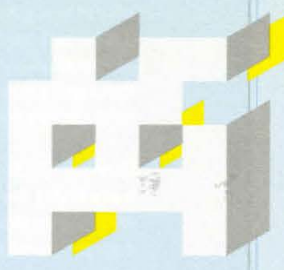
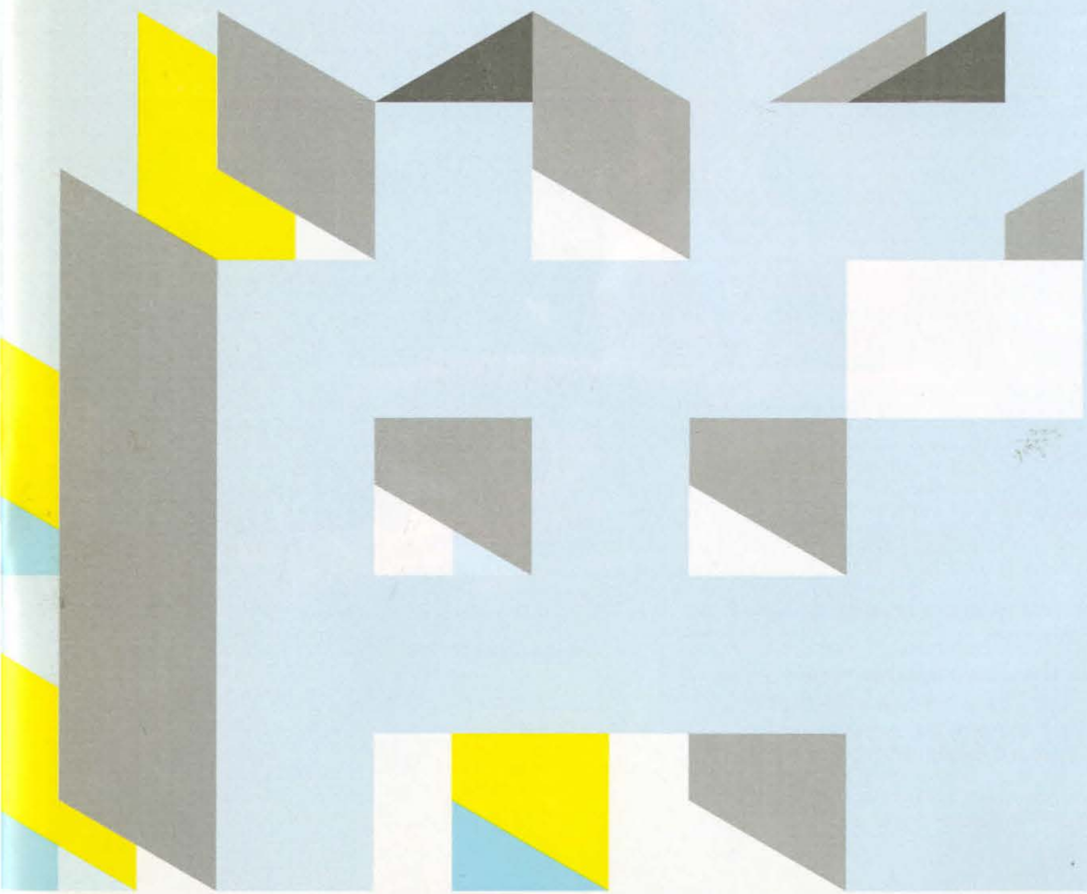
CURJ

FALL 2003
VOL.3 NO.2

CALTECH UNDERGRADUATE RESEARCH JOURNAL

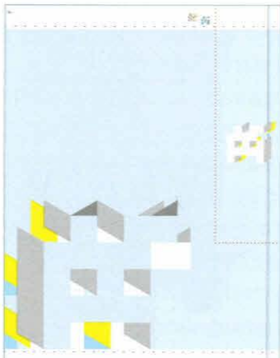


BREAK!
EXECUTING STRATEGIES WITH ROBOT TEAMS



Changing The Way You See Things™
DuPont™ olight®
At DuPont Displays we see displays differently: Not as mere afterthoughts or component parts, but as the defining feature of most every device – the "face" through which people connect with technology.
DuPont™ Olight® displays will transform the way people interact with information. These displays capitalize on the power of OLED to make products that are more vivid, efficient, responsive, and light. DuPont™ Olight® displays are more than just screens, they enhance the performance of the entire device – from cell phones to consumer and industrial electronics.
Because of our unique view, we look for unique people – people who want to make a difference in a leading edge technology and be part of a team that makes our business successful. We are looking for candidates with skills and knowledge in the following areas:
• Material Science
• Process Engineering
• Mechanical Engineering
• Device Science
• Electrical Engineering
• Electronic Engineering
• Marketing
Science
Marketing
Engineering
If you would like to know more about our business or to apply online, visit our website at http://olight.dupont.com.
DuPont is a fortune 500 company with excellent benefits. We are proud to be an EEO/AA employer. If you would like to become a member of our team, please send your resume to Employment, DuPont Displays Inc, 120 Cremona Drive, Suite C, Santa Barbara, CA 93117-5511.
No phone inquiries please.
DUPONT
The miracles of science™
© Copyright 2003 E.I. du Pont de Nemours and Company. All rights reserved. The DuPont Oval Logo, Dupont™, The miracles of science™, Olight®, and Changing The Way You See Things™ are trademarks or registered trademarks of E.I. du Pont de Nemours and Company.

CURJ



LETTERS

REVIEWS

RESEARCH

FINIS

Cs *Computer Science*
A *Astronomy*
Be *Bioengineering*
C *Chemistry*
Ec *Economics*
G *Geology*

FROM THE EDITOR

Six issues after the Caltech Undergraduate Research Journal was established with the goal of showcasing some of the best undergraduate research performed, CURJ has become one of the premier undergraduate research journals in production. Combining captivating research essays with stunning graphic design, each issue of the journal now reaches thousands of students worldwide at over three dozen universities and libraries across four continents.

Our growing distribution network is the backbone of our outreach to collegiate programs interested in the development of future researchers. We are now working toward strengthening our ties to undergraduate research programs in the United States and overseas in order to encourage collaboration in the growing undergraduate research community. It is our hope that in addition to publishing a journal that is both entertaining to read and composed of the highest quality undergraduate research, young scientists can also use CURJ to expand their research interests and to better perceive research opportunities at other universities.

The publication of a journal dedicated to the arena of undergraduate research relies on motivated students interested in entering the vigorous publication process. We publish writing from undergraduates worldwide and encourage interested students to submit their work for review. Please consider contributing to CURJ and let us know what you think of our journal.

Philip Wong
Editor-in-Chief

CALTECH UNDERGRADUATE RESEARCH JOURNAL VOL.III NO.2

Original CURJ Design by Aniko Hullner Grau

www.curj.caltech.edu

ArtCenter College of Design

SPACE AND EARTH EXPLORATION: 2020 VISION

BY CHARLES ELACHI

Forty five years ago, mankind took a great leap forward (and upward) with the launches of Sputnik (USSR) and Explorer (USA). The Explorer satellite developed at Caltech's JPL started a very active American space and earth science program that has led to amazing intellectual and utilitarian benefits ranging from telecommunication, Global Positioning Systems, global topographic mapping, Earth hazard and resource monitoring, robotic exploration across the solar system, and new insights in the formation and evolution of the universe and the laws of physics that govern it.

As we started the new century, we saw major new advances and highly visible setbacks, but also a strong determination to expand our knowledge articulated by the NASA mission: to understand and protect our home planet, to explore the universe and search for life in it, and to inspire the next generation of explorers as only NASA can.

How are we planning to accomplish these goals and where do we envision being in 2020? The following are some illustrative examples.

A few months ago, JPL completed the processing of a global digital topography data set acquired by the 11-day Shuttle Radar Topography Mission (SRTM) jointly sponsored by NIMA (National Imaging and Mapping Agency) and NASA. This data set provided digital height measurements of the whole earth land mass between ± 60° latitude to about 5 meter height accuracy and at about 25 to 90 meters posting. One cartographer recently called this is the most important breakthrough in cartography since Gerald Mercator's projection more than four centuries ago. With a single breakthrough, we developed a whole new way to provide three dimensional views of our planet with more than an order of magnitude improvement relative to present stereo and ground techniques developed more than 100 years ago. Over the next decade, the combination of interferometric space radar and GPS techniques will enable us to measure and map, globally and continuously, surface motion to better than a centimeter, thus enabling a whole new way of studying the dynamic of our planet's surface. By 2020, a new field of space tectonics will reach a level of maturity similar to that achieved by space oceanography in the last 20 years.

As we develop more capable telescopes, sensitive detectors and arrays, and picometer meteorology systems, we will be able to view the early "days" of the formation of the universe and observe it not only via the electromagnetic energy it emits but also via the gravitational waves resulting from extremely violent phenomena. The recently launched Spaceborne Infrared Telescope (SIRTF); the planned James Web Space Telescope (JWST); a number of ongoing and planned cosmic background telescopes; UV, X-ray, and gamma ray telescopes; and gravitational wave "antennas" will allow us to go back in time and view the "life history" of our universe in significant detail.

The search for life will not only focus on direct detection of present or past life, but also on exploring environments which are amenable to biological activities so we can understand how life evolved and how common (or rare) it is in our universe, how common it is that simple molecules somehow get together to form complex molecules, then cells, then the intelligent beings reading this article (and writing it). By 2020, we should have surveyed at least the neighboring 1000 stars, acquired "family portraits" of their surrounding planetary systems, analyzed the spectra of their individual planets and determined their size, composition, temperature, etc. Closer to home, we would have fully explored the bodies in our solar system which might be favorable for life (Mars, Europa, Titan, comets), established permanent robotic scientific stations on the surface and around Mars and laid out the groundwork for human expeditions to our neighbor in the solar system.

I see the inspiration of the next generation of explorers every day. When I see the sparkle in the eyes of the thousands of school children that visit JPL, when I see the smile on the face of the blind touching a Braille encyclopedia of astronomy, the excitement of a young engineer watching "her or his" rover moving on the surface of Mars or when a "mature" scientist grabs me in the Mall at JPL with a huge grin on his face and child-like excitement to show me the first images from SIRTF (true story).

Yes, failures and tragedies do occur. This is most unfortunate, and they should be avoided, but in any endeavor where you are exploring new worlds with more and more sophisticated machines, you have to push the limit and, every once in a while, you will stumble. But that is what exploration is all about. The most important thing is to never fear the risk of pushing the limit and expanding your knowledge because the payoff is worth it.

Charles Elachi (Ph.D '71, Caltech) is the Director of the Jet Propulsion Lab and Professor of Electrical Engineering and Planetary Science at Caltech. He is a member of the National Academy of Engineering.

THE PARADOX OF CHRONIC PAIN MANAGEMENT

BY THOMAS CHELIMSKY

The management of chronic pain in our society epitomizes the fundamental incongruities of our health care system. The problem is overwhelming, both from a human suffering and an economic perspective. Clinically effective and financially viable medical solutions are available. Yet the incentives that drive healthcare insurers in today's healthcare market cause them to deny payment for these services, which cost no more than a gallbladder operation.

Pain is one of the few disorders for which Congress has considered direct legislation (H.R. 1863: to place research, education and treatment of pain among our national public health priorities). It was the cover story of the May 19th 2003 issue of *Newsweek*. Costs are staggering, with unconventional treatments alone accounting for several billion dollars expended by patients out of pocket each year, and US business and industry losing about $90 billion annually to sick time, reduced productivity, and direct medical and other benefit costs. Prevalence estimates range widely from 2% to 40% of the population, with cautious estimates suggesting 10% prevalence, and a more recent meta-analysis estimating severe chronic pain at 11%. Approximately 23.5 million are disabled (based on data from the National Institute on Disability and Rehabilitation Research), with pain directly causative in 75%. The high toll of pain in the United States has prompted the Joint Commission on Accreditation of Health Care Organization (JCAHO) to require all accredited organizations, whether inpatient or outpatient facilities, to develop and implement programs that screen for and effectively manage pain.

Despite such extensive commentary and potential legislation, the facts regarding optimal treatment remain poorly disseminated, not only in the lay press, but even in the medical community. Neither medical schools nor training programs afford any significant education regarding chronic pain management. Unquestionably, the interdisciplinary pain program is demonstrably superior to every other strategy directed at treating chronic pain. This holds true for an array of different carefully examined outcome measures, including pain reduction, increased physical activity, rate of return to work, management of depression and emotional distress associated with pain, and particularly, overall health care utilization. In addition, when one factors in lifetime disability and productivity effects, the savings related to interdisciplinary pain programs approach $2.5 billion savings over twenty years, being an astonishing 9 times more cost-effective than conventional medical treatment.

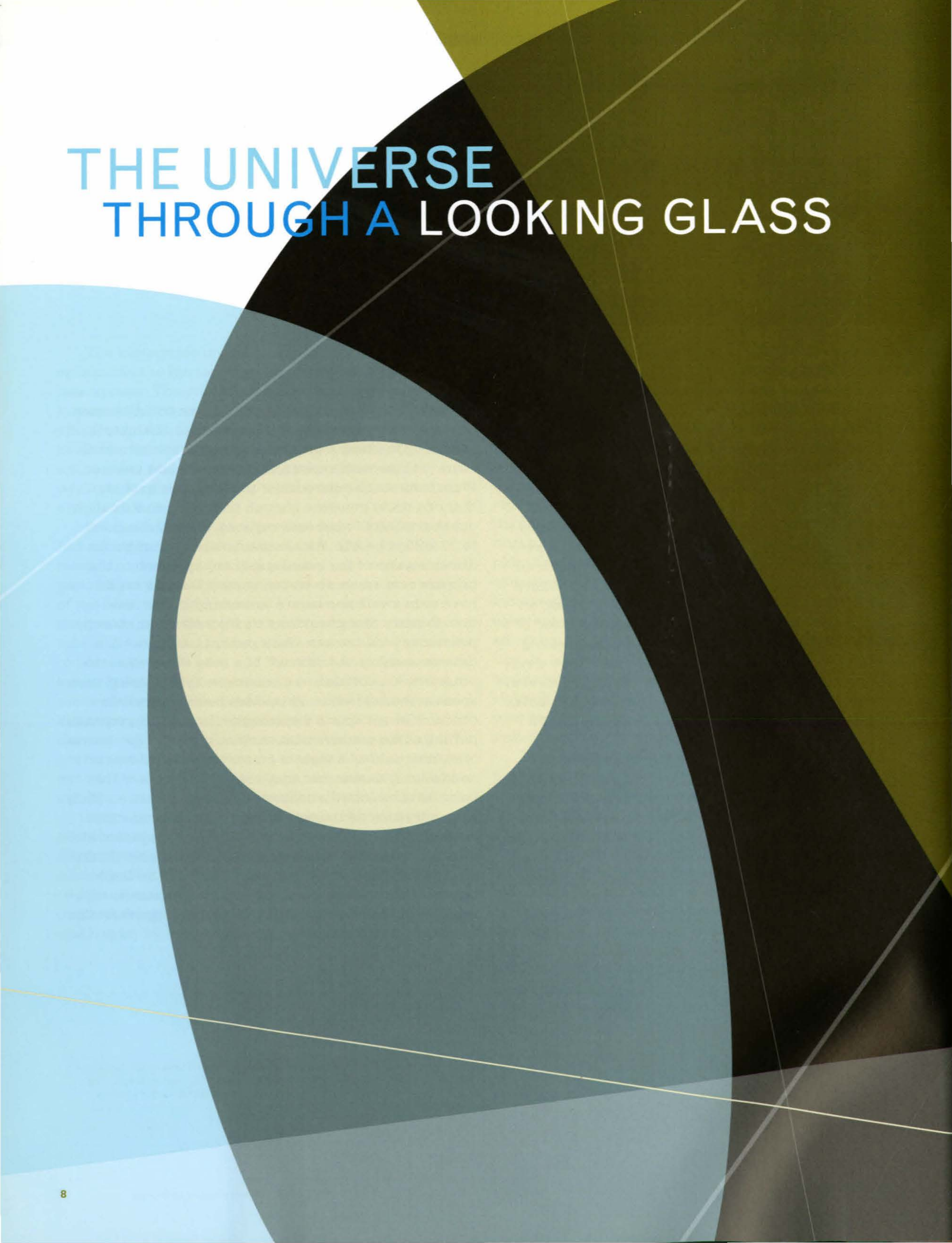
What is an interdisciplinary program? It is a 3 to 6 week period of coordinated, intensive (8 hours/day, 5 days/week), physical, psychological and behavioral training combined with medical treatment. An interdisciplinary team, composed of the patient, occupational and physical therapists, a psychologist, and one or more physicians, manage the program. The result is physical reconditioning, return of life functions such as work and family, empowerment toward a sense of freedom and hope, pain reduction, sleep restoration, and remission of the depression that usually accompanies chronic pain. More than 2/3 of patients will return to productive employment after such a program, the real source of savings to our society. These programs run anywhere from $2500 to $3500 per week, depending on geographic location and the services included, with total cost of $8000 for the least expensive 3-week program, and $22,000 for the most expensive 6-week program.

Even with such impressive evidence that the interdisciplinary treatment approach is effective, the Commission on Accreditation of Rehabilitation Facilities (CARF) lists only 85 accredited pain centers nationwide, and the number continues to dwindle. Assuming that only half the programs seek accreditation, and that the average pain program can treat 100 patients per year in an intensive program, at most 20,000 patients can receive interdisciplinary programmatic treatment in a year, an insignificant portion of the 30 million people in need of services. In addition, insurers are less and less willing to support this avenue of care, because they have a very short-term interest in their enrollees (18 months on average); they usually do not cover the disability payments; or the system offers no economic incentive for health insurers to optimize work productivity.

At this point, if we wish to address the issue of chronic pain at all, two logical approaches come to mind, given the facts regarding optimal treatment. New incentives, either financial or legal, can encourage insurance companies to pay for appropriate treatment. For example, health insurance and disability insurance could become linked in some way, so that the burden of disability is shared by the health insurer, and causes the insurer to "care" for the outcome of healthcare, not just the cost. Were insurers to reverse their thinking, it is unlikely that one could generate enough pain programs throughout the country to increase capacity from 20 thousand to 30 million people. An alternative strategy might be dissemination of the interdisciplinary approach to the primary care arena. However, several barriers would have to be overcome before embarking in this direction. Primary care physicians traditionally shun chronic pain for several reasons. Such patients are viewed as time-consuming and difficult, at a time when economic pressures are certainly not conducive to increasing time spent at medical visits. Physicians have appropriate concerns about opiate dispensation (actually a very small portion of the management of chronic pain). The absence of chronic pain as a topic in current medical education and training means that they may be unaware of the benefits of an interdisciplinary approach, and most likely have not received training in running an interdisciplinary team. This approach would require an experimental pilot regional program. Since the traditional post-graduate education lecture format has essentially no impact on practice, critical components of this program would include experiential learning and clinical support for the formation of interdisciplinary teams.

Thomas Chelimsky, M.D., is Associate Professor of Neurology at Case Western Reserve University, Cleveland, Ohio. He was Director of the University Pain Center until its closure at the end of 2003. He is also Director of Clinical Autonomic Disorders and the Autonomic Laboratory at University Hospitals of Cleveland.

THE UNIVERSE THROUGH A LOOKING GLASS



BY BEN GRANETT

A PHOTON TRAVELING TO EARTH FROM A GALAXY billions of light-years away usually has an uneventful trip. Mostly, it travels through intergalactic space, rarely running into anything. But if it happens to pass by another galaxy or other high mass object, it will get pushed around quite a bit. The gravity from a star, for instance, can bend the trajectory of a beam of light, just as it bends the trajectory of an orbiting planet. This phenomenon is known as gravitational lensing.

Conventional lenses use refraction to bend light. A fishbowl filled with water, for example, acts as a lens refracting light through the air-glass-water interfaces. When you observe a fish through the glass of a round bowl, its image is warped. Gravitational lensing, on the other hand, is a consequence of the curvature of space and is understood through the general theory of relativity. It is a subtle effect, but is observed over astronomical distances. Stars or galaxies act as gravitational lenses, magnifying and distorting the images of objects behind them.

Gravitational lensing has become a powerful tool for studying some of the greatest unsolved problems in astronomy. Researchers use the phenomenon to look for dark matter based on its gravitational effects, as well as to measure the expansion rate and age of the universe.

It was Einstein who originally discovered gravitational lensing in 1911 using only mathematical logic and physical principles. Immediately, he sought confirmation by experiment. Since it would be impossible to test the prediction in a laboratory, Einstein considered the nearest object with the strongest gravitational field: the Sun. He calculated that a light ray passing very near to the Sun would bend through a tiny angle, 1.74 arc seconds, or less than a thousandth of a degree [FIGURE 1]. This would be a very slight change in trajectory, but astronomers could measure it.

Detecting background stars in the glare of the Sun is impossible, but such an observation would be possible during a total solar eclipse. In those few minutes of darkness, background stars appearing very near the Sun could be photographed. If Einstein were correct, the positions of the stars during the eclipse would be different from their positions when measured at night.

Einstein presented his general theory of relativity to the Royal Prussian Academy of Sciences in 1915. World War I prevented direct communication of the discovery to British scientists, but news leaked quickly. Arthur Stanley Eddington, the director of Cambridge Observatories, received copies of the paper and recognized the importance of Einstein's work. He immediately promoted

"Stars or galaxies act as gravitational lenses, magnifying and distorting the images of objects behind them."

the theory and sought means to confirm it. The astronomer Sir Frank Watson Dyson realized that the May 29, 1919 total solar eclipse would provide the ideal opportunity. A cluster of stars, known as the Hyades, would pass near the Sun during this eclipse providing many bright stars to measure. The path of totality would cross the Atlantic Ocean from Brazil to West Africa. Dyson planned two expeditions to observe the eclipse, one to the island of Principe, off the coast of Spanish Guinea in West Africa, and one to Sobral in northern Brazil.

Both teams faced unexpected weather hardships and equipment problems. The best observations came from photographs taken through a modest telescope with a 4-inch objective lens. After careful analysis of the images, Eddington noted that the positions of the stars differed from their usual positions by precisely 1.74 arc seconds. The experiment immediately turned the 40-year-old Einstein into an international celebrity. Einstein himself, in a postcard to his mother, expressed his excitement: "...joyous news today. H.A. Lorentz telegraphed that the English expeditions have actually measured the deflection of starlight from the Sun."

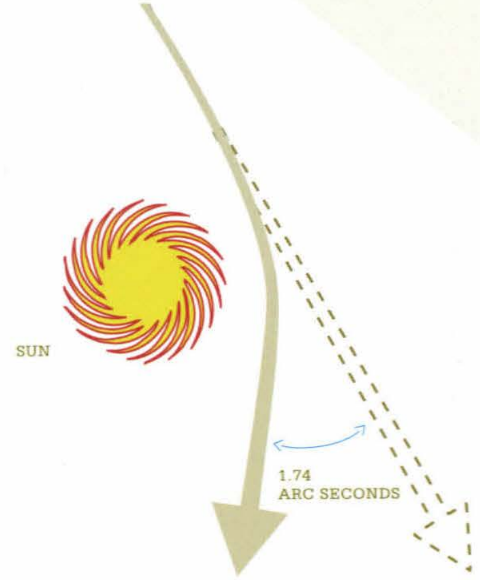


FIGURE 1 A light ray passing the Sun will be bent through an angle of 1.74 arc seconds by the gravitational field. The trajectory is shown by the solid arrow. Without the Sun, the light ray would follow a straight path shown by the dotted arrow. For comparison, the Sun is about one half of a degree in diameter or 1800 arc seconds in the sky. The angle is greatly exaggerated in the diagram.

RELATIVELY SPEAKING

The physics behind gravitational lensing is general relativity, a field invented and developed by Einstein at the beginning of the 20th century. Relativity described the dynamics of space, time, and matter in a completely new way that demolished the foundations laid down by Isaac Newton in the 17th century.

Newton viewed the universe as a giant machine ticking at a constant rate and functioning in a regular way. Through the application of the laws of physics, Newton believed that all future states of the machine could be predicted. Newton was very successful; with a few simple laws he could describe the motion of billiard balls on a table or calculate the positions of the planets orbiting the Sun.

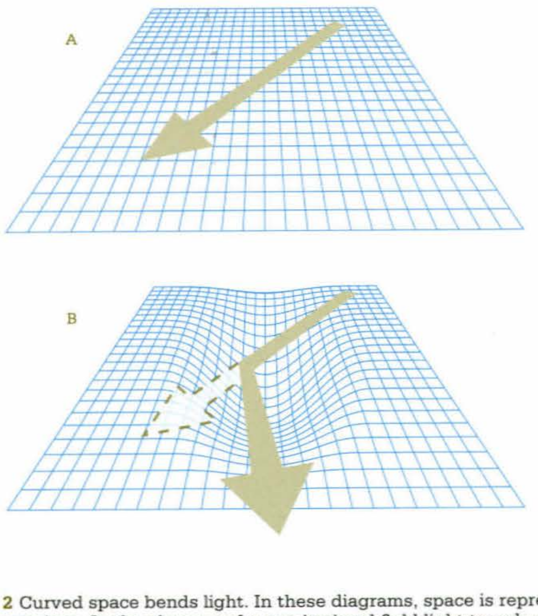
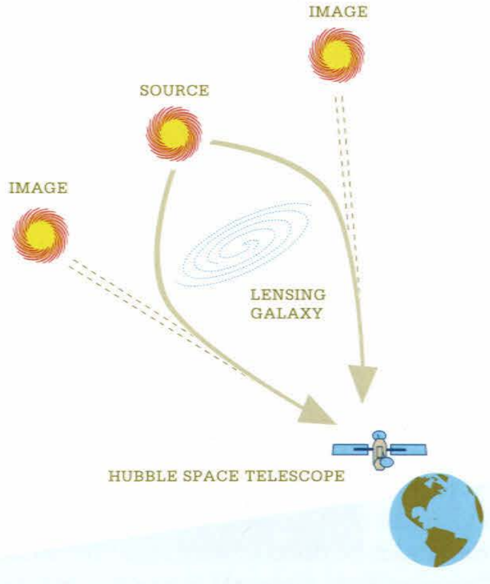


FIGURE 2 Curved space bends light. In these diagrams, space is represented in two dimensions. In the absence of a gravitational field light travels along a straight path (A). A gravitational field produced by a massive object will bend the trajectory of the light ray (B) like a ball rolling along a warped table.

> "If Einstein were correct, the positions of the stars during the eclipse would be different from their positions when measured at night."

Einstein changed this concept of the universe. Instead of a static, steady ticking, clockwork machine, Einstein proposed a dynamic universe. Space itself, Einstein claimed, responded to the matter it contained. Matter warps space, just as an acrobat warps the surface of a trampoline. Not only is the geometry of space uncertain, but time itself, Einstein postulated, is not always measured consistently. The simple laws of motion suddenly became extremely complex. In Einstein's universe, everything depends on the observer: what you see in the universe relative to yourself is different from what I see relative to myself.

How does this cause light to bend? In the context of relativity, gravity is not considered to be a force but is a consequence of the warping effect that matter has on space. A massive object warps space causing other objects to fall into the indentation. In this way, gravitation is an effect of the geometry of space. Light, bound by space, must also follow the curvature. As light travels, the rays follow the contours of space, and thus the trajectory bends in the same way that the Sun bends a planet's path [FIGURE 2]. Although the concept is abstract, astronomers have beautifully observed the bending of light by gravity to an accuracy of nearly 0.1 percent.



COSMIC GRAVITATIONAL LENSES

The shifts in star position measured by Eddington were only a few thousandths of a degree. More dramatic are the deflections of light by galaxies with masses a trillion times the Sun's mass and millions of light years away. Galaxies act as extreme lenses, focusing the light of distant sources. The distortions produced depend upon the alignment of Earth, the lensing galaxy, and the background object being lensed. In a perfect alignment, the background source appears as a ring around the lensing galaxy. This ring is named the Einstein ring. If the alignment is not perfect, we see multiple images of the background source due to the various paths the light can take to reach us, as illustrated in FIGURE 3. In intermediate cases, the images formed can be distorted into arcs.

Arcs are a tell-tale sign of a gravitational lens, but identification is often not so straightforward. In picture-book cases, symmetry reveals a gravitational lens. For instance, if there are four images, they would be positioned at the corners of a diamond with the real object in the center. However, in complicated systems in which there are multiple components to the lens, the symmetry can be difficult to find. Most importantly, identification of gravitational lenses requires analysis of the spectra, or chemical signatures, of the images. Each image of the source always shows identical chemical features in its spectrum.

FIGURE 3 Gravitational lenses produce mirage images of distant objects. This figure diagrams two light rays (solid arrows) originating from a distant source. As the light heads towards Earth, it does not follow a straight path, but the lensing galaxy bends its trajectory. From our perspective on Earth, it looks like the light is originating from two separate objects, and traveling along straight paths (dashed arrows). In this situation, we see two images of the same source.

The first gravitational lensing effects besides those produced by the Sun were not discovered until 1979 when a distant quasar was imaged twice by an intervening galaxy [**FIGURE 4, frame A**]. Quasars are galactic nuclei powered by massive black holes. Because they are extraordinarily luminous, quasars can be seen from great distances. Dennis Walsh, Bob Carswell, and Ray Weymann serendipitously discovered the first lens outside our solar system from Kitt Peak National Observatory in Arizona. The pair of quasars was immediately considered peculiar because of the similarities of the two objects. In the original paper, the discoverers stated, "The two sources show great similarity in their spectral characteristics. A conventional interpretation could regard as coincidence the similarity of emission spectra. ... A less conventional view would find the quasars to be two images of the same object produced by a gravitational lens." It turned out that their suspicions were correct. New observations have revealed the intervening galaxy, confirming that the two images actually represented one quasar split by a gravitational lens.

Since that initial discovery, many other gravitational lenses have been identified. Today, one of the leading tools in their discovery is the Hubble Space Telescope orbiting Earth. Above Earth's atmosphere, the Hubble Space Telescope is able to obtain much higher resolution observations than ground-based telescopes, which are affected by atmospheric blurring. In one survey completed by the Hubble, astronomers discovered ten new gravitational lenses over an area the size of the full moon.

MICROLENSING: SEARCHING FOR THE INVISIBLE

Gravitational lenses come in two sizes, micro and macro. Macrolenses are high-mass galaxies or clusters of galaxies composed of billions of stars; microlenses are compact, low-mass objects, such as a single star or even a Jupiter-sized planet. Both types of lenses produce the same imaging effects, but to different degrees. While macrolenses can reveal large objects in other parts of our universe, astronomers detect microlenses within our own galaxy. Stars are in constant motion and as they cross paths along our line of sight, they produce short-lived microlensing events. The star in back will appear distorted as it is lensed by the foreground object. Multiple images can form, but they will overlap and astronomers simply observe a brightening of the lensed star.

Microlensing events occur randomly and cannot be predicted. Yet despite the great difficulties in actually catching them, there are many reasons to try. With microlensing, astronomers can detect objects that may be impossible to detect in a more direct way. For instance, free-floating, Jupiter-sized planets in our galaxy are difficult to observe because they do not emit their own light. Generally they can only be detected via their gravitational effects. If a planet moves in front of a star, the star is lensed and the star may appear to brighten from our perspective on Earth for a short period while it and the planet are aligned. As the planet moves away, the star dims to its normal luminosity. The change in brightness of the star occurs in a predictable way. A number of ongoing research programs monitor the sky for microlensing events. Since microlensing events are extremely rare, millions of stars must be monitored over a period of years to get useful results. Using the 50-inch telescope on Mount Stromlo in Australia, the MACHO (Massive Compact Halo Objects) group has monitored 18 million stars since 1992 and has identified 55 microlensing events. The results of this and other groups' research suggest that hard-to-detect objects, like free-floating planets and faint brown dwarf stars, may contribute significantly to the mass of the galaxy.

> "In one survey completed by the Hubble, astronomers discovered ten new gravitational lenses over an area the size of the full moon."

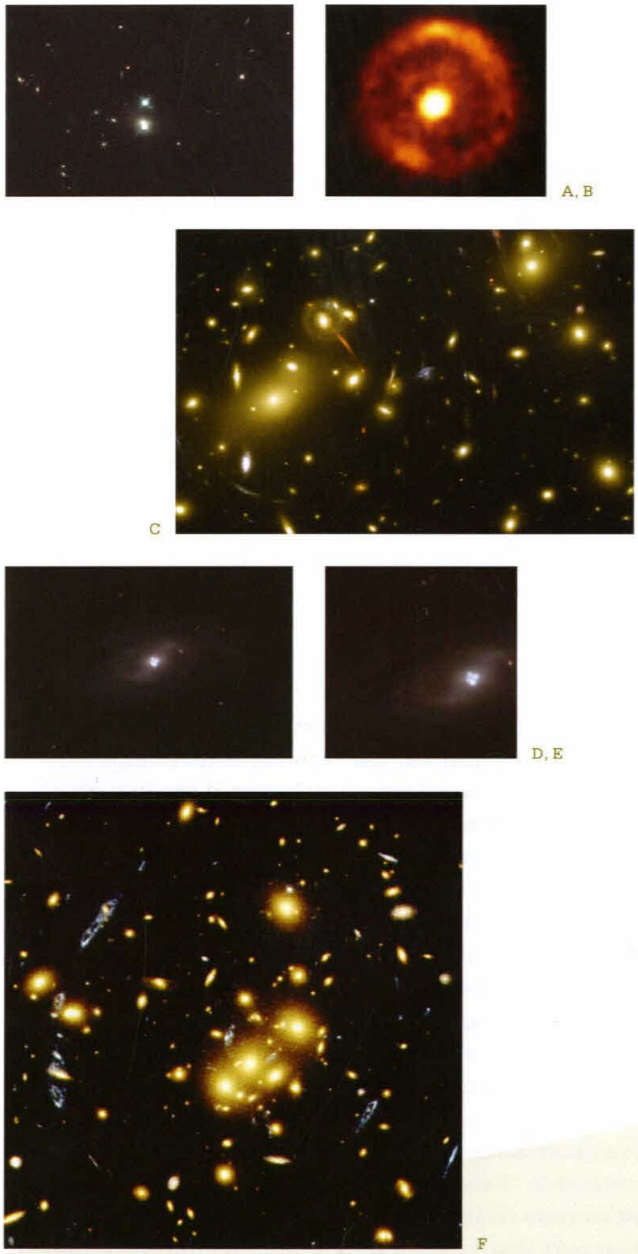

A, B
C
D, E
F

FIGURE 4 A gravitational lens zoo. (A) shows the first gravitational lens discovered by Walsh and collaborators in 1979. It is a distant quasar lensed by a nearby galaxy. (B) shows an Einstein ring. (C) is an example of arcs. (D) and (E) show symmetries. (F) shows a galaxy cluster made up of yellow galaxies lensing a distant blue galaxy. The blue galaxy is multiply imaged and distorted into arcs. *Sources: Fisher 1997; J. Rhoads, S. Malnotra, I. Dell'Antonio, N. Sharp, WIYN telescope at Kitt Peak National Observatory, NASA, Hubble Telescope, and the Space Telescope Science Institute.*

UTILIZING GRAVITATIONAL LENSES IN COSMOLOGY

Cosmology is concerned with answering the grandest questions in nature, such as, "How old is the universe?", "Where did it come from?", "Where is it going?", and "What is its shape and size?". Extraordinary progress in physics has been made over the past few decades to the point where astronomers can make consistent estimates of the size, shape, mass, and age of the universe. Gravitational lenses promise to be an important tool in further solving these problems.

A number of ongoing and planned experiments should end debate on such vital issues as the geometry and rate of expansion of the universe. It is well established that the universe is expanding as a result of its origins in the Big Bang. Edwin Hubble first quantified the expansion in the 1920s, yet today the rate of expansion is not well known. Different methods of measurement arrive at answers that disagree by up to 30%.

As a consequence of the Big Bang, space is expanding and all galaxies (ignoring local gravitational attractions) are being carried along for the ride. As space expands, every galaxy moves away from every other galaxy, as if they were points on an expanding balloon. From Earth, we observe all galaxies receding from us, but observers elsewhere in the universe would see the same sight. The velocity of the recession is proportional to the distance and the rate of expansion; that is, we see distant galaxies moving away from us faster than nearer galaxies. By measuring the rate at which a galaxy recedes, astronomers can infer the distance of the galaxy from the Earth.

"With microlensing, astronomers can detect objects that may be impossible to detect in a more direct way."

Velocities of galaxies are measured using a Doppler effect, the redshift. Light emitted from a source moving away from Earth is shifted to lower energies, or toward the red portion of the spectrum. Blue light is shifted toward red; red light is shifted toward infrared; and so on. The result is similar to the drop in pitch of a receding siren. By measuring the color of light coming from a galaxy, astronomers can find the redshift and infer the galaxy's recessional velocity and its relative distance from Earth. The fundamental problem is the conversion of redshift into true physical distances. To make the conversion, the expansion rate must be known.

Not only is the expansion rate necessary to measure distance, it is needed to find the age of the universe. To find the age of the universe from the rate of expansion, the universe is merely "run backwards" and deflated to find the time when the universe was of zero size.

Gravitational lenses provide an excellent tool for measuring the expansion rate because they give a direct measurement of distance. First, the relative distances between Earth, the lens, and the source can be determined by measuring the redshifts. Then, the true distance can be found from the time delay between the multiple images, if it can be measured. In a system that is not symmetric, the rays of light that follow different paths to the observer will travel different distances. This results in what may seem like a strange effect: the multiple images will show the source at different times. Imagine an alien civilization viewing multiple images of Earth through a gravitational lens. One image might show Earth in the year 2000, while another image might take longer to arrive and would show Earth in the year 1900, entertaining (and possibly scaring) the aliens with a convenient rerun of two world wars!

If the time delay between images and the geometry of the system are known, the true distances can be derived. Time delays are easy to measure if the source object shows regular variability. FIGURE 5 shows an example. This gravitational lens system (B0218+357) is one of the few that have been measured precisely enough to obtain a time delay. The system consists of a distant quasar lensed by a foreground galaxy. Two images are produced and the distance the light travels for the two images is different. Consequently, variations in the brightness of the quasar are seen in one image before the other. The time delay has been estimated at 10.5 days. Using the speed of light, this time can be translated into a distance.

Few gravitational lenses have given good distance measurements. Few sources show variability, without which the time delay cannot be found. Most multiply imaged subjects are frustratingly static, and no estimate can be made. Even if a source is regularly variable, the prerequisite asymmetry must be met so that the time delay between the images is appreciable. Finally, when a good candidate is discovered, the hundreds of observations necessary to track the time delay are difficult to make accurately.

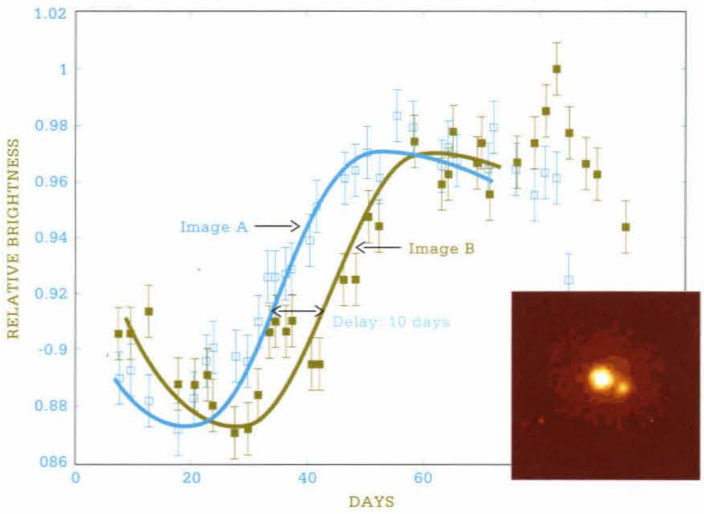


FIGURE 5 This gravitational lens is a distant quasar split into two images (A and B) by a foreground galaxy. The brightness of the quasar fluctuates in time and astronomers have measured the variations in brightness of the two images (blue and green data points on the plot). The two images represent two paths of light of different length: the light in image A gets to us faster than the light in image B. Thus fluctuations are seen first in image A and later in image B. The time delay is about 10 days and this can be used to calculate distance. *Sources: Biggs 1999 and the Space Telescope Science Institute.*

Even when the time delay for a source can be found, the troubles are still not over. A good distance approximation requires an understanding of the gravitational field of the lens. When the lens is fairly simple, such as a single galaxy, the gravitational field can be estimated well. However, in most cases, the lens is made up of contributions from many galaxies, as in a galaxy cluster, and the mass in the lens cannot be well traced. Modeling the lens is a challenging problem that requires both theory and good observations.

In a few cases, gravitational lenses have provided enough information to estimate the expansion rate, and thus the age of the universe. However, estimates vary—data from lens system Q0957+561 results in the universe estimated at 16 billion years old, lens system PG1115+080 gives an estimate of 18 billion years, and the measurements from system B0218+357 lead to an estimate of 14 billion years. Although the current predictions are not consistent and do not surpass the accuracy of other techniques, the ideal gravitational lens could potentially beat the competition. The race is on to find the perfect lens.

GRAVITATIONAL LENSES AS ASTRONOMICAL TOOLS

After formulating his theory of general relativity, Einstein realized the scientific potential of cosmic gravitational lenses, but, noting the technology of the day, he remarked, "there is no great chance of observing this phenomenon." Yet, with today's telescopes, gravitational lenses are within astronomers' reach. Microlensing studies are providing clues to the identity of dark matter in our galaxy, and on a cosmological scale, observations of gravitational lenses are suggesting estimates for the age and expansion rate of the universe. C

Ben Granett is a fourth year undergraduate in Astronomy at the California Institute of Technology. He would like to thank Dian De Sha for her helpful insights on the clarity and style of the paper and Professor Roger Blandford for sharing his expertise.

FURTHER READING

1 Blandford, R.D. and R. Narayan. Cosmological Applications of Gravitational Lensing. *Annual Review of Astronomy and Astrophsyics* **30**, 311–358 (1992).

2 Coles, P. Einstein, Eddington and the 1919 Eclipse. *Proceedings of International School on the Historical Development of Modern Cosmology, ASP Conference Proceedings* **252**, Ed. V.J. Martinez, V. Trimble, M.J. Pons-Borderia. 21 (2001).

3 Williams, L.L.R. and P.L. Schechter. The Quest for the Golden Lens. *Astronomy and Geophysics* **38**, 10–14 (1997).

MICROARRAYS
A MACRO SOLUTION

BY FATIH OZSOLAK

WHEN SOL SPIEGELMAN AND DAVID GILLESPIE PUBLISHED their paper describing the first successful quantitative assay for DNA-RNA hybridization in 1965, they had no idea what their discovery would later produce. DNA-RNA hybridization, the process by which single strands of DNA and RNA are combined to form a double helix, was one of the first nucleic acid technologies explored by scientists. Many later developments were based on the principles established by Spiegelman and Gillespie. A significant breakthrough in the 1970s was the invention of southern blot hybridization, a method for detecting the presence of DNA fragments in a nucleic acid sample. Subsequent advances in DNA sequencing, robotics, and computer technology in the early 1990s led to the emergence of a novel technology, microarray analysis—a protocol which promises to revolutionize biomedical research more than any technique introduced in the past forty years.

Although microarrays initially evolved as a nucleic-acid-based technology, recent developments have allowed scientists to utilize them for the efficient analysis of thousands of tissue samples and proteins. These important advances have expanded the application of microarray technology to many other fields.

WHAT EXACTLY ARE MICROARRAYS?

A microarray, also known as a "chip," consists of thousands of tissue samples or molecules of known identities that are fixed in an orderly fashion on a glass, silicon, or nylon substrate. They are classified based on the substance arrayed—DNA, tissue, or protein.

DNA microarrays contain thousands of DNA probes deposited on a solid support so that the exact location of each sample is known. There are two main types of DNA microarrays: cDNA (complementary DNA) and oligonucleotide. cDNA microarrays are mainly used to compare gene expression profiles of a sample of interest. cDNA is synthesized from an RNA template strand by an enzyme called reverse transcriptase. The resulting cDNA is a single-stranded molecule complementary to the specific RNA molecule from which it was created. In this approach, high-speed robots position copies of the synthesized cDNA (termed cDNA clones) on a substrate to produce a cDNA microarray containing up to 50,000 different cDNA fragments. Subsequently, the arrays undergo a comparative hybridization reaction: a sample of interest, such as diseased tissue, is compared to a reference sample, generally normal tissue. RNA is isolated from the two samples and labeled with different fluorescent dyes during their reverse transcription into cDNA. Both labeled cDNA strands are seeded onto the same microarray in equal amounts [FIGURE 1]. Fluorescence measurements are then performed for both of the dyes, and intensity ratios can be calculated to determine relative gene expression levels. This set of fluorescence intensity measurements is called a gene expression profile.

On the other hand, oligonucleotide arrays contain gene-specific oligonucleotides (short chains of DNA or black RNA) that are either synthesized on the array *in situ* or synthesized *in vitro* and deposited onto arrays by high-precision robots. This technology allows the production of chips containing up to 100,000 probes, providing a method for analyzing gene expression. The preparation of RNA for hybridization to oligonucleotide chips is somewhat different from the process used to prepare cDNA samples. An RNA fragment of interest is first reverse transcribed to single-stranded cDNA, which is then converted into double-stranded cDNA. This double-stranded DNA is transcribed *in vitro* to cRNA (complementary RNA) in the presence of biotin-laced ribonucleotides. Biotin, a common biological marker molecule, can be detected easily by scanning techniques. Thus, the amount of nucleic acid present can be quantified by analyzing the presence of biotin. Each biotin-labeled cRNA is then hybridized individually to oligonucleotide chips and hybridization signals are measured using a laser scanner. Unlike cDNA microarrays, the labeling reaction involved in working with oligonucleotide arrays takes place during the *in vitro* transcription step and samples are hybridized to independent oligonucleotide arrays.

cDNA and oligonucleotide microarrays are also used for large-scale genotype determinations. One example is the identification of single nucleotide polymorphisms, or SNPs, which are single base pair positions in genomic DNA where individuals in a given population show different sequence alternatives. Researchers associate SNPs with differences in disease development among patients. Another example is the measurement of gene copy number changes, alterations in the repetition of all genes in a cell's genome, by a method called Comparative Genomic Hybridization (CGH) of DNA microarrays [FIGURE 2].

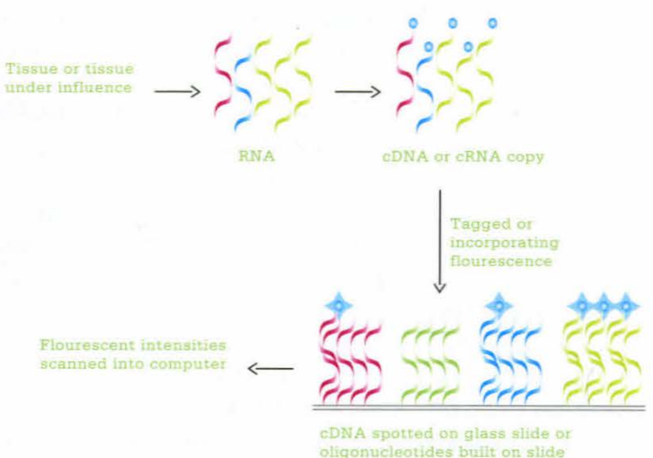


FIGURE 1 A typical approach to microarrays first involves isolating RNA from biological samples. Subsequently, numerous cDNA or cRNA copies are made and a fluorescent marker is incorporated (light blue stars in diagram). These fluorescent copies are spotted onto a glass slide and the resulting microarray is scanned under laser light. *Source: Butte A. The Use and Analysis of Microarray Data. Nature Reviews Drug Discovery* 1(12), 951-960 (2002 DEC).

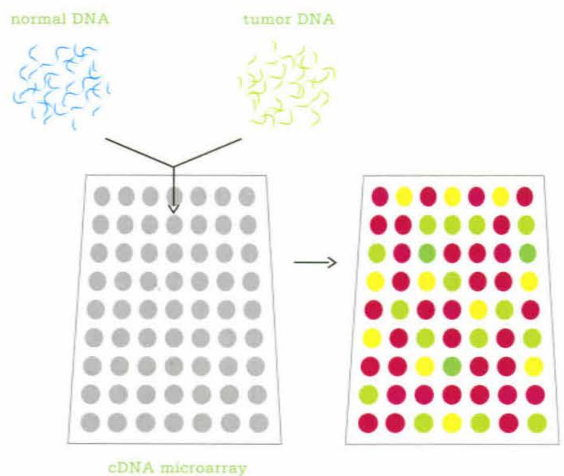


FIGURE 2 Schematic depiction of CGH. Genomic DNA isolated from tumor cells and normal cells are labeled with different fluorescent markers. They are then hybridized together to a single cDNA microarray already containing oligonucleotides. The chip is imaged by fluorescence microscopy to detect the presence and multiplicity of genes. *Source: Pollack et al. Analysis of DNA Copy-Number Changes using cDNA Microarrays. Nature Genetics* 23(1), 41-46 (1999 SEP).

"The use of DNA microarray technology in gene expression measurements **has changed the linear approach** that scientists use to study pathogenesis."

In CGH, oligonucleotides are first arrayed onto a chip. Scientists then isolate DNA from both normal and diseased (tumor) tissue and label the two DNA samples differently. These DNA samples are hybridized to the array; analyzing the signal of the labeling markers quantifies the presence of genes in the DNA of the tissues. In this way, scientists determine the differences in the copy number of genes between the normal and diseased tissues. Advances in microarray technology, as well as the completion of the Human Genome Project, have allowed the hybridization target to change from large DNA samples to small DNA probes representing regions of interest in the human genome. This improvement has significantly increased the reliability and resolution of CGH.

A tissue microarray (TMA) is a high-throughput technology that facilitates the detection of DNA copy number changes as well as individual gene expression levels at both RNA and protein levels. In this technique, a large number of tissue sections are arrayed on slides. While nucleic acid microarrays are useful for analyzing many genes in a few tissue samples, TMAs are useful for analyzing a few genes in many tissue samples. TMAs are powerful for the validation of findings obtained from microarray analysis [FIGURE 3]. Since the fabrication of the first TMA in 1998, TMAs have been used in the basic understanding of many types of cancers, Alzheimer's disease, and AIDS.

Protein microarrays contain thousands of protein molecules printed on a glass surface. The first comprehensive protein chip, containing approximately 10,000 proteins, was produced three years ago to determine the substrates for certain protein kinases (enzymes that regulate the activity of proteins). Since then, protein chips have been used in breast and prostate cancer research to determine the expression levels of cellular proteins associated with tumorigenesis. However, due to difficulties in sufficiently preserving protein structure and function when printing protein molecules onto solid surfaces, the use of protein chips has remained limited.

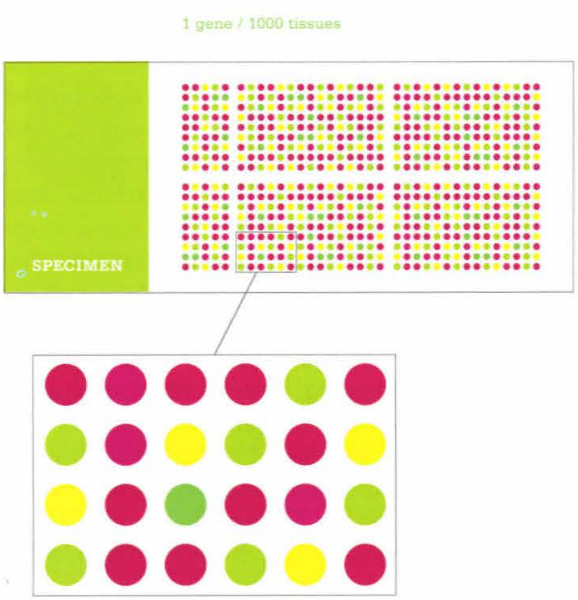


FIGURE 3 A tissue microarray (TMA). cDNA microarrays enable one to analyze thousands of genes in one specimen at a time. In contrast, TMAs are applicable to the analysis of one target at a time, but in up to 1000 tissues on each slide. This makes TMA a powerful validation tool. *Source: Kallioniemi O.P. et al. Tissue Microarray Technology for High-Throughput Molecular Profiling of Cancer. Human Molecular Genetics* 10(7), 657-662 (2001 APR).

"The roles of many genes associated with tumorigenesis have been discovered in the past decade by using DNA microarrays."

APPLICATIONS OF MICROARRAY TECHNOLOGY

Although the use of microarrays has so far been limited to global gene expression analysis, recent advances have led to applications in molecular pharmacology, toxicology, and clinical practice. One such use of microarrays is the exploration of the basic biology of diseases. The biochemical processes that cells employ to maintain their regular functions are altered during disease pathogenesis. These cellular alterations are usually the result of fine-tuning the expression levels of particular genes. Therefore, to understand the disease, it is critical to identify those genes whose expression levels are altered during disease pathogenesis.

Scientists have become accustomed to assessing expression levels of genes by measuring the quantity of their protein products. However, efficient techniques for determining protein levels on a global scale are not yet available. Advances in DNA microarray technology have allowed rapid and reliable determination of gene expression profiles by measuring mRNA levels. Since scientists can now obtain global gene expression levels far more efficiently than previously possible, DNA microarrays have become fundamental tools for global gene expression analysis.

The use of DNA microarray technology in gene expression measurements has not only sped up the experimental work of scientists, but has also changed the linear approach that scientists use to study pathogenesis—discovering and investigating one disease-related gene at a time. Now scientists can consistently and quickly monitor the expression levels of thousands of genes at different stages of disease progression and learn when and how disease-causing genes are activated or deactivated. They can also determine the interactions between genes and deduce important pathways involved in pathogenesis. In addition, methods employing DNA microarrays, such as CGH, can determine gene copy number changes on a global scale. Comparison of gene expression changes with DNA copy number alterations can provide further clues about functions of genes and their participation in pathogenesis. This use of DNA microarray technology has been predominantly employed for the understanding of cancer. The roles of many genes associated with tumorigenesis have been discovered in the past decade by using DNA microarrays.

Microarrays are also indispensable for drug discovery and development. Currently, half of the candidate drugs that perform perfectly in preclinical and early clinical tests fail when they are tested on a larger group of patients. This is often due to an insufficient understanding of the molecular actions or unforeseen toxicity of the drug. Identification of good drug targets is a particularly big challenge for pharmacologists when treating a disease because the effectiveness of treating potential targets is often limited by a poor understanding of the disease.

Microarray technology is being used in drug discovery and development to overcome these obstacles. First, DNA microarrays are versatile tools for the identification of molecular targets for drug development because they can easily compare healthy and diseased tissues and detect differences in gene activities. Secondly, scientists can use microarrays to compare desired and side effects of new drugs. Comparisons of gene expression profiles of relevant tissues before and after drug introduction with DNA microarrays as well as tests to determine drug-protein interactions with protein chips will assist in confirming that drugs are working in the way that they were designed and eliminate the possibility of side effects that might not be detectable by conventional methods.

Through similar applications, scientists can use microarrays for genotyping, disease diagnosis, and toxicogenomics—the science of identifying potential human and environmental toxins and explaining their mechanisms of action. Unexpected gene expression patterns in a test organism due to the effect of a chemical, as measured by DNA microarrays, can be informative and helpful in the identification of suspected toxins [FIGURE 4].

MICROARRAYS IN PERSPECTIVE

Although microarrays are important tools today, there are three major obstacles preventing further applications of this technology. The primary limiting factor is the high cost and the resulting limited access to the technology. Costs should decrease in the near future, but they will remain prohibitively high for many interested scientists. The second problem is the lack of satisfactory computer software to analyze the enormous data obtained from microarray analysis, causing data overload and decreasing the efficiency of each experiment. In addition, the use of many different kinds of microarray systems by researchers makes it hard to combine the data obtained by different investigators. Different systems show very different levels of reproducibility and consistency and are usually incompatible. Therefore, standards must be developed in order to make the obtained data more meaningful and accessible.

Microarray technology is one of the most rapidly evolving technologies used in biomedical science. Since the construction of the first DNA microarray in the early 1990s, technological advances as well as the completion of the Human Genome Project have allowed microarrays to move out of specialized genomics labs and into biology labs, where they are being used to answer many questions. This evolution is likely to accelerate the technology's entrance into other clinical areas. An analysis of microarray technology estimates that the microarray market will grow over fourfold from its current size, $500 million, to $2.21 billion by the year 2005. Microarrays are already having a major impact on biomedical research and are likely to play an even more important role in pharmacology and clinical practice in the foreseeable future. C

Fatih Ozsolak is a third year undergraduate in Biochemistry and Genetics at Washington University in St. Louis. *The author wishes to thank Mark A. Watson, M.D., Ph.D., Assistant Professor of Pathology and Immunology at Washington University Medical School.*

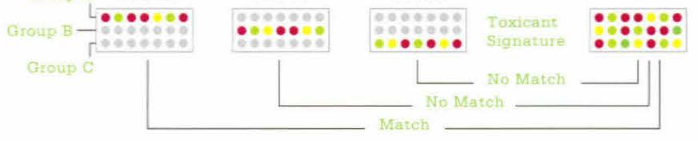


FIGURE 4 A potential array-based method for identifying a toxin's mechanism of action. Gene-expression data derived from exposure of model systems to toxins are analyzed; in the case of oxidant stressors, a consistent set of changes is detected in Group A. This pattern is then compared to the gene expression changes invoked by the suspected toxicant, and a positive match provides the basis for a hypothesized mechanism of action. *Source: Nuwaysir E.F., Bittner M., Trent J., Barrett J.C., Afshari C.A. Microarrays and Toxicology: the Advent of Toxicogenomics. Molecular Carcinogenesis* 24(3), 153-159 (1999).

FURTHER READING

1 Nuwaysir E.F., M. Bittner, J. Trent, J. C. Barrett, and C. A. Afshari. Microarrays and Toxicology: the Advent of Toxicogenomics. *Molecular Carcinogenesis* **24**(3), 153-159 (1999).

2 Kallioniemi O.P., U. Wagner, J. Kononen, and G. Sauter. Tissue Microarray Technology for High-Throughput Molecular Profiling of Cancer. *Human Molecular Genetics* **10**(7), 657-662 (2001).

3 Chakravarti A. Population Genetics—Making Sense out of Sequence. *Nature Genetics* **21**(1 SUPPL), 56-60 (1999).

4 Mills J.C., K. A. Roth, R. L. Cagan, and J. I. Gordon. DNA Microarrays and Beyond: Completing the Journey from Tissue to Cell. *Nature Cell Biology* **3**(8), E175-178 (2001).

BREAK!

EXECUTING STRATEGIES WITH ROBOT TEAMS

THE FIGHTER SQUADRON STREAKS LOW ACROSS THE COUNTRYside, fast approaching its target. Suddenly antiaircraft fire explodes skyward, destroying the lead jet. Undaunted, the remaining fighters quickly modify their defensive formation as two peel off to hunt the threat. But no casualties... it's the year 2020 and the jets are autonomous robots.

Like their human equivalents, robot teams accomplish things that an individual robot cannot. Robot technology promises utility in dangerous or uncomfortable jobs that are difficult or impossible for humans. Programming a group of robots to work together to achieve an objective is now a popular and difficult research problem. NASA's Jet Propulsion Laboratory (JPL) continues multi-robot research where two or more rovers work together to move large objects (such as solar cells for a Mars outpost), rappel down cliffs, and link together to enhance mobility on rough terrain. The military funds research into control techniques for groups of Unmanned Aerial Vehicles that currently fly alone and require careful human supervision. If a task requires many individuals with different abilities, robot teams are candidate solutions.

"Programming a group of robots to work together to achieve an objective is now a popular and difficult research problem."

BY LYLE CHAMBERLAIN

There's a problem. Programming even a single robot is a daunting task. Go to any research lab and ask for a demo. If you stick around long enough, you will inevitably see a hint that the robot you are watching is really an idiot savant. It performs amazingly well under the right circumstances, but adeptly embarrasses its creators when faced with the complexity and unpredictability of the world outside the carefully controlled lab. Sensors become uncalibrated, the floor seems more slippery than expected, or a motor overheats—these kinds of problems are difficult to anticipate. Even worse, consider an adversarial combat environment where the enemy actively tries to confuse the robot. Such problems stifle conventional attempts at control, in which the robot carefully constructs an internal model of the world and plans its actions based on this imaginary construct. The complexity of the real world makes this traditional approach ineffective in all but the most carefully controlled environments, such as a factory floor. Having several robots work together amplifies the uncertainties and computational complexity involved in this type of control. For these reasons, researchers are working hard to create new robot control techniques that don't rely on rigid assumptions about the environment.

UPPING THE ANTE: A BETTER GAME FOR BETTER ROBOTS

Recent years have seen the rise in popularity of the "Robocup" game, where university students create teams of small mobile robots and pit them against each other in a robot soccer league. Both robot teams must be autonomous—once they are turned on, it's "hands off." Cornell University has dominated the Robocup World Championships for several years now. Their leader, Professor D'Andrea, explains that the best solutions to the problem lie within conventional control techniques. Many of the problems that plague real robots don't exist in the Robocup domain. For example, a central computer that receives feedback from the game field via a camera can control all the robots. There are few to no problems of inter-robot communication, local sensor inaccuracies, team coordination, and leader designation.

These oversimplifications in Robocup led D'Andrea to design a new game: RoboFlag. The RoboFlag environment is similar to Robocup, but introduces real-world complexities. Each robot must run on its own without an all-knowing central computer, can only detect objects in its immediate vicinity, and must communicate with teammates over slow data connections in order to coordinate team actions and goals. Adding to the complexity, two humans are allowed to help the robot team of six achieve the goal of playing capture the flag against another team. How the humans should or shouldn't help the semi-autonomous team presents another real-world challenge.

"Go to any research lab and ask for a demo. If you stick around long enough, you will inevitably see a hint that **the robot you are watching is really an idiot savant.**"

In the inaugural year of the project, a team of three Caltech SURF students spent a summer at Cornell to develop the programming for a RoboFlag team. A counterpart team of Cornell students worked at Caltech to develop an opposing robot team. At the end of a two-month period, the teams met to play RoboFlag on adapted Cornell robots originally built for Robocup.

DEFINING STRATEGY WITH A 3-LAYER ARCHITECTURE

Strategies for RoboFlag gameplay determine how the individual robots, the entire team, and also the human overseers behave and interact to achieve goals. There is a difference between implementing a strategy on an actual system of robots and merely defining an abstract strategy. A strategy definition technique must employ real control techniques in defining and pursuing its goals. From the many aspects of RoboFlag we could have explored (such as human-robot interaction or high-level strategy planning), we focused on developing a multi-robot control architecture definition in order to discover exactly what information is needed to direct several independent robots to complete abstract tasks such as defense, scouting, deception, or attack. Classifying this information provides a basis upon which abstract strategies can be built.

We developed a control architecture that follows a popular three-layer hybrid control technique [FIGURE 1]. The three-layer control method combines several complementing control techniques to overcome the inherent weaknesses of the individual methods.

FIGURE 1 Three-layer hybrid control. Control handles low-level continuous problems, Sequencing generates sequences of goal-achieving actions, and Deliberative plans how to achieve high-level goals (such as "steal enemy flag") by requesting subgoals from Sequencing.

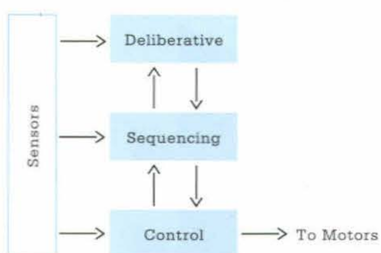
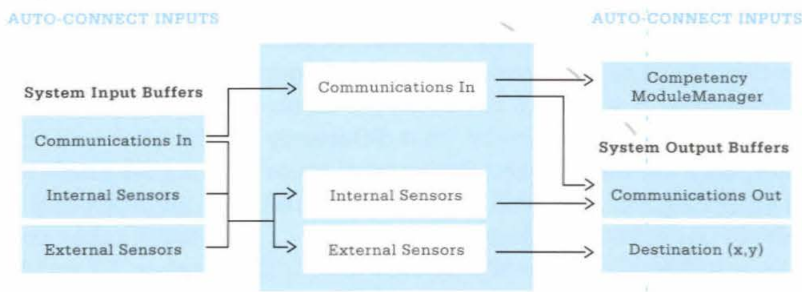


FIGURE 2 Our Three-Layer Architecture Implementation. Destination connects to the first layer, Competency Module represents the second layer, and Strategy Module represents the third.



The architecture is arranged into three ascending levels of control. The first level, "Control," provides low-level motor control and solves other continuous control problems. In our system, this level was realized as simple point-to-point navigation, handling all the dynamics of the robot mobility system.

The second level, "Sequencing," deals with specific tasks such as tracking an enemy, approaching the flag, running from an aggressor, and avoiding obstacles. It is called "Sequencing" because it implements sequences of actions designed to achieve a goal. This level provides output in the form of action requests to the Control layer.

The third level, "Deliberative," views the world in a more abstract manner. It is less concerned with specific moment-to-moment events, but plans the overall goals and actions of the robot. Its output consists of requests to the Sequencing layer for specific types of behavior on the field. For example, a request may be as abstract as "go to lower mid-field and be a defender." The Sequencing layer takes care of the specific actions, including cooperating with other robots working on the same task. Traditional planners, which use world models, work well for the Deliberative layer, where the plans are composed of goals needed to achieve a task. The actual execution of goal-achieving behavior is then performed by the Sequencing layer, which does not rely on a strong world model and is less susceptible to the ongoing changes in the environment. The Deliberative layer can thus manipulate abstract concepts such as defense, offense, power management, and cooperation without worrying about the small details and inaccuracies that can cripple a normal planning-based controller.

Our implementation of the three-layer approach is shown in FIGURE 2. "Destination" represents the Control layer, accepting an x, y coordinate and driving the robot to that location. "Competency Module" represents the Sequencing layer, providing sequences of goal-achieving behavior. "Strategy Module" represents the Deliberative layer and can activate different Competency Modules depending on the current goal. For instance, it may activate the "Faker" competency module, which rushes towards the enemy flag but always stops and runs from enemy defenders. The human operator can choose between Strategy Modules with the communications link.

PROVIDING COMPETENCY IN A VOLATILE ENVIRONMENT

Three-layer control doesn't intrinsically solve the problems of complexity and uncertainty mentioned in the introduction; however, the structure allows the introduction of another technique in the Sequencing layer, "Behavior Based Control."

Behavioral control operates under the premise that robot actions don't have to be optimal, but rather, "good enough." For example, if you were to run across an obstacle course several times, you would do it differently every time. You may not have picked the optimal route and planned each step, but you would have completed your objective: to get across the field.

In traditional artificial intelligence, actions are planned linearly [FIGURE 3A]. Sensors gather data, a model is created, and responses are planned. Behavioral control, also called "Reactive Control," is parallel instead of linear [FIGURE 3B]. Several behaviors watch the sensory input and respond immediately to what they see. Because the behaviors are basically stimulus-response mappings, they use the real world as their model. Behavioral control techniques are thus less susceptible to inconsistencies in sensor data and fast-changing environments.

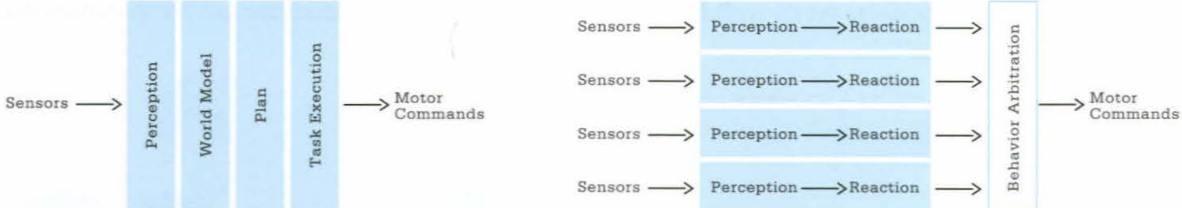
An arbitration scheme is used to decide which behavior ultimately controls the robot. Techniques exist for combining the outputs of several behaviors to generate a compromise action, but we used a simple priority-based scheme. Behaviors are given a priority level, and the highest priority behavior that wants control receives it while the others are ignored. The robot will always react to its environment in some way. Though the action may not be optimal, a carefully designed group of behaviors can provide "good enough" control that still works in an unpredictable environment.

Groups of behaviors can provide complex goal-oriented behavior. The Competency Modules consist of groups of behaviors designed to work together to achieve a specific goal. The output of the dominant behavior in the Competency Module controls the robot by commanding a navigation point to the Destination module (Control layer).

For example, suppose we want a robot to patrol an area and chase enemies. A simplified Competency Module is shown in FIGURE 4. B1 (Explore) drives the robot forward to explore its surroundings. B2 (Avoid) has higher priority than B1 and steers the robot away from obstacles and team members when necessary. B3 (Attack), takes control whenever sensors detect an enemy robot. B2 and B3 could be switched to create more cautious behavior.

We implemented a behavioral control environment which multi-tasks behaviors in parallel and arbitrates their outputs based on priority. Behaviors can be grouped into easily manipulated Competency Module objects. These objects are also smart enough to automatically connect their member behaviors to the appropriate sensors, communications, and command outputs in the environment shown in FIGURE 2. We developed a library of over forty basic behaviors that can be used to compose new Competency Modules quickly.

FIGURE 3 Traditional world modeling and planning (A) vs. behavioral control (B).



A. MODELING AND PLANNING CONTROL

B. REACTIVE BEHAVIORAL CONTROL

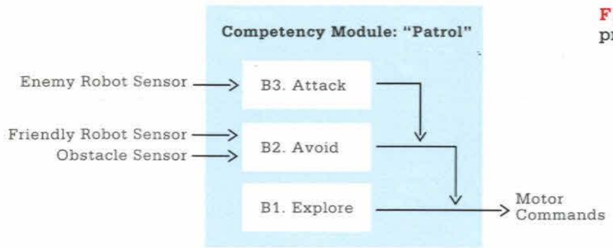


FIGURE 4 Example Competency Module using priority-arbitrated behavioral control.

ALLOWING ROBOTS AND HUMANS TO WORK TOGETHER

Inter-robot communication and human control operate through the use of virtual sensors. A virtual sensor is a communications link with another entity. The sensor becomes active when it is receiving a message. For example, if a team member sees an enemy, it may broadcast its position to the other team members. The entire team can then "see" the enemy robot as if each had detected it itself. Depending on which behavior the virtual sensor activates, the other team members may or may not respond. Human control can be inserted at many levels using this method. Virtual sensors connecting to high-priority behaviors provide direct human intervention, while virtual sensors connecting to mid-to-low-priority behaviors may be viewed as human suggestions.

The human also has direct control of the Deliberative layer [FIGURE 1]. He can request goal-achieving behavior by activating Competency Modules. For instance, he may direct four robots to defend the flag, one to tease the offense, and then manually drive a "striker" to steal the flag. Or, under attack, he may manually control the defenders and let the offense take care of itself.

We developed an interface to interact with the robots [FIGURE 5]. The interface provides various methods for selecting and controlling the robots, including selecting destinations. In addition, the human can select robot strategies as well as modify the strategy parameters during the game via the panel on the right side of the figure. FIGURE 5 demonstrates the robots executing the "Circle Defense" strategy. The "EditParam" button has been pressed to allow modification of team behavior by adjustment of the values of virtual sensors on each robot. The interface also provides a status panel that indicates each robot's state, fuel level, and currently running Strategy Module and Competency Module. The users are provided with the robot's area of visibility, or "Fog of War," that indicates sensory line-of-sight information. Furthermore, the interface provides a persistent display of enemy robots and obstacles after they have left the viewing regions of the team's robots. This persistence is represented by a version of probabilistic fading.

FIGURE 5 Human-robot interface. The blue circles surrounding the small white circle (the flag) represent robots running the Circle Defense Competency Module. The gray circles are roving robot obstacles. Light green areas represent the sensing regions of the robots. The large gray circle on the right side of the field expands and shows possible location of a previously detected mobile obstacle.

TWO APPROACHES TO WIN THE GAME

The final competition against Cornell provided insight into the utility of the system. From the beginning, our work focused primarily on developing a distributed control scheme; consequently, the human-robot interface was not as developed as the opposition's. This was a calculated risk. The intention was that our robot team would require minimal human supervision, enabling it to perform complicated plays in which every robot significantly contributed. The primary problem was associated with avoiding obstacles. The obstacle avoidance algorithm used was provided as part of the RoboFlag environment, so we had to incorporate it as part of the First Control Layer. Behaviors in the Second Layer at times contradicted the obstacle avoidance feedback. With no way to coordinate these opposing behaviors, the robots occasionally became unstable near obstacles, veering into them or rapidly oscillating in place. The obstacle avoidance routine necessitated constant human supervision of the entire robot team—negating the advantage that the high level of autonomy should have provided. Future versions will incorporate obstacle avoidance in the Second Layer.

We played three games against Cornell, whose control technique consisted of a central commanding agent that accepts sensory information from all robots, reconstructs an interpretation of the field based on this data, and then commands destination for each robot. They had a polished videogame-like robot-human interface and were very good at using it to manually control all of their robots simultaneously. We lost the first two matches as we struggled with a behavior that was acting incorrectly due to unexpected sensor noise. We were able to prototype and test alternative control networks quickly in the short inter-game periods with the library of over forty basic behaviors and the modular nature of our infrastructure. After repairing the controllers, we soundly beat Cornell in the third game. In an effort to salvage this game, they attempted a risky strategy to block access to our home base. This effort failed and resulted in most of their defense being completely incapacitated. We achieved an abnormally high score for normal game play. The first two games were close, and the third was biased by a mistake. Thus, in the end, it was hard to tell which team's approach was better suited for playing the game.

"Behavioral control operates under the premise that robot actions don't have to be optimal, but rather, good enough."

In another test, we played an informal game without obstacles. We found that suddenly our robot team's operation became more autonomous. The robots required much less "babysitting," and the team was able to attack and defend simultaneously on multiple fronts with minimal human supervision. We also played a completely autonomous game with no human intervention and no obstacles. Caltech won easily, as the Cornell team relied heavily on human intervention.

We achieved our objective of understanding what information could be used to define strategies for RoboFlag play by creating a multi-robot control architecture. Game play consisted of selecting a competency module for each robot and assigning it an area of the field in which to execute this module. In these scenarios, the human acted as the Deliberative layer. Other Caltech teams will be able to build Competency Modules quickly with the infrastructure and library of behaviors and move on to studying deliberative planners, inter-robot coordination, and human-robot interfaces.

Officially, Cornell emerged as the winner of the competition. Strong human control provided an advantage in the current system, especially where obstacle avoidance hindered autonomy. However, future versions of RoboFlag may favor a more autonomous team, and we have created tools to develop such a team. C

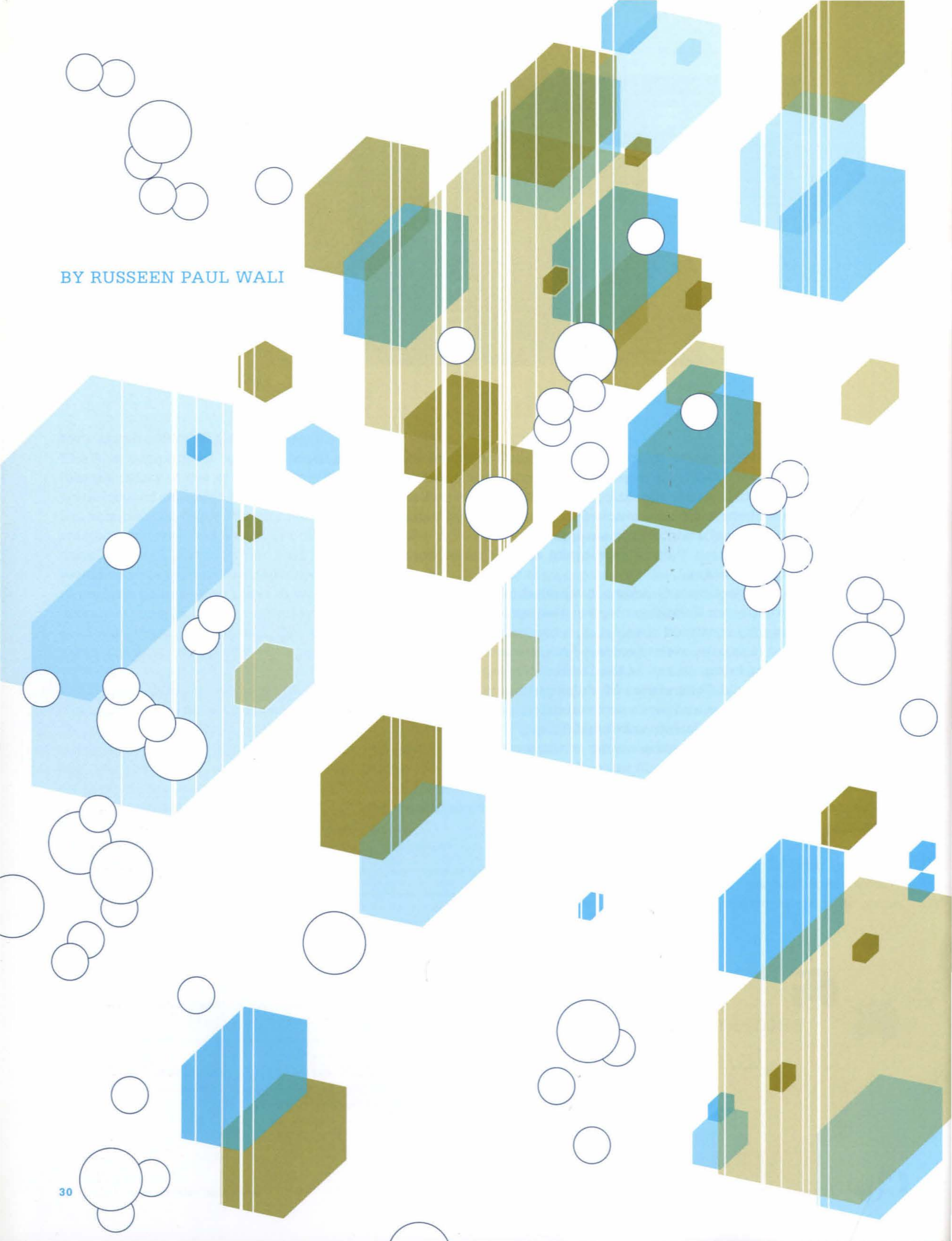
Lyle Chamberlain is a third year undergraduate in Engineering and Applied Science and minoring in Control and Dynamical Systems at the California Institute of Technology. *He would like to thank his teammates Megha Watugala, who designed and implemented the human-robot interface and the GUI; Japeck Tang, who wrote the large library of behaviors and strategies; and his mentors, Professor Julie A. Adams (Vanderbilt) and Michael Babish for their insight and patience. Many thanks also go to Professor Richard Murray (Caltech), Professor Raffaello D'Andrea (Cornell), Steve Waydo, Adam Hayes, and Jeff Sullivan. The work described in this paper was supported in part by the Defense Advance Research Projects Agency; the Air Force Research Laboratory, Information Directorate, a PECASE award, and the Caltech SURF Program.*

FURTHER READING:

1 Brooks, R.A. A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation.* **2**(1), 14-23 (1986).

2 Pirjanian, P. Multiple Objective Behavior-Based Control. *Journal of Robotics and Autonomous Systems.* **31**(1-2), 53-60 (2000).

3 Arkin, R.C. and T. Balch. Cooperative Multiagent Robotic Systems. *Artificial Intelligence and Mobile Robots.* Ed. D. Kortenkamp, R.P. Bonasso, and R. Murphy. (AAAI Press: Menlo Park, California, 1998).



BY RUSSEEN PAUL WALI

A CLEANSING CACOPHANY

TOM LEHRER ONCE JOKED "THE BREAKFAST garbage that you throw in to the Bay, they drink at lunch in San Jose." As much as we would like to forget about our sewage once it goes down the drain, the water cycle works its way back to the water flowing out of the faucet. While organic waste is usually easy to handle, more pernicious contaminants like home cleansers, medicine, and plastics require more elaborate methods. Phenol is a key component in the synthesis of many of these products, but degrades into toxic products such as benzoquinone, catechol, pyrocatechol, and hydroquinone [FIGURE 1]. These contaminants, especially the quinones, can cause nausea in small concentrations and cause chronic liver and kidney problems in larger concentrations. Our goal is to isolate and describe a key part of the degradation pathway of phenol and develop an effective means to neutralize products.

CLEANING WATER WITH SOUND AND FURY

Our work focuses on oxalate ($C_2H_2O_4$), a step in the degradation pathway of phenol, which has been successfully degraded by the use of ultrasonic waves. Like all sound waves, ultrasonic sound travels by compressing the media through which it travels. When ultrasonic waves hit water, they stretch and press together the water, creating pressure differences that in turn create tiny bubbles. These bubbles are filled with gases, water, and the dissolved molecules in the water. With each sound wave that passes through a bubble, the bubble grows until it becomes too big and consequently collapses. The collapse creates temperatures of up to 5000 K and pressures of up to 1000 atm. These conditions break down the molecules caught in the bubble and also cause water vapor to split into reactive OH^- and H^+ ions, which can then react with other molecules to form H_2O_2 and HO_2. These additional products caused by the ultrasonic waves are also thought to help in the degradation of oxalate.

"These contaminants, especially the quinones, can cause nausea in small concentrations and cause chronic liver and kidney problems in larger concentrations."

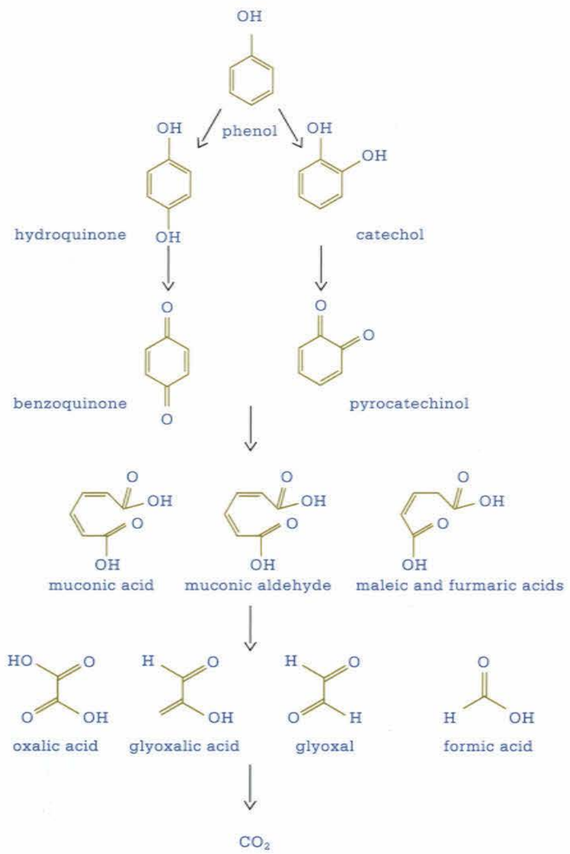


FIGURE 1 Degradation pathway of phenol

The first step in the process was to verify that all of the contaminants degraded to carbon dioxide or another compound that would bubble out and leave the solution. We prepared samples of oxalate dissolved in water and then divided them into three categories that were exposed to ultrasound, infused with ozone (O_3), or both. The total organic carbon (TOC) and the oxalate concentration were measured at thirty-minute intervals. The TOC data were consistent with readings of the oxalate concentration [FIGURE 2], so oxalate concentration was taken as a gauge of the total amount of carbon in the solution. This implies that oxalate immediately degrades into carbon dioxide and leaves the solution.

"When ultrasonic waves hit water, they stretch and press together the water, creating pressure differences that in turn create tiny bubbles."

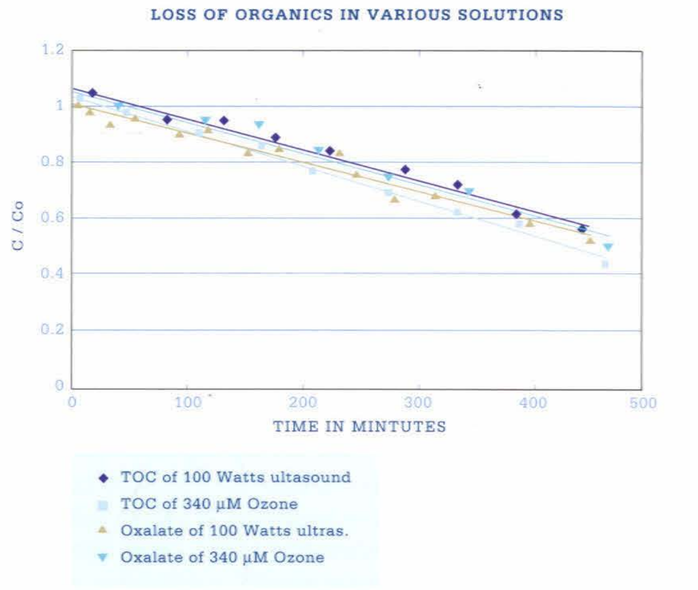


FIGURE 2 Degradation of oxalate and Total Organic Carbon (TOC) in ozonated and sonicated solutions.

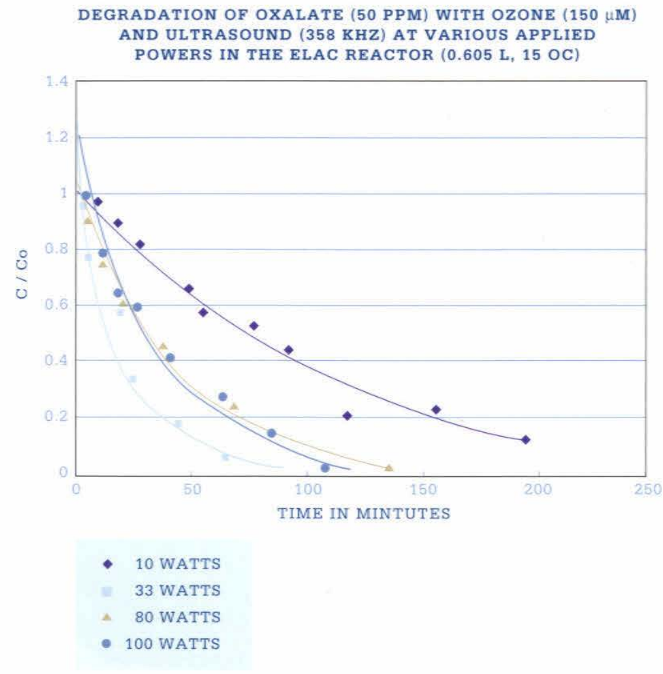


FIGURE 3 Degradation of 50 ppm oxalate with various ultrasound powers

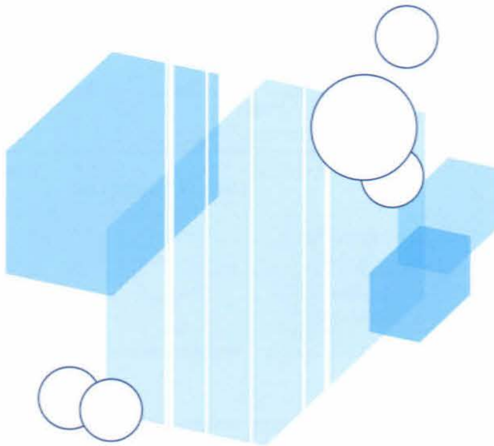
OZONE'S SHADOW

We then examined the level of ozone concentration with varying levels of ultrasound power from 0 Watts (W) to 100 W and oxalate concentrations varying from 50 to 800 ppm. At low ultrasound powers and oxalate concentrations, ozone concentration decayed exponentially, quickly reaching negligible levels. After verifying that the oxalate degradation via ultrasound was effective with this setup, we moved on to determining analytically the optimal methodology for reducing oxalate contamination.

A series of reactions were run at 50 ppm oxalate with 150 micromolar O_3 and ultrasound powers varying from 10 to 100 W. Because the rate of decay of oxalate in solution was proportional to the concentration of oxalate left, this reaction is said to follow first order kinetics. If the rate of decay were proportional to the concentration of oxalate remaining squared, it would follow second order kinetics; if the rate of decay were constant, it would follow zeroth order kinetics; and so on. We determined that 33 W was the most effective ultrasound power [FIGURE 3].

Initially, it was hypothesized that reactions would proceed more rapidly as ultrasound power increased. Prior data have shown that at 350 micromolar O_3, reaction rates were fastest for 100 W of ultrasound, followed by 50 W, 33 W, and 10 W, respectively. These preliminary data led to two working hypotheses. Higher powers of ultrasound may disturb the water matrix to such an extent that mass transfer of the ozone out of solution is so great that there is limited ozone available for reaction. The other possibility is that the ultrasound produces more radical species such as OH, which react with both the oxalate and the ozone molecules and limit the ozone available from reaction. The reactor does have lower steady-state concentrations of ozone in solution as ultrasound power increases, but it is unclear if this reduces reaction rates at higher powers.

"We found that the optimal conditions for oxalate degradation **depend mainly on oxalate concentration and ultrasound power.**"



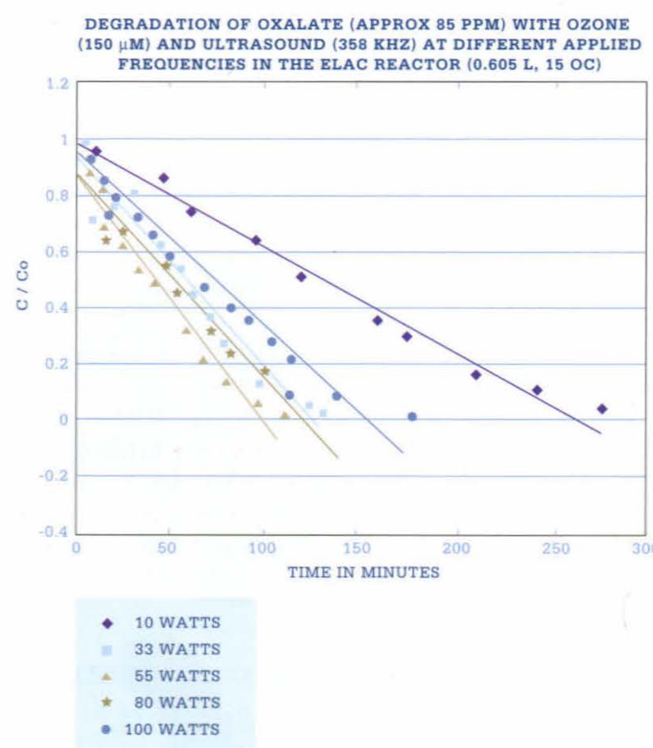


FIGURE 4 Degradation of 85 ppm oxalate with various ultrasound powers.

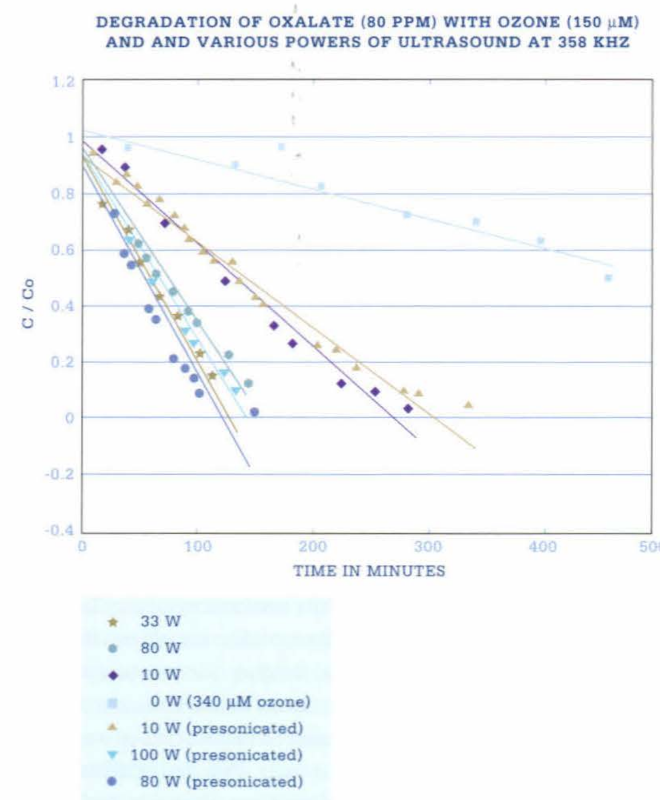


FIGURE 5 Degradation of 85 ppm oxalate with various ultrasound powers and conditions. Some samples were pre-sonicated, but this seemed to make little difference in the overall reaction rate.

A POOR PLAYER'S CHOICE OF OZONE AND ULTRASOUND

Another series of reactions was run at 85 ppm oxalate, with the same ozone concentration and ultrasound power gradations. These reactions followed zeroth order kinetics, and 50 W was the optimum ultrasound power at this concentration [FIGURE 5]. This discrepancy between the 50 ppm and 85 ppm oxalate solutions leads us to believe that a combination of oxalate concentration and ultrasound power that determines the type of kinetics that the reaction follows. This implies a point of inflection between zeroth order and first order kinetics along a continuum from 33 W and 50 ppm to 50 W and 85 ppm and further [FIGURES 3 & 4]. Another series was run in which a reaction vessel full of water and ozone was pre-sonicated for roughly one and one half hours before the oxalate solution was added. Previous work has demonstrated that significant amounts (50-200 micromolar) of hydrogen peroxide (H_2O_2) can be produced from water in the ultrasound reactor. Because this reagent also reacts with ozone we suspected it might lead to the formation of more radical species that would aid in the rate of oxalate removal. This was in fact the case, but the effect of pre-sonication was very small and certainly would not be a cost-effective way to degrade oxalate [FIGURE 5].

A TALLY SIGNIFYING SOMETHING

We found that the optimal conditions for oxalate degradation depend mainly on oxalate concentration and ultrasound power. Although it is best if ozone concentrations increase in proportion to ultrasound power, additional ozone does not increase reaction rates. For a 50 ppm oxalate solution, it is best to use about 133 µM ozone and 33 W ultrasound at 358 kHz. For a slightly higher concentration, with 85 ppm oxalate, we found that the same concentration of ozone and a slightly higher power ultrasound, 50 W, work best.

There are some difficulties in creating an industrial application: sound waves might not travel completely through large amounts of water, large amounts of ozone would be needed, and distributing ozone through a large system is not trivial. Nevertheless, there are some advantages for investigating such a technique for industrial applications. Because of the relatively short time scales involved, this method could easily clean large amounts of water in a single day. While it would not replace the traditional techniques used to purify water, it could lighten the load for methods already in place. We could then forget about the checkered past of our drinking water. C

Paul Wali is a third year undergraduate studying Chemistry at the California Institute of Technology. This work was done with Michael R Hoffmann, Professor of Environmental Engineering at Caltech, and Timothy Lesko, a graduatestudent in Environmental Engineering. *It was funded by a 2002 Caltech Summer Undergraduate Research Fellowship. The author wishes to thank Timothy Lesko, Michael Hoffman, and the Caltech SURF office.*

FURTHER READING:

1 Leighton, T.G. *The Acoustic Bubble* (Academic Press: London, 1994).

2 Mason, T.J. and J.P. Lorimer. *Sonochemistry: Theory, Application and Uses of Ultrasound in Chemistry* (John Wiley & Sons: New York, 1988).

3 L. K. Weavers, N, Malmstadt, and M. R. Hoffmann, Kinetics and Mechanism of Pentachlorophenol Degradation by Sonication, Ozonation, and Sonolytic Ozonation. *Environmental Science Technology* 34, 1280-1285 (2000).

SAVING FOR A RAINY DAY

BY ERIC CHUA

AN UNEXPECTED LAY-OFF. A SICK PARENT. A PERSONAL INJURY. It seems like a good idea to save part of your income so that, in the event of (cross your fingers!) such a crisis, you will not find yourself staring in the face of bankruptcy. So what does economic theory have to say about such precautionary savings?

Theory suggests that the best strategy is to hold an ample stock of savings to ensure that one's standard of living does not fall to unprecedented, uncomfortably low levels in the case of a financial emergency. But how much should people actually save?

THE SAVINGS MODEL

When economists first began to investigate this problem, they encountered great difficulties in mathematically solving the savings model with realistic assumptions of consumer behavior and plausible notions of income uncertainty. Solving the complex model required high levels of mathematical expertise and tedious numerical methods. Once the problem was finally solved, economists were baffled that the average household in society made saving decisions that were remarkably similar to the optimal solution! In other words, the average household was exhibiting near-optimal behavior long before economists were able to find the correct solution to the mathematically challenging problem. Does this imply that individuals actually solve the complex savings problem, or do they employ some other mechanism in which they learn the optimal savings behavior?

People can learn about optimal saving behavior via two mechanisms. First, one can individually learn about the optimal solution through repeated experience. Economist Milton Friedman suggests that repeated experience in attempting difficult problems can build good intuition about the optimal solution. For example, an experienced billiard player may not know Newtonian mechanics, but has an excellent grasp of where the balls will go when he hits them. Likewise, repeated experience with the complex savings problem can allow individuals in society to learn the optimal saving behavior. The second learning mechanism is called social learning. In this scenario, individuals observe the spending and saving decisions of others before making a decision about what kind of saving strategy is best for them. This mechanism mirrors the real world, where one can observe the strategies and consequences of others' saving decisions and take advice from parents, siblings, and friends.

One research paper by Todd Allen and Christopher Carroll suggests that the individual learning mechanism is ineffective. They use computer simulations to show that individuals can find a good spending strategy through trial and error methods only if they spend a million years of simulated time searching for such a strategy! Another paper points out a weakness of social learning. Such a mechanism can lead to "information cascades" or "herd behavior" where participants ignore their own intuition and simply

OPTIMAL SPENDING PATH

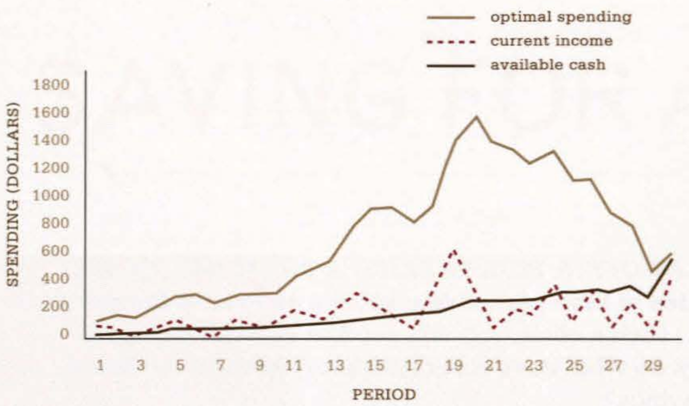


FIGURE 1 Simulated Optimal Consumption Path. The best strategy, represented by the "Optimal Spending" line, is to save in the beginning and spend later. "Current Income," or the income draw, fluctuates greatly in each period. "Available Cash" represents the sum of previously made savings and current income—in other words, the maximum amount that one can spend each period.

imitate the saving strategies of others, thereby making decisions that are sub-optimal for their own situation. However, we argue that systematic individual and social learning may be used together to make good saving strategies. To prove this, we employ experimental techniques that examine the effectiveness of both learning mechanisms.

CALLING ALL SAVERS

In our experiments, participants made saving and spending decisions in seven sequences (or "lifetimes," consisting of thirty periods each) based on the following given information: the income draw (the level of income obtained in that period), the corresponding amount of cash available, and the habit stock. By making income draws in future periods uncertain to the participant, we tested whether the participants are able to make sufficient precautionary savings. The habit stock resembles a standard of living index—people feel "pain" (lose points) when they cannot afford to maintain previous higher levels of spending. Adding the habit stock factor makes the already complex savings model even more difficult to solve.

Allowing each participant to play seven sequences of the saving experiment enabled us to investigate the effectiveness of individual learning. It is important to note that the income draws in each sequence were different. If individual learning were effective, participants should be able to formulate a good saving strategy after a few trials and apply this strategy to different income draws in future sequences.

To incorporate social learning, participants were split into two groups. The first group (individual learning) played the experiment independently. The second group (social learning), however, had additional information on how the participants in the first group (without social learning) played their experiments. This information included the spending/saving decisions of the subject who obtained the highest payoff, the one who obtained the lowest payoff, and one random player. It also showed the income draws experienced in those sequences. It is important to note that participants in the group with social learning experienced different income shocks in their seven sequences from the group without social learning. This was done to ensure that participants could not merely imitate the spending decisions of the highest point sequence. Participants with social learning should be able to form good consumption rules to use in their own income draws. Thus, by providing participants with the best and worst spending strategies from the previous group, this experiment created an accelerated social learning mechanism.

"Risk-averse participants in the social learning group relied heavily on the 'tried and tested' strategies of the previous group, even though they may have been sub-optimal for their own income realizations."

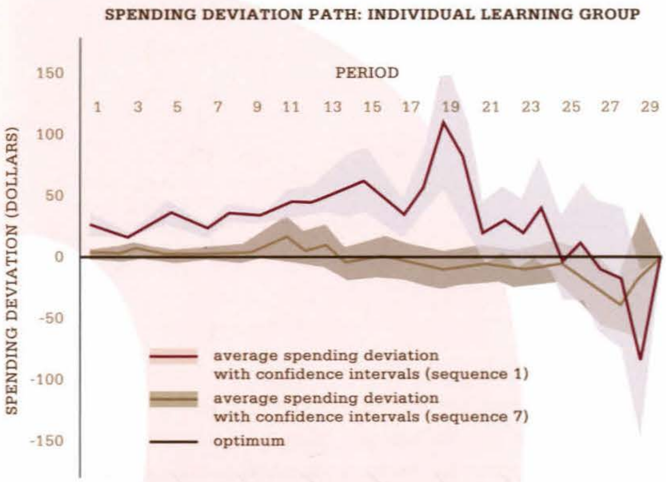


FIGURE 2 Individual Learning Group, sequences 1 and 7. Participants make better decisions over time in the individual learning group. The tan line (indicating spending behavior in the last sequence of the game) is much closer to the optimal behavior (horizontal dark brown line) than the red line (which shows spending behavior in the first sequence of the game).

"Social learning causes a staggering 73% improvement in the saving decisions of our experiment participants."

As shown in FIGURE 1, optimal behavior in our savings model requires participants to save sufficiently to protect against future bad income draws and also to ensure that the habit stock remains at affordable levels. Over-spending in any period is undesirable because spending (and consequently one's well-being) will drastically decline if future income is reduced. During early periods, the optimal strategy is to save a large portion of available cash since there are more periods of future uncertainty. As the end of the sequence approaches, participants can consume a greater proportion of their available cash, until the last period where the optimal strategy is to consume everything.

STRAYING FROM THE PATH

35 undergraduates from the National University of Singapore and 37 undergraduates from Caltech participated in this experiment. We compared participants' actual spending decisions to the mathematically optimal spending behavior. To test the effectiveness of individual and social learning, we calculated the deviation of each participant's actual spending decision from the optimum. The "Average Spending Deviation" paths in FIGURE 2 are averaged across the 35 participants in the individual learning group for sequences ("lifetimes") 1 and 7.

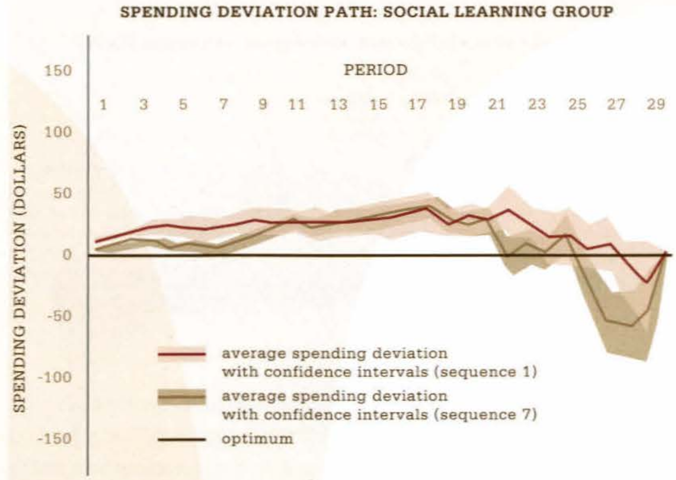
FIGURE 2 shows that participants who have little or no experience in the complex savings experiment (and no social learning) perform extremely poorly with respect to the optimum. The average spending deviation path of sequence 1 [red line, FIGURE 2] shows large deviations for much of the sequence, indicating that actual spending is significantly different from the optimum.

Individual learning, however, effectively brings spending decisions towards the optimum—in other words, individuals learn to make better spending decisions over time. The average spending deviation path of sequence 7 [tan line, FIGURE 2] is much closer to zero for much of the sequence, implying that participants learned from their mistakes in previous sequences and formed good consumption rules quickly enough to be used in this sequence. Hence, the spending decisions made in the last sequence were nearly as good as the optimum.

To examine the effect of social learning, we analyzed the average spending deviation paths for sequences 1 and 7 in the second group [FIGURE 3]. Participants in the social learning group made much better spending decisions than individual learning participants in the first sequence of the game. In other words, social learning played an extremely effective role in improving spending behavior.

However, even though individuals in the social learning group started off making better spending decisions than the individual learners, the social learners did not improve much over time. In other words, they did not seem to learn from their mistakes in order to make better financial decisions in the later sequences.

FIGURE 3 Social Learning Group, sequences 1 and 7. Participants with social learning information make better spending decisions than the individual learners. The spending deviation paths in the first and last sequences (indicated by the red and tan lines, respectively) are much closer to zero than the spending deviation paths in Figure 2. Note, however, that there is no significant improvement over time in this group.



The ineffectiveness of individual learning (that is, the lack of improvement) in the social learning group may be caused by the informational cascade phenomenon mentioned earlier. Risk-averse participants in the social learning group relied heavily on the "tried and tested" strategies of the previous group, even though they may have been sub-optimal for their own income realizations. In other words, participants put less weight on individual learning, thereby keeping their spending decisions far from the optimum.

MAKING GOOD SAVINGS DECISIONS IN OUR LIFETIMES

As expected, results show that participants without social learning and with no prior experience in playing the complex savings experiment performed extremely poorly with respect to the computationally optimal solution. If university undergraduates are unable to solve the complex problem mathematically, it is even more unlikely that the average individual in society can do so.

If this is the case, how do consumers exhibit saving behavior that is remarkably similar to the optimal solution? Our results shed some light on this problem. Contrary to earlier findings, we showed that individual learning (or repeated experience) helps participants make good saving decisions in a relatively short period of time—not the million years suggested by one paper. Regression analysis also revealed that most of the improvement in savings behavior occurs during the first two sequences, implying that participants can quickly correct their errors to make near optimal saving decisions.

We also found that social learning can be very effective in bringing saving decisions towards the optimal. Social learning causes a staggering 73% improvement in the saving decisions of our experiment participants. However, in the group with social learning, individual learning is not effective in bringing saving decisions towards the optimum. One possible explanation is that social learning causes an informational cascade that leads to persistent sub-optimal saving behavior.

Although experimental results show that individual learning can cause rapid convergence towards the optimum, we must acknowledge that, in reality, consumers only have one trial, or lifetime. However, it is possible that consumers can learn from mistakes in the past and start consuming optimally during a later point in their lives.

Social learning is a more realistic explanation for why empirical data fits the mathematical solution so well. In the real world, consumers can learn from the experiences of family, friends, and financial advisors in order to make better spending decisions. Our findings suggest that if governments are concerned that their citizens are making sub-optimal spending decisions, particularly in our difficult financial times, they should start making such social learning resources more readily available to citizens. C

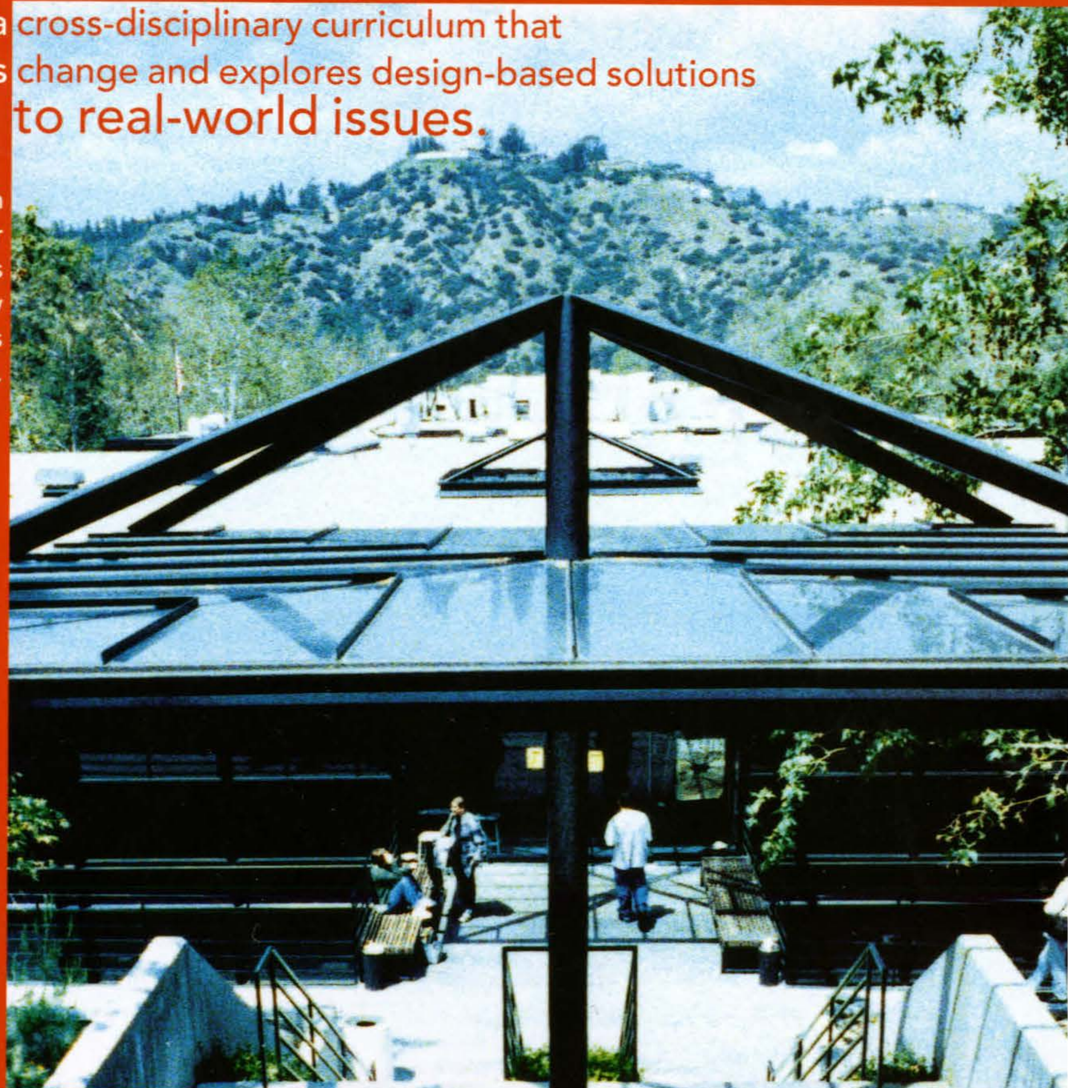
Eric Chua recently completed his undergraduate studies at the University of Cambridge, United Kingdom, where he majored in economics. *His research was funded by the SURF and the Caltech-Cambridge exchange programs. He wishes to thank his SURF mentor Colin Camerer, Professor of Economics at Caltech, as well as Chong Juin Kuan and Julie Malmquist for assistance in laboratory arrangements.*

FURTHER READING:

1 Carroll, C.D. A Theory of the Consumption Function, With and Without Liquidity Constraints. *Journal of Economic Perspectives* **15**, 23-45 (2001).

2 Allen, T.W. and C.D. Carroll. Individual Learning about Consumption. *Macroeconomic Dynamics* **5**, 255-71 (2001).

3 Fehr, E. and P.K. Zych. Do Addicts Behave Rationally? *Scandinavian Journal of Economics* **100**, 643-62 (1998).



CURJ IS A COLLABORATIVE EFFORT OF STUDENTS AT ART CENTER COLLEGE OF DESIGN
AND THE CALIFORNIA INSTITUTE OF TECHNOLOGY

MORNING GLORY WAVES

BY VICTOR TSAI

"The Morning Glory phenomenon is not a **katabatic flow**, but rather a series of **solitary waves**."

FOR YEARS, PEOPLE LIVING NEAR THE GULF OF Carpentaria in Australia have observed spectacular roll clouds. Because they frequently occur in the morning, natives dubbed the strange atmospheric phenomenon the "Morning Glory" [FIGURE 1]. Despite having been reported to the rest of the world in the 1930s, very few people took any notice of the phenomenon. Reginald Henry Clarke was the first to analyze the effect in 1972, and through work by Clarke and other researchers in the 1970s and early 1980s, the phenomenon is regarded as relatively well understood.

In 1988, Dr. Hiroo Kanamori, professor of geophysics at the California Institute of Technology, observed an unusually long period wave using a broadband seismograph station in Pasadena. However, he could not determine the speed of the wave because the observation was made only at one seismic station and because there was a lack of high-quality barometric data about the wave.

In October of 2001, Dr. Sharon Kedar of NASA's Jet Propulsion Laboratory noticed a similar long-period, pulse-like disturbance on a seismograph in the Los Angeles Basin. Dr. Kanamori, Dr. Juliette Artru, and I investigated the phenomenon further and discovered that it traveled across the Los Angeles Basin at approximately 10 meters per second (about 20 miles per hour). Since this is much slower than typical seismic waves, we hypothesized that this disturbance was probably atmospheric in origin and perhaps was related to the Australian Morning Glory atmospheric waves.

FIGURE 1 The Australian Morning Glory clouds often appear as long arches across the sky, running from horizon to horizon, though this bowing effect is caused by the observer's perspective. The clouds are only 100 to 200 meters thick and hover close to the ground, with their bases anywhere between 50 to 500 meters in altitude.

> "Three out of the five morning-glory events studied occurred within 10 hours of a distant, large earthquake."

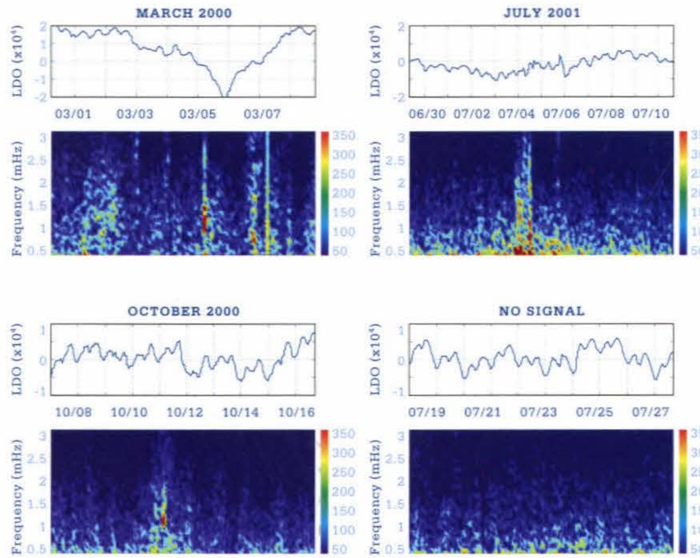


FIGURE 2 Four spectrograms of barometric data, three from data taken during Morning-Glory events and one not during a Morning-Glory event with the original barograph records shown on top. All data were recorded at the Pasadena station. Blue areas indicate low amplitudes and red areas indicate large amplitudes at a particular time and frequency. The last graph shows the background reading, with no large signals.

CLOUDCHASERS

Clarke's 1972 paper was the first to describe and model the Morning Glory's main attributes. Working mainly from eyewitness accounts and meteorological data from a single weather station on the coast of the gulf, Clarke and his colleagues described the effect as a propagating wave-like pressure jump formed by a katabatic flow—that is, a flow of cold air down the slopes around the gulf.

Clarke searched for more definitive observations of the Morning Glory in two expeditions to the gulf in the late 1970s. Collecting detailed data on the ground, taking time-lapse photographs, and chasing the clouds over 20 miles in a car to measure their speed, the team collected and analyzed a large amount of data which confirmed theories developed by Doug Christie and his colleagues in a paper published that same year.

Christie's work suggested that the Morning Glory phenomenon is not a katabatic flow, as Clarke first thought, but rather a series of solitary waves. You can picture a solitary wave by imagining a long baking pan half full of water and giving the pan one hard shake. A single bulge will run down the length of the pan without changing shape. That bulge is a solitary wave. The bottom of the pan and the air pressure form a waveguide to confine the wave to a layer just above the pan. In the same way, the Morning Glory is a wave traveling on a waveguide in the atmosphere. This waveguide arises from a temperature inversion in the atmosphere—where the temperature of the air becomes warmer with higher elevation—and guides the Australian Morning Glory waves. The air in the wave is held in this layer by buoyant force since the air immediately below it is colder and denser and the air just above it is warmer and less dense.

LOOKING FOR WAVES

The wave's characteristics suggest that it is probably atmospheric in origin, so I searched through two and a half years of barometric data taken at stations in Southern California's TriNet seismic network. However, because the TriNet network is primarily used to study earthquakes, only a few of the stations have barographs; most have only seismographs. Luckily, when I filtered the seismograph data to show only the wave's characteristic long period signal (between 500 and 2,500 seconds),

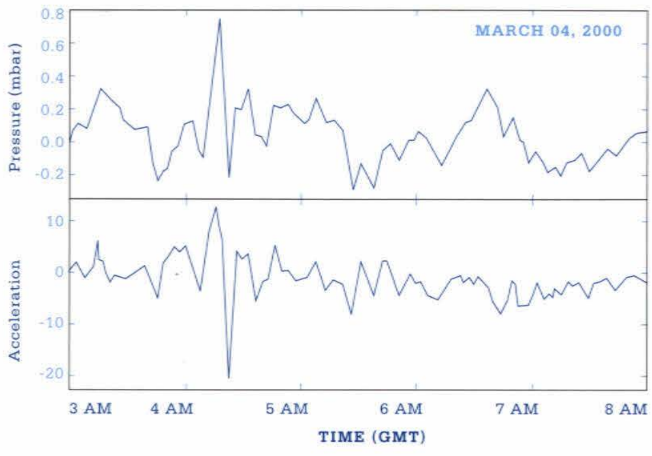
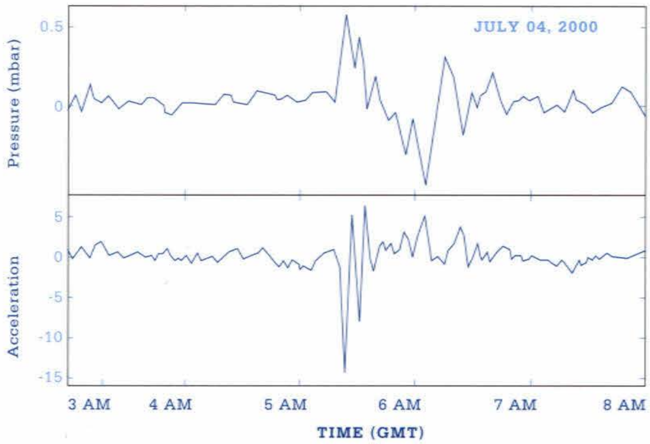
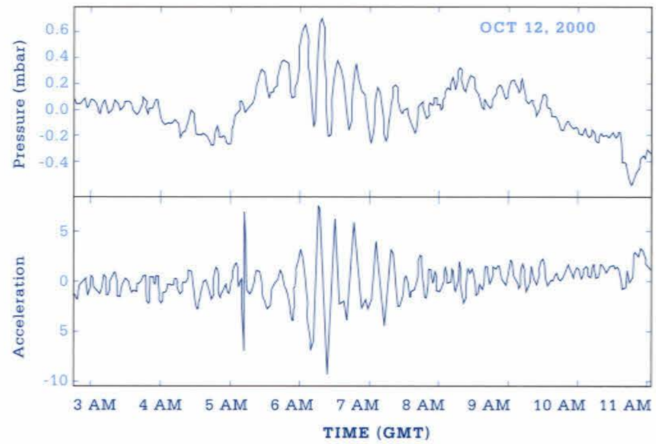


FIGURE 3 Comparisons between barograph and seismograph records of events observed in Pasadena, California. Acceleration is given in arbitrary units. The pressure changes correlate well with the activity picked up by the seismograph.





I found that the seismograph data correlated well with the barograph data [FIGURES 2 AND 3]. This correlation is due to complex interactions between the atmosphere and the seismograph's surroundings.

The complexity of these interactions makes it hard to interpret the seismic data quantitatively, but I was able to use the general shape of the seismic record to track the propagation of the pressure wave. I searched the seismic and barometric data for long period waves, then used a least-squares fit on the data to determine the movement of that wave through the Los Angeles Basin.

To verify the propagation speed, I plotted the seismograph data from each station on one graph, with time along a horizontal axis and the distance between the stations on the vertical axis [FIGURE 4]. The amplitudes of the plots are on an arbitrary scale. By looking at these record sections, I could pick out the groups of traveling waves and then confirm the speeds calculated earlier.

We decided to call this effect the Los Angeles morning-glory wave because these waves have a speed and period that is similar to the waves observed in Australia. The Los Angeles morning-glory waves do not have roll clouds because Southern California is much drier than the Gulf of Carpentaria and so there is not enough moisture in the air to make the pressure waves visible as moving clouds.

One of the most prominent atmospheric features of the Los Angeles Basin is the frequent smog. Much of the smog is due to the natural temperature inversion that commonly occurs in the basin. In most places, temperature drops with increasing height. The Los Angeles Basin, which is surrounded by mountains on one side and ocean on the other, is atypical in that the temperature

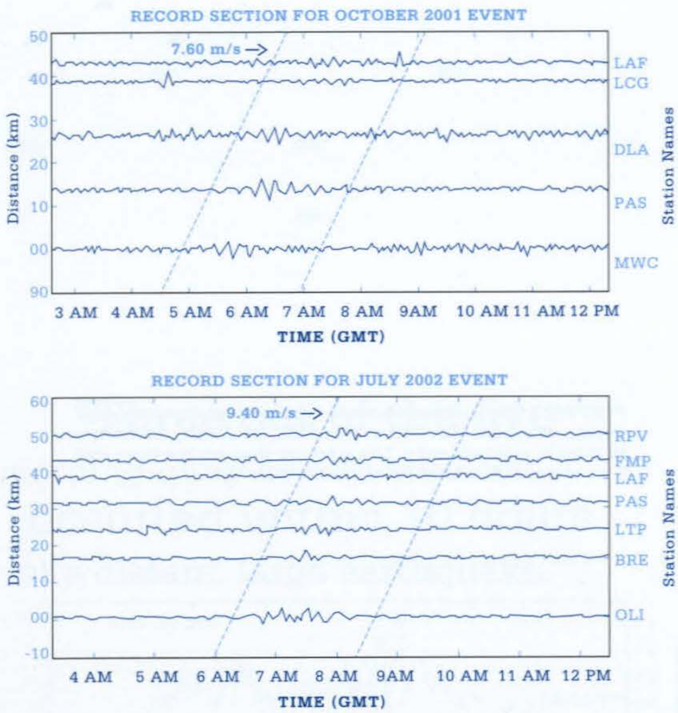


FIGURE 4 Record sections of the waves. The vertical axis gives the distance between the various stations, and the seismograms from each of those stations are plotted against time in the proper place. From these seismograms, it is easy to identify the sets of traveling waves. It is also easy to determine the speed of the waves by looking at the slope of the dotted lines.

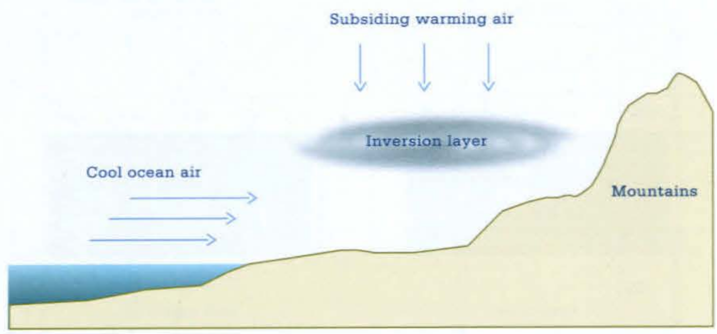


FIGURE 5 Upper air inversions in Los Angeles. The temperature inversion layer over the Los Angeles basin develops because much of California is on the eastern edge of a tropical high pressure cell in the Pacific Ocean. The presence of the high pressure means that the air in the region is subsiding from high altitudes. This subsiding air is compressed by the increasing pressure as it descends, causing it to warm up. The combination of this warm, subsiding air with the onshore flow of cool air creates the inversion layer. The mountains help to trap the air in the basin. *Source: J. Thorngren.*

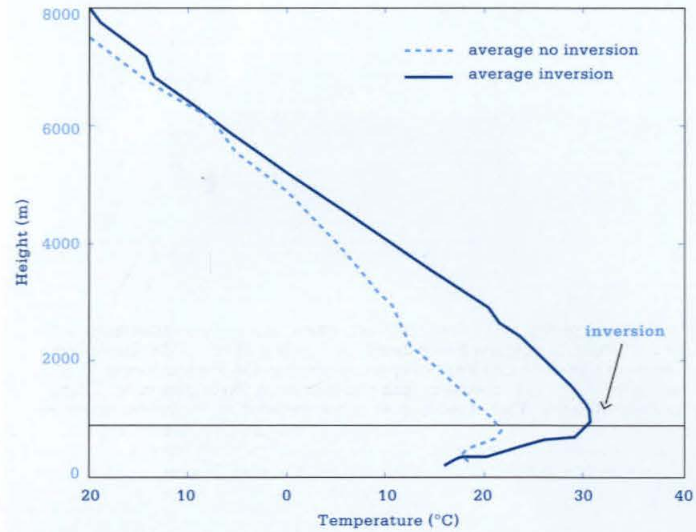
first increases with height before finally decreasing as expected [FIGURE 5]. The top of the inversion layer is typically at an altitude of 1.6 miles [FIGURE 6] and the density change is about 1% of the original density. This decrease in density forms a stable layer which acts as a waveguide.

I interpreted the Los Angeles morning-glory wave following Christie's simple two-layered model. Examining three mathematical relations in the model, I could take the measured quantities of speed, pulse width, pressure, and difference in ground pressure as known values. This left three unknowns: wave amplitude, waveguide height, and change in pressure. Solving for these values, I found that the wave amplitude is a significant fraction of the waveguide height, putting it out of the linear regime, which is consistent with our hypothesis that the wave is a nonlinear excitation of the temperature inversion layer. Even with this mathematical confirmation, we still don't fully understand what causes these waves.

STORMS, WINDS, AND EARTHQUAKES

The Los Angeles morning-glory wave of July 11, 2002 was accompanied by another atmospheric disturbance [FIGURE 7]. This disturbance showed up on the seismograph records as high frequency noise and traveled through California at a well-defined speed. The arrival of the disturbance in the Los Angeles basin coincided with the appearance of the morning-glory wave, implying a cause and effect relationship between the two phenomena. However, it isn't clear what caused this high frequency noise. Weather maps for July 10 indicated monsoonal weather, which suggests stormy weather as a possible mechanism. Strong winds, such as the Santa Ana Winds, might also be able to excite morning-glory waves in the inversion layer, but we currently do not have enough data to make any definite conclusions.

Three out of the five morning-glory events I studied occurred within 10 hours of a distant and large (magnitude greater than 6.0) earthquake. However, because such earthquakes move the ground in the area near the waves by only a tiny amount, it is difficult to believe that these seismic events can directly affect the atmosphere.



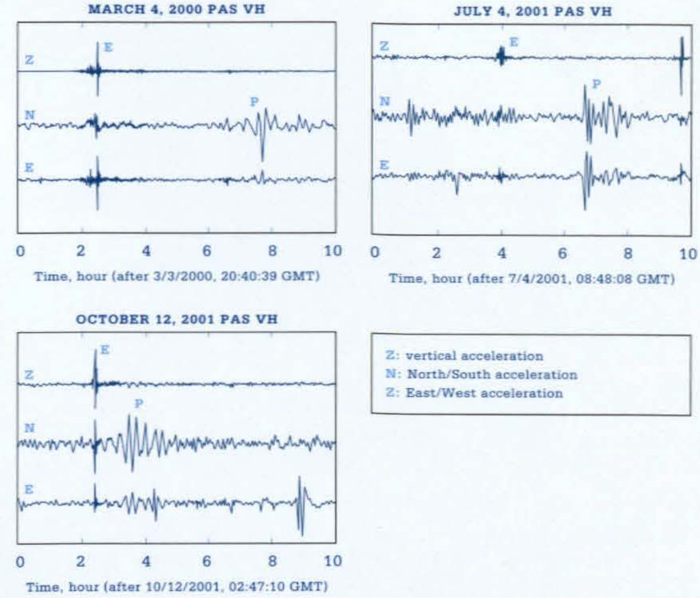


FIGURE 6 Temperature as a function of altitude. In places without a strong temperature inversion, the temperature only rises by at most 5°C between ground level and around 1000 meters. However, in a region with a strong temperature inversion, the temperature rises almost 15°C before falling again with increasing altitude.

FIGURE 7 Four of the morning-glory events I studied occurred within a few hours of a moderate-magnitude earthquake. The "E" and "P" indicate the earthquake and the Morning-Glory pressure change, respectively. The time delay between the earthquake and the pressure signals ranged from one to four hours.

On the other hand, given that about a hundred magnitude 6.0 earthquakes occur every year, and assuming that the morning-glory waves and the earthquakes are independent, the probability that three out of five Morning Glories should happen to occur within ten hours of a large earthquake is 1.2%. This makes us think that these earthquakes could somehow be partially responsible for the morning-glory waves, though we don't understand how or why.

CONSEQUENCES OF THE WAVES

Although the source of the wave remains a mystery, the discovery of the waves has consequences for both aviation safety and for the earth's internal oscillations. Researchers have demonstrated that certain "microburst" events (small, concentrated downdrafts caused by rain showers and thunderstorms) can cause airplanes to lose control and crash. Christie pointed out that the morning-glory wave is another transient, high-energy atmospheric event that might pose an aviation hazard.

More recently, Naoki Suda and his colleagues showed that the Earth constantly oscillates, but could not identify any obvious sources of excitations. These oscillations occur in the same frequency band as the morning-glory event, which suggests that the morning-glory waves may contribute to the Earth's background oscillations. Because many places in the world have inversion layers that are capable of developing Morning Glories, enough of these events around the world could potentially transfer sufficient energy to the ground to excite the Earth's modes of oscillation. C

Victor Tsai is a fourth year undergraduate in Geophysics at the California Institute of Technology. *He wishes to thank Dr. Hiroo Kanamori and Dr. Juliette Artru for their constant support and advice. This research was supported by the SURF program at Caltech, as well as by generous funding from Marcella Bonsall.*

FURTHER READING:

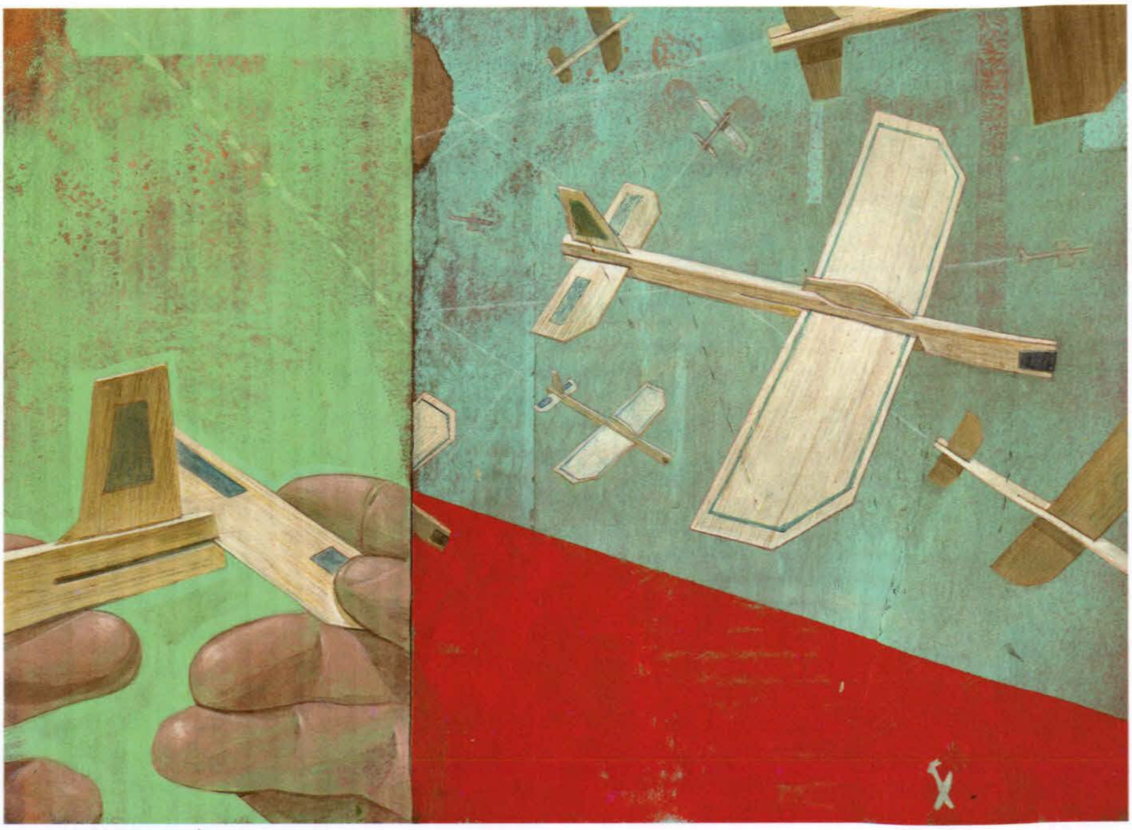
1 Clarke, R.H. The morning glory: an atmospheric hydraulic jump. *Journal of Applied Meteorology*. **11**, 304-311 (1972).

2 Christie, D. R. and R. White. The Morning Glory of the Gulf of Carpentaria. (http://www.dropbears.com/brough/Aopa.htm 29 OCT 2003).

3 Christie, D. R., J. K. Muirhead, and A. L. Hales. On Solitary Waves in the Atmosphere. *Journal of Atmospheric Science* **35**, 5 (1978).

CE N'EST PAS UN MODÈLE

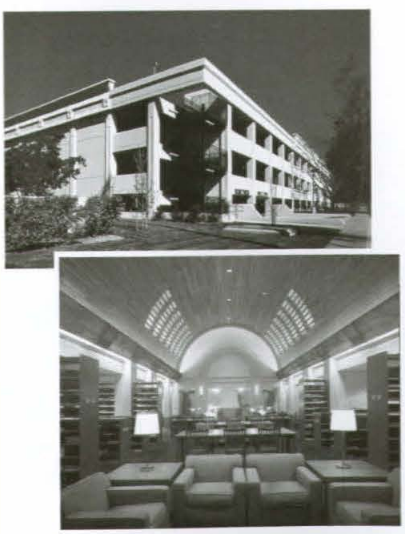
BY THOMAS LEE BAKOFSKY

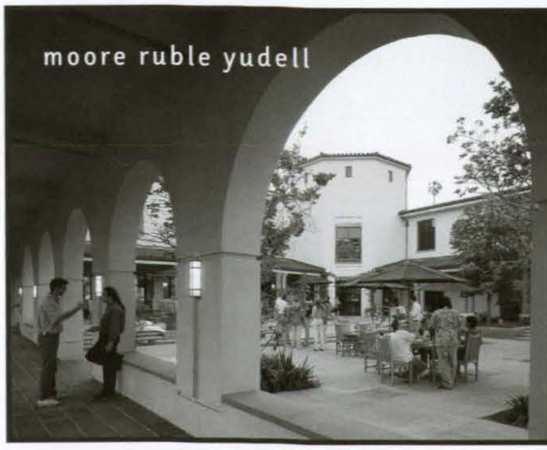


Ce N'est Pas Un Modèle
mixed media, 14" x 10"
Artist's Private Collection

The essences of art and science are an amalgamation of exploration, experimentation and innovation. I see little distinction between chip designer, oil painter, cosmologist, and sculptor. While yields of individual efforts differ, they are driven by curiosity and a love of problem solving. In the context of this illustration I've attempted to distill the intricacies of aeronautics, engineering, and perhaps even science in general into something that can be grasped at a glance; simple yet delicate wooden forms. The hands embody attention to craft, deliberateness, and organization. As a whole the work is intended to highlight the energy and inspiration inherent in both creative and scientific endeavors.

Thomas Lee Bakofsky
Art Center College of Design





THE WORLD IS GETTING SMALLER.

WHICH LEAVES MORE ROOM FOR YOUR IMAGINATION.

To us, imagination has always represented the most exciting frontier. At Northrop Grumman, we use the power of imagination to push our defense and aerospace capabilities years into the future. With projects ranging from the James Webb Space Telescope to Space Based Infra-Red and Millimeter Sensor Technology to the Joint Strike Fighter, we think there are plenty of areas left to explore. Join us and discover a place where the adventures are just beginning. To view our current opportunities, please visit our website at: www.northropgrumman.com. U.S. Citizenship is required for most positions. An Equal Opportunity Employer M/F/D/V

NORTHROP GRUMMAN DEFINING THE FUTURE™

www.northropgrumman.com