to being driven at high levels through heterologous promoters, the *xbp1*-4-*xbp1*-EGFP construct has shown to be more physiological than the previous UPR markers.

Moreover, while previous studies have focused on the unconventional *xbp1* mRNA splicing as the primary mechanism of UPR regulation, our research suggests that this pathway is heavily regulated at the level of *xbp1* transcription. In many unstressed tissues, xbp1 transcription is dormant, as shown in Figure 5D, where *xbp1*-RFP was not expressed universally. However, upon ER-stress, *xbp1*-luciferase is induced, indicating that mechanisms other than IRE-1-induced *xbp1* mRNA splicing assist in the UPR pathway.

Paving the way to understanding another branch of the UPR

It has already been established that the transcription factor *xbp1* activates the transcription of ER-stress responsive genes by binding to ER stress response elements within their promoter regions [24] after *xbp1* mRNA is spliced by IRE-1 [20, 24, 3]. Another branch of the UPR consists of the transcription factor ATF6 [6]. In order for ATF6 activation to occur, the transcription factor must be translocated from the ER to the Golgi apparatus in order to be cleaved by site 1 and site 2 proteases (S1P and S2P) [21].

Although many believe that ATF6 has an essential role in the UPR, this has yet to be validated by gene silencing. Using the *xbp1*-4-*xbp1*-EGFP construct developed here, it is possible to see if ATF6 RNAi treatment blocks *xbp1*-EGFP activation. The inhibition of *xbp1*-EGFP activation through ATF6 RNAi treatment would be a good indication that ATF6 does in fact play an essential role in the UPR. Moreover, while ATF6 cleavage is known to be a factor in *xbp1* transcription, other unknown components involved in *xbp1* transcriptional regulation may exist. The identification of a small *xbp1* regulatory DNA sequence would allow for the establishment of assay systems to help identify new mechanisms involved in UPR.

Acknowledgements

I thank Prof. Hyung Don Ryoo for introducing me to the topic of the unfolded protein response and allowing me to conduct research in his laboratory, and Dr. Min-Ji Kang for her time and patience in teaching me the laboratory skills needed to complete my project. I also thank Richard Lee for his comments and suggestions on the manuscript.

References

1.Armknecht, S., Boutros, M., Kiger, A., Nybakken, K., Mathey-Prevot, B., Perrimon, N. (2005) High-throughput RNA interference screens in Drosophila tissue culture cells. Methods Enzymol 392: 55–73

2.Brand, A.H., Perrimon, N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118: 401–415

3.Calfon, M., Zeng, H., Urano, F., Till, J.H., Hubbard, S.R., Harding, H.P., Clark, S.G., Ron, D. (2002) IRE1 couples endoplsmic reticulum load to secretory capacity by processing the XBP-1 mRNA. Nature 415: 92–96

4.Forman, M.S., Lee, V.M., Trojanowski, J.Q. (2003). Unfolding pathways in neurodegenerative disease. Trends Neurosci. 26(8): 407-410

5.Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.D., Dubchak, I. (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Research 32: 273-279

6.Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., Ron, D. (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol Cell 6: 1099–1108

7.Harding, H.P., Calfon, M., Urano, F., Novoa, I., Ron, D. (2002) Transcriptional and translational control in the mammalian unfolded protein response. Annu Rev Cell Dev Biol 18: 575–599

8.Hollien, J., Weissman, J.S. (2006) Decay of endoplasmic reticulum localized mRNAs during the unfolded protein response. Science 313: 52–53

9.Iwawaki, T., Akai, R., Kohno, K., Miura, M. (2004) A transgenic mouse model for monitoring endo¬plasmic reticulum stress. Nature Med 10: 98 102

10.Kaufman, R.J., Scheuner, D., Schroder, M., Shen, X., Lee, K., Liu, C.Y., Arnold, S.M. The unfolded protein response in nutrient sensing and differentiation. Nat Rev Mol Cell Biol 3:411–421, 2002

11.Marciniak, S.J., Ron, D. (2006). Endoplasmic reticulum stress signaling in disease. Physiol. Rev. 86(4): 1133-1149

12.Mori, K., Ogawa, N., Kawahara, T., Yanagi, H., Yura, T. (1998) Palindorme with spacer of one nucleotide is characteristic of the cis-acting unfolded protein response element in Saccharomyces cerevisiae. J Biol Chem 273: 9912–9920

13.Mori, K. Tripartite management of unfolded proteins in the endoplasmic reticulum. Cell 101: 451–454, 2000

14.Oyadomari, S., Araki, E., Mori, M. Endoplasmic reticulum stress-mediated apoptosis in pancreatic β-cells. Apoptosis 7: 335–345, 2002

15.Patil, C., Walter, P. (2001) Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Curr Opin Cell Biol 13: 349–355

16.Pichaud, F., Desplan, C. (2001) A new visualization approach for identifying mutations that affect differentiation and organization of the Drosophila ommatidia. Development 128: 815–826

17.Ruegsegger, U., Leber, J.H., Walter, P. (2001) Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell 107: 103–114

18.Ryoo, H.D., Domingos, P.M., Kang, M.J., Steller, H. (2006) Unfolded protein response in a Drosophila model for retinal degeneration. EMBO J 26: 242-252

19.Schroder, M., Kaufman, R.J. (2005) The mammalian unfolded protein response. Annu Rev Biochem 74: 739–789

20.Shen, X., Ellis, R.E., Lee, K., Liu, C.Y., Yang, K., Solomon, A., Yoshida, H., Morimoto, R., Kurnit, D.M., Mori, K., Kaufman, R.J. (2001) Complementary signaling pathways regulate the unfolded protein response and are required for C. elegans development. Cell 107: 893–903

21.Shen, J., Snapp, E.L., Lippincott-Schwartz, J., Prywes, R. (2005) Stable binding of ATF6 to BiP in endoplasmic reticulum stress response. Mol Cell Biol 25: 921-932

22.Thomasson, W.A. Unraveling the mystery of protein folding. Breakthroughs in Bioscience. Accessed online.

23.Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. Science 267, 1456–1462 (1995)

24.Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., Mori, K. (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell 107: 881–891

25.Yoshida, H., Oku, M., Suzuki, M., Mori, K. (2006) pXBP1(U) encoded in XBP1 pre-mRNA negatively regulates unfolded protein response activator pXBP1(S) in mammalian ER stress response. J. Cell Biology 172(4):565-575

Web Abstracts

Find more articles online at murj.mit.edu!

Using Chemical Spectroscopic Analysis to Develop an Ultraviolet/Infrared Heat Mirror

Mic Byrne, Class of 2014, MIT

This experiment aims to find a variety of inexpensive chemicals which can reflect and absorb ultraviolet and infrared light. These chemicals would be used to create "heat mirrors" which could be engineered to function for many practical uses, such as cooling solar panels and preserving energy in the temperature control of buildings. The intent was to find a chemical that can transmit 30% of the incoming infrared and ultraviolet light. Based on research and careful analysis of MSDS information, a final list of chemicals was formed: ZnO, TiO2, and CaOH. Although only three of many, these chemicals demonstrated sufficiently different properties to provide a wide base for analysis. Suspensions were created using different concentrations of these materials. The suspension was applied to a glass slide and the solvent was evaporated, leaving a thin, even layer of solute. These slides were run through the infrared spectrometer and the ultraviolet-visible spectrometer. Although no efficient way to measure whether the 30% transmission goal was achieved, transmission spectrums were generated for the different chemicals. These results can be used in future research when deciding which chemicals to use to block or transmit different wavelengths of light.

Heat mirrors can be used to heat buildings efficiently without wasting energy.

Demobilization, Demilitarization and Reintegration in Iraq

Cody Zoschak, Class of 2013, Department of Political Science

Part of the standard operating procedure for any foreign intervention is a systematic program of Demobilization, Demilitarization, and Reintegration (DDR). This is intended to neutralize armed groups such as militias without a costly and drawn out military campaign, which will rarely end successfully. In years following the power vacuum left by Saddam Hussein's Baath party, the majority of the power in Iraq was held by leaders of various armed militias such as Muqata al-Sadr, yet the Coalition Provisional Authority (CPA) did not take serious action to neutralize these militia. In a unstable area of operations such as Iraq, the first priority must always be to establish a secure environment, after law and order has been restored, nation building can take place. The CPA rushed through the steps and attempted to restore services to Iraqis and establish governmental structures before they had pacified the country. DDR is a key part of this pacification, especially if one takes into account the enormous number of soldiers that became unemployed when Saddam's army was dissolved. This paper will examine the steps that were taken to neutralize the militias and other armed groups, and will also note where DDR was absent and speculate as to the damage that was done.

World-changing technologies. Life-changing careers.

Imagine your career here.

Sandia is a top science and engineering laboratory for national security and technology innovation. Here you'll find rewarding career opportunities for the Bachelor's, Master's, and Ph.D. levels in:

- **Electrical Engineering**
- **Mechanical Engineering**
- **Computer Science**
- **Computer Engineering**
- **Systems Engineering**
- **Mathematics, Information Systems**
- **Chemistry**
- **Physics**
- **Materials Science**
- **Business Applications**

We also offer exciting internship, co-op, post-doctoral, and graduate fellowship programs.

Learn more >>

www.sandia.gov

Sandia National Laboratories

Operated By

LOCKHEED MARTIN

A Best Diversity Company Chosen By The Readers Of Diversity/Careers 2010

Sandia is an equal opportunity employer.
We maintain a drug-free workplace.