CONTENTS

Cover by Max More & Nanc Clark

EXTROPY #13

3rd quarter 1994 (vol 6, no 2)

EDITORIAL

There is beauty in space, and it is orderly. There is no weather, and there is regularity. It is predictable... Everything in space obeys the laws of physics. If you know these laws and obey them, space will treat you kindly. And don't tell me man doesn't belong out there. Man belongs wherever he wants to go.

— Werner Von Braun

The sky is no longer the limit.

— Richard M. Nixon

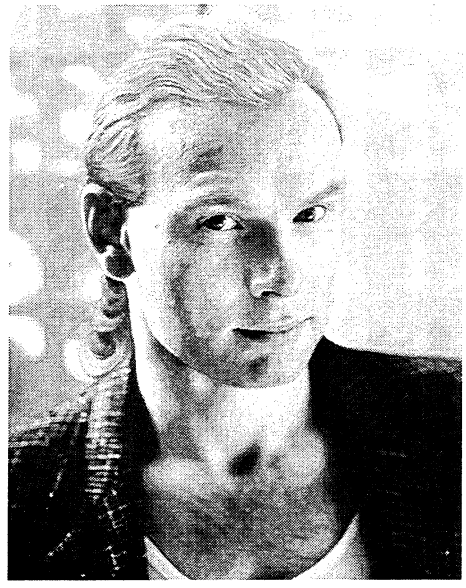


Photo by Nanc Clark

Even politicians can understand some breakthroughs. Nixon may not have been farsighted enough to have had himself frozen but he did appreciate at least some of the importance of humanity's expansion into space. We Extropians also appreciate how vital it is to promote space development. Boundless expansion, limitless lifespans, and the spontaneous ordering of experimental communities requires the room and resources to be found off-Earth.

Until now (apart from the introductory section in #1), *Extropy* has been missing a direct discussion of space technologies and cultural possibilities. With Nick Szabo's "Boundless Constellations: The Emergence of Celestial Civilization", we begin to rectify this omission. Expect to see more coverage of space issues in future issues and at next year's EXTRO conference. Nick presents an alternative to the centralized, state-controlled approach.

The idea of "uploading" human consciousness from brains to faster and more powerful hardware is something of a theme this issue. Robin Hanson's "If Uploads Come First" gives us a hardheaded, economic analysis of the effects of uploaded persons on economy and society. Part Two of Dave Krieger's interview with David Ross presents more of Ross's unorthodox thinking, and encourages us to think about the similarities and differences between the religious concept of a soul and the technologically-based idea of self as embodied software.

J. Storrs Hall examines a fascinating future nanotechnology tool known as *utility fog*. Uploading is only one of the diverse applications made possible by this technology - a technology that blurs the boundary between virtual reality and standard physical reality. Going from future to current tech, Simon! D. Levy continues his series of introductory articles on neurocomputing, this time explaining the workings of sequential neural nets.

Charles Platt (whose interview with me - "Taking the N Out of Entropy" - appears in the current issue of *Science Fiction Eye*) raises "Two Questions for Extropians" to which I offer brief responses. The first question - regarding how we are to view the idea of "pure intellect" and whether it is separable from or superior to the inherited structure of human motivation and mentation - needs open discussion because Charles is not the only one to be concerned with the way at least some Extropians seem to think of our future selves.

Contributing a change of pace and a dash of humor, Carl Feynman's "advertisement" for the Galactomatic-1000 was inspired by the wormholes discussion running on the Extropians e-mail list and in this journal (See Michael Price's "Traversable Wormholes and Interstellar Civilization" in *Extropy* #11 (5:1:14-23) and Robin Hanson's "Wormhole Warfare" in *Extropy* #12 (6.1:38-39).

In our book review section, The Transhuman Taste, Reilly Jones reviews Stuart Kauffman's important work on complexity theory, *The Origins of Order, Self-Organization, and Selection in Evolution*. Reilly's keen interest in the topic revealed itself in his comments on the anti-dogmatization panel at the recent EXTRO 1 conference. I review Julian Simon's latest groundbreaking work, *Good Mood* - and examine what happens when an exceptional economist tackles practical psychology.

Max More

EXTROPY — a measure of intelligence, information, energy, life, experience, diversity, opportunity and growth. Extropians are those who consciously seek to increase extropy. The Extropian Principles are: (1) Boundless Expansion; (2) Self-Transformation; (3) Dynamic Optimism; (4) Intelligent Technology; (5) Spontaneous Order. [See *Extropy* #11 for Extropian Principles v.2.5]

TRANSHUMANISM — Philosophies of life (such as Extropianism) that seek the continuation and acceleration of the evolution of intelligent life beyond its currently human form and human limitations by means of science and technology, guided by life-promoting principles and values, while rejecting dogma and religion. [See *Extropy* #6]

Boundless Constellations

The Emergence of Celestial Civilization

Nick Szabo

Illustrations by Nancie Clark

In tsarist Russia, Konstantin Tsiolkovsky first calculated and articulated the Extropian dream of boundless expansion through the cosmos. In free space, his "mansion-conservatories", far grander than paltry cottages like Versailles, would not collapse under their own weight. Vast fluxes of solar energy streaming through space could be tapped by gossamer-thin parabolic reflectors. Elaborations on space colony designs since Tsiolkovsky include science fiction's wide variety of imaginative but often fanciful space stations, planetary colonies, and pioneer ships, Dandridge Cole's bubbled-asteroid colonies in the 1960's, and Gerard O'Neill's sleek manufactured suburbias in the 1970's.

But attempts to implement space habitats, even tiny dependent space stations, have crumbled into disarray. We've seen the continued failure of government agencies and their obsolete plans for space stations and planetary bases. Saturn rusts; the Shuttle and Energia production lines have been shut down. Out of the ashes of this failure springs a new vision, where the solar system is recognized to have potential for far more than mere re-creations of Earth, where central planning and narrow goals are replaced by a wide diversity of means and ends. Out of this vision will emerge a celestial civilization far greater than that for which any human can plan.

Constellations

The NASA approach, the Von Braun vision of "next logical steps" for "the space program," is exemplified by the Shuttle and Space Station "Freedom" (SSF). These projects failed to meet their targets by a wide margin. The military and commercial users took most of their payloads off the Shuttle after wasting much effort to customize their satellites for that vehicle, and SSF has failed in a jumble of disorganization and miscommunication. Over $50 billion has been spent on these two projects with no reduction in launch costs and no improvement in commercial space industrialization. Russia's space stations have consumed similarly large sums with little to show for the effort. Meanwhile, military and commercial users have come up with a superior strategy for space development: the constellation.

One fundamental problem with the concept of a space station is that a large "stepping stone" or "centerpiece" of "the space program" will by its very nature be in the wrong orbit. If we choose 28.5 degrees, we lock out participation by the Russian launch sites and the largest users of space, our military in polar orbit. If we put it at 50 degrees the penalty for using it as a "way station" to Clarke (24-hour) orbit, the Moon, Mars, or asteroids is prohibitive. In turn, 28.5 degrees still puts a significant penalty over going straight to Clarke orbit, the Moon or Mars. If we put it in polar orbit, it is useful for the military, useless as a way station, and we can't get to it from the world's main spaceport, the USAF's Canaveral launch site.

In the new approach, different functions are broken down into different constellations placed in the optimal orbit for each function: thus we have the GPS/Navstar constellation in 12-hour orbits, comsats in Clarke and Molniya orbits, etc. Secondly, a task is distributed amongst several spacecraft in a constellation, providing for redundancy and full coverage where needed. SSF's different functions — satellite repair, life sciences research, space manufacturing research, etc. — require quite different environments and orbits. For example, by far the largest market for spacecraft servicing is in Clarke orbit. For a tiny fraction of the cost of a large station in the wrong orbit, we can put up a fleet of small teleoperated robots and small test satellites on which ground engineers can practice their skills. Once in place, robots can pry stuck solar arrays and antennas, attach solar battery power packs, inject fuel, and take on more sophisticated tasks as experience is gained and AI improves. Once the fleet is working, it can be spun off to commercial companies, who can work with the comsat companies to develop comsat replaceable module standards.

Space travel is expensive. $500 buys a ticket to the other side of the planet, but it costs over $10 million for a cut-rate, subsidized ride on the Russian low-Earth-orbiting space station. Automation and miniaturization are improving far faster than launcher and space habitat technology, so it will remain much cheaper to travel in space by robot proxy. In the first decades of the 21st century, instead of Mars bases with domed bubbles and spacesuited astronauts, we will see hundreds of insect-sized robots equipped for telepresence. Our entire solar system will be saturated by instruments: cheap, legion, and everywhere. Virtual reality will be less expensive and more effective than space suit reality.

To get around time lags of seconds (the moon), minutes (Mars), or hours (the outer planets), highly realistic virtual colonies will be built on Earth, starting with high-resolution, fractally enhanced 3D maps of planetary surfaces, created from data returned by tiny spacecraft. A teleprogramming protocol will be used to reduce the time-lag problem to a lag in transmitting corrections to simulation errors.

We will work and play in many virtual space colonies before we get around to building any real ones. Automated space tourism will live nicely beside automated science, prospecting, and mining.

Boundless Expansion

The solar system provides an impressive venue for expansion. Astrophysicist David Criswell (Finney & Jones, 1985) notes that we could use space to sustain a 20% per year growth in our use of bulk materials for many centuries to come. At this rate asteroids would be upgraded by 2140, and Jupiter taken apart and converted to space colonies between 2200-2400. This is only a tentative first step. If Sol's outer layers could be taken off, we could turn it into 15 white dwarf stars, each with an expected life of over 20 trillion years. Alternatively, some of the mass could go into build-

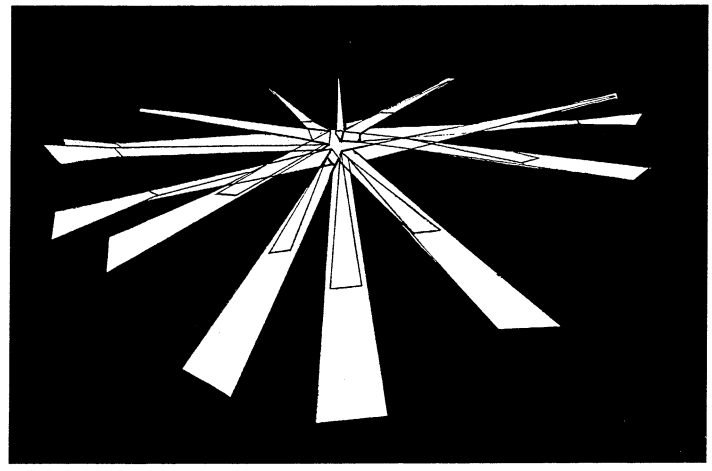
ing further space colonies, once Jupiter has been upgraded. Lifting solar plasma from the surface might be accomplished by Criswell's planet-sized, sun-straddling magnetohydrodynamic machines. These require 1.9e14 (hundred trillion) joules of energy per ton, and providing 10% of the solar flux for this task permits an upgrade rate of 6.5e18 (million trillion) tons per year.

At this rate it would take 300 million years to convert the sun to a collection of white dwarfs and more cool matter for space colonies. Building only one white dwarf would leave a distribution of life-forming elements from this upgraded matter sufficient to make 6,650 biospheres the size of Earth's every year for 300 million years. Alternatively, most of this mass could be used to build nanomachines or to make additional white dwarfs. We can take solar decentralization even further, once we master the art of confining hot dense plasmas in magnetic bottles, we can subdivide the sun to as fine a granularity as we like.

The sun currently emits 81 trillion kilowatts per person, while the "wasteful" developed countries consume only 20 kilowatts per person. Thus, we can increase some combination of population and energy consumption by up to a factor of 4 trillion within our own solar system. Gossamer mirrors in microgravity can concentrate thermal energy with orders of magnitude less mass and pollution than energy production on Earth, so there is also short-term economic incentive to tap into this resource.

Unfortunately, converting sunlight to electric power is inefficient and requires a gargantuan investment of capital. A much cheaper source of near-term electric power may come from tapping Jupiter's magnetic field, a dynamo 19,000 times stronger than the Earth's, with electrodynamic tethers. Metis is the innermost known moon of Jupiter, probably a captured asteroid dusted with sulfur from Jupiter's famous volcanic moon, Io. Traveling through the magnetic fields and inner Van Allen belts of Jupiter at 9,100 meters per second, Metis creates an electric potential of .68 volts per meter. Io, which generates 400,000 volts across its surface out where the magnetic field is weaker, sweeps up enough plasma to create a 5 million amp flux tube though the plasma between itself and Jupiter's poles. A tether with a good plasma collector may be able to tap 1 million amps at Metis, giving 10,000 kilometers of conducting cable, or the same generating capacity as 680 large nuclear plants on Earth. Generating electricity at Jupiter is almost as simple as laying down the cable, making it by far the cheapest source of electric power in the solar system.

Ultimately we would be tapping the orbital energy of Metis, which is enough for us to generate 1 billion megawatts for 630 years, before Metis falls into Jupiter. Similar amounts of energy await in Metis' neighbor moons, Amalthea, Thebes, and Andrastea. Furthermore, we can arrange to perturb asteroids and comets so that they are captured into retrograde Jupiter orbit. In this orbit, the power generated per meter of cable would be over ten times the power generated at Metis, because the orbit is traveling against instead of with Jupiter's rotating magnetic field.



Solar sail

The sun currently emits 81 trillion kilowatts per person, while the "wasteful" developed countries consume only 20 kilowatts per person. Thus, we can increase some combination of population and energy consumption by up to a factor of 4 trillion within our own solar system.

With its copious supply of volatiles and organics, the ability to capture metal asteroids, and its cheap electric power supply, Jupiter may become the industrial center of the solar system. With cheap power we can transmute elements, make antimatter, perform kilometer-scale arc welding, electroplating, vapor and plasma deposition, and much else. Laser beams based at Jupiter might power interstellar spacecraft or transmit power to various points around the solar system.

Not only can we tap cheap energy, we can also reduce the energy cost to travel around the solar system to nearly zero. Elevators use counterweights so that only frictional energy is expended in taking people to the proper floor; we can use the same principle to transport cargo between orbits.

One energy-conserving system is called the reciprocating mass driver. These electromagnetic catapults not only launch payloads, but also catch payloads and slow them down, tapping their energy to accelerate other payloads in the opposite direction. The movement of payloads around the solar system can be scheduled so that momentum and energy are conserved at each station. Any inefficiencies in the solar system mass driver net work can be made up by launching payloads with excess velocity at energy-rich Jupiter.

Are there any ultimate limits in our potential use of the solar system? Freeman Dyson notes that the rate of energy metabolism falls with the square of the temperature. This has the consequence that, in an expanding universe, life of any fixed degree of complexity can survive forever upon a finite store of energy. Cold environments are fundamentally more hospitable to complex forms of life than hot environments. Life depends less on an abundant supply of energy than on a good signal-to-noise ratio. It is easier to keep warm on Pluto than cold on Venus. Dyson has calculated that the total energy reserve contained in the sun would be sufficient to support forever a society with a complexity 10 trillion times greater than our own.

Crossing the Product Desert

Thus we see the vast potential for expansion in the solar system, in the intermediate term by first industrializing and then dismantling Jupiter, and in the long term by upgrading the sun itself. But how can we get there from here? Future technology has been described as a "product space", full of metaphorical mountains, deserts, fertile valleys, and oases. Many long-range visions, from nanotechnology to space colonization, suffer from a "product desert" between current art and future potential. Business is not willing to invest in the long-range vision, and there is a dearth of intermediate profitable businesses. How can we cross the product desert between ourselves and the vast celestial communities we want to build?

Today the biggest commercial space business worldwide is communications: between $4 billion and $20 billion per year, depending on whether you choose to count military comsats, ground stations, and the like. The industry has grown at 10% annually during a

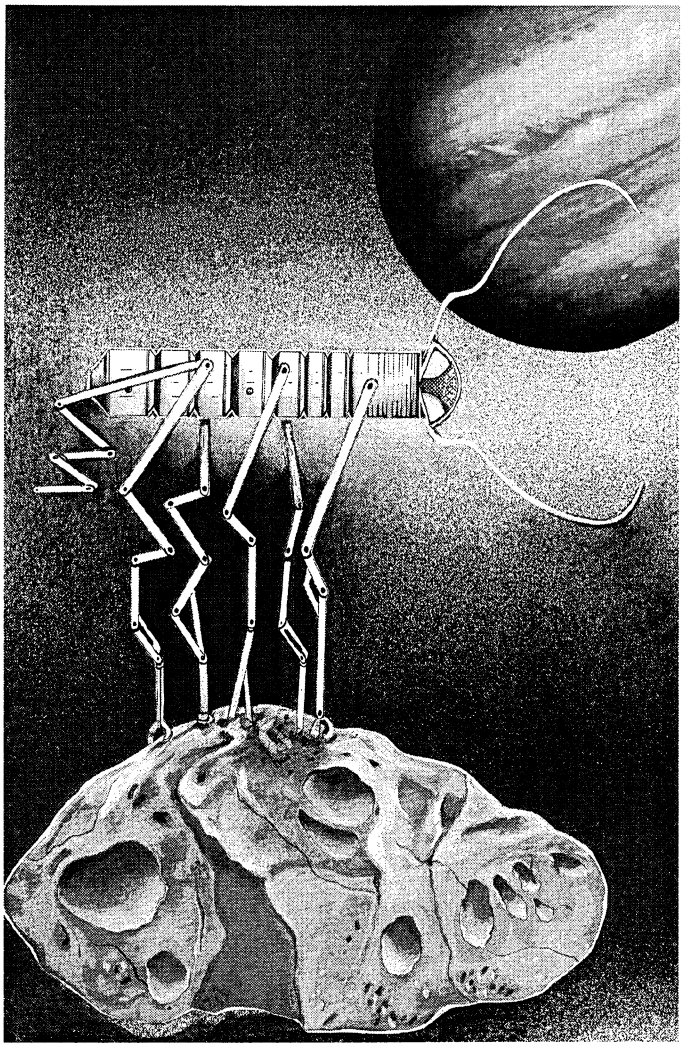
worldwide recession. Despite continued predictions of its imminent doom at the hands of fiber optics, stocks like Comsat are trading at all-time highs. Communications satellites are moving into niches quite different than those served by fiber, and their throughput has improved almost as rapidly as that of fiber. Former billboard salesman Ted Turner found the ultimate billboards, TV transmitters in Clarke orbit to beam advertisements, and incidentally great propaganda for capitalism in general, across the globe. Did a billboard salesman topple the Soviet empire?

Space tourism has been widely touted, but the Russians manage less than $30 million in revenue a year (from other governments), a small fraction of the cost of the Mir program, and only 1% of the projected annual cost of NASA's space station. Automated, virtual tourism is possible but will be a similarly small market. Militaries will continue to play a leading role in space with their automated systems, but they may cost as much in bureaucratic obstacles as they provide in capabilities. Solar power satellites require $10's of billions in high-risk investment while offering no substantial cost reductions. However, there are many intermediate uses for solar power: energy for weapons, space factories, high-power communications, emergency power or light delivered to specific locations on Earth, etc.

Microgravity promises many long-term advantages. On Earth, every activity involving large weights is dominated by the cranes, rails, engines, and other machinery needed to handle heavy objects in Earth-normal gravity. The energy/capital ratio of solar energy collectors in microgravity can be many orders of magnitude lower than any thermal power source on Earth. We can build kilometer-sized gossamer structures, mix materials without convection or separation of immiscible phases, manipulate vast volumes of plasma, etc. Single crystals can be ten or twenty times as strong for their size as the same materials in less ordered form. Several separation processes, such as electrophoresis, work much better in free space.

Unfortunately at launch costs of $8,000 per kilo to even the lowest orbit, industry is far from being able to afford to take advantage of such possibilities. The chance of large industries along these lines is small unless the cost of raw materials in orbit is reduced by several orders of magnitude. Asteroids have long been known to contain vast quantities of elemental iron and nickel for steel and could serve as a source of petrochemicals as good as oil shale. One recently suggest source is volatile ice (water, methane, ammonia, etc.) delivered via ice rocket from Jupiter-family comets, those with elliptical orbits between Earth and Jupiter.

An ice rocket consists of a long cylinder about the same size and shape as a Space Shuttle's solid rocket booster, but made out of ice and coated with a thin insulating paint. To this is attached a tiny thermal rocket, about the size of a fist, and a tiny nuclear reactor, or few square meters of mirror, which concentrates sunlight on the rocket engine. The engine slowly eats the ice, converting it into a high-velocity vapor exhaust. The rocket engine is small and simple, so that dozens of them can be built and launched on a commercial budget at launch costs not much lower than today's.



To mass-produce the ice rockets we melt cometary ice and purify it with a centrifuge. We form the ice cylinder in two steps. First we freeze a thin shell by wetting a large, cold cylindrical form. As this ice gets thicker, it freezes further layers more slowly, so we start squirting small spheres across a shaded vacuum. These spheres freeze on the outside, then accumulate on the inside of the cylinder. Soon the cylinder is filled with partly frozen water, which will continue to freeze over several years while the rocket travels towards its destination.

The ice maker is the most important part of the system. It must produce a very high ratio of ice mass to equipment mass, and it must be automated and reliable; think of a tiny automaintained sewage treatment plant. Other parts of the comet (organics, dirt, etc.) can be gathered and attached as payload. The cylinder is then attached to the small rocket engine, whose tiny thrust over the course of two or three years delivers the payload to a variety of destinations: orbits around the Earth, Jupiter, or Mars, the surface of Earth's Moon, or to asteroids. To get to high Earth orbit we must exhaust about 90% of the ice, or 80% if we take a couple extra years to use a gravity assist. We might also find ice hidden in some Earth-crossing asteroids, in Martian moons, or at the lunar poles, in which case more than 10% can be obtained.

If the output of the ice maker is high, even 10% of the original mass can be orders of magnitude cheaper than launching stuff from Earth. This allows bootstrapping: the cheap ice can be used to propel more equipment out to the comets, which can return more ice to Earth orbit, etc. Today the cost of propellant in Clarke orbit, the most important commercial orbit, is fifty thousand dollars per kilogram. The first native ice mission might reduce this to a hundred dollars, and to a few cents after two or three bootstrapping cycles.

The cost of other materials for space industries would also be drastically reduced. Besides vacuum and microgravity industries, one industry with a vast market — recreational and other officially unapproved drugs — gains a substantial, little noticed advantage from operating in space. The military has been smuggling information through space for years. It flies over enemy airspace, shooting photos, dropping them in film cartridges or transmitting them home. It beams through Clarke orbit and around the planet a wide variety of data. Some say memetic warfare via comsat played a major role in bringing down the Soviet Union.

Just as a spy satellite knows no border, so a reentry vehicle knows no border guards. Just as it has proved difficult to defend against even small numbers of incoming nuclear war-

heads, so it may be prohibitive to strike down large numbers of cheap, disposable reentry vehicles, sintered from lunar or asteroid regolith, sometimes as decoys and sometimes carrying their vital payload directly to the local distributor, anywhere on the planet. Within minutes of its detection by NORAD, a reentry vehicle has arrived at its exact location and its valuable (>$1,000/kg) payload taken away by the local dealers. In return, the dealers finance resupply launches to the space industries via an encrypted digital black market, perhaps fronting as a legal space pharmaceuticals manufacturer.

Cheap water and organics are essential to drug manufacture. Jupiter-family comets are the leading potential source. Such volatiles may also be available at the lunar poles, Mars' moons, or certain Earth-crossing asteroids. Alternatively, the crops can be grown in large bubbles on Mars itself. A Mars-fueled shuttle combined with an ice rocket makes cargo shipment to Earth very cheap, a few dollars per kilogram.

Once the volatiles and organics have been separated, they are fed to a series of chemical microreactors and converted to essential nutrients and construction materials for greenhouses. Greenhouses are made in a very simple, automated fashion, for example by pumping air into liquid polymer spheres which are then solidified and filled with nutrients and trellises for the crop. The crops grow not only drugs, but also fiber and resins to provide structural strength for further greenhouses, and genetically engineered enzymes are extracted and used in the chemical microreactors. Early nanotechnology, in the form of "techno-ribosomes" or assemblers might also help to construct the greenhouse and reactors. This self-replicating greenhouse system might expand exponentially across the volatiles of the inner solar system, converting them into drugs and reentry vehicles for delivery to Earth.

At final approach to Earth, the cargo vessel makes last-minute adjustments and screams down to its destination, until the last minute when high-g parachutes brake the cargo to a gentle landing. At the same time other elements of the "meteor shower", decoys, rain down at various nearby locations, diluting law enforcement resources in the area. For further stealth, the deliveries might be timed to correspond with actual meteor showers.

This business might be large, in the $10's to $100's of billions of dollars per year, but there will also be political pressure to stamp it out. It is not clear how this will be accomplished. By the terms of the Outer Space Treaty, no nation can claim any region of outer space. Except for treaties specifically regulating certain regions, e.g. Clarke orbit, space is an anarchy. From space, national boundaries are shown to be a mere figment of Earth-bound culture, and we can decide anew whether to take these coercive organizations with us into space.

Most of the tech needed for self-sufficient space colonies is developed in the self-replicating greenhouse system for growing crops. Interestingly, recreational drugs also dominated the settlement of our last great frontier, the Americas. Most of the early English new world colonies were started to make tobacco or rum.

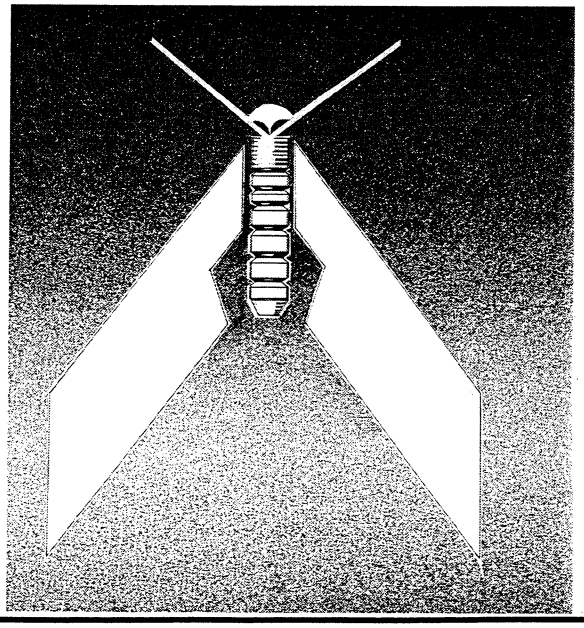
Another major New World export was precious metals. Might we find quality ores in space? Many iron meteors, and by extension metallic asteroids, have platinum ores of higher grade than any found on Earth, but unfortunately the capital costs are high and the market ($3 billion/year) is small.

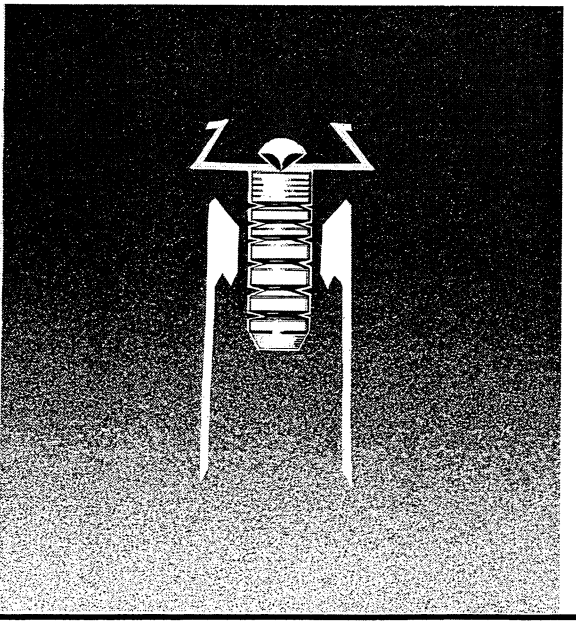
Mars in its ancient state of volcanism and running water may have formed many valuable ores. Mars Observer may have started returning its 1.5 meter resolution pix of Mars by the time you are reading this. Gold deposits of the same quality as those which once existed on prehistoric Earth could trigger a bonanza; the gold mining market is over $10 billion per year.

Several authors have proposed making CH_4/O_2 propellant from Mars atmosphere. This procedure is even simpler than ice mining: simply draw CO_2 from the atmosphere and react it with hydrogen to form methane and oxygen. Hydrogen can be shipped from Earth or extracted from Martian ice. The propellant can power rovers, mines, and single-stages surface/orbit shuttles. These boost the ore off the planet, and ice rockets ship it back to Earth. If we find high quality ore (nuggets, à la Sutter's Mill) a start-up automated gold mining operation could be cheaper than NASA's proposed grandiose astronaut mission. Unlike NASA, the project could bring in a an impressive 30% annual return, assuming only conservative launch cost reductions at Earth. After the ice rockets, Mars SSTOs, and automated greenhouses are in place, people can travel to and live on Mars for years or even lifetimes, and earn massive salaries tending the burgeoning greenhouse and mining industries located there or in orbit above.

New Visions

The traditional scenario involved mining the Moon to build kilometer-scale colonies and solar power satellites (SPS), but the capital costs are enormous, SPS would probably be more expensive than second-generation nuclear and natural gas plants on Earth, and the Moon probably lacks the volatiles and organic materials essential to life and industry. The mining of comets and asteroids, bootstrapping ice rockets and self-replicating greenhouses to supply a large number of big markets, allows a different path to space colonies, and more diverse markets for funding them. If space industry infrastructure can be established by a different business, and costs come in significantly lower than Earth-based sources, the electricity market is very large (>$500 billion/





year globally) and might provide a vast long-term space colony export market.

Mars and free-space colonies will likely compete for attracting colonists. Both kinds will initially resemble grim, Antarctic-style outposts, but as nanotechnology matures it can take advantage of the vast material and energy in space to build vast biospheres and cities unparalleled on Earth. Space will move from being an outpost for hardy workers to being a venue for pioneers wishing to vastly expand their capabilities. Jupiter may end up as the epicenter of space colonization, with cities both in orbit and on the Galilean moons. Eventually, Jupiter itself will be upgraded, turned into thousands of celestial cities strung around Sun.

Freeman Dyson's vision of affordable space colonies starts with genetic engineering to enable colonies of plants and animals to grow and spread in alien environments, and advanced automation or AI to allow machines to go out ahead of life and prepare the ground for life's settlement. His Martian potato lives deep underground, its roots penetrating layers of subterranean ice while its shoots gather carbon dioxide and sunlight on the surface under the protection of a self-generated greenhouse. A comet creeper is a warm-blooded vine which spreads like a weed over the surface of comets and keeps itself warm with a super-insulating fur as soft as sable.

The space butterfly, is fed on Earth like a caterpillar, launched into space like a chrysalis, and metamorphoses itself in space like a butterfly. It will sprout solar sails instead of wings, grow telescopic eyes to see where it is going, gossamer-fine antennae for receiving and transmitting radio signals, long springy legs for landing and walking on the smaller asteroids, chemical sensors for tasting the asteroidal minerals and the solar wind, electric-current generating organs for orienting its wings in the interplanetary magnetic field, and a high-quality brain enabling it to coordinate its activities, navigate to its destination, and report its observations back to Earth. The butterfly might also have a chemical rocket. To refuel itself, it first navigates to a comet or planetary ring and browses there, eating ice and hydrocarbons and replenishing its supply of propellant. If one ring tastes bad it can try another, moving around until it finds a supply of nutrients with the right chemistry for its needs. After eating its fill, it will use internal metabolic processes with the input of energy from sunlight to covert the food into chemical fuels. [See illustrations.]

Dyson envisions small Mayflower-style settlement expeditions to the asteroid belt. Settlers might finance their cyborg tool set with small niche businesses that thrive on isolation, like purebred breeding and genetic engineering. In the long run the gossamer mirrors built by these space dwellers form a complete Dyson sphere around the sun, lest any of its photons go to waste. Unlike O'Neill's cylindrical space colony or Larry Niven's Ringworld, Dyson's sphere is an emergent structure, not a preplanned construction. The Dyson sphere coalesces out of millions of space colonists trading surface-area real estate in a peaceful celestial market. Going beyond Dyson and O'Neill, extropian visionaries have conceived of space colonies as electronic posthuman communities, shucking biospheres for vast computer brains manufactured in space. In space, semiconductor manufacturing processes can be scaled up by factors of a million or more. After conquering the solar system with self-replicating nano-assemblers, spores for transmitting stations are shot across the galaxy and the immortal explorers beam themselves from star to star, eventually meeting on the far side of the galaxy for the greatest Extropian party of them all. At the same time, many Extropians might choose "boundless implosion", seeking ever small computers and ever faster transmission times, until the very bottom of physics, if such a bottom exists, is reached. With currently known physics, Hans Moravec estimates our solar system could contain more than 10^{30} (1 million trillion trillion) cities, each providing brain storage to a million posthumans (see "Pigs In Cyberspace", *Extropy* #10).

GLOSSARY

Clarke orbit: An equatorial orbit with a period equal to the planet's rotation, so that satellites in this orbit appear fixed in the sky from the planet's surface. Also called "geosynchronous orbit" (GEO), it is the home of most communications satellites.

Molniya orbit: High-inclination, 12-hour orbit. Satellites in Molniya orbit spend most of their time over the north or south pole, so this orbit is used for satellites linking extreme northern or southern latitudes.

Jupiter-family comets: Comets that have been captured by Jupiter into orbits resembling Earth-Jupiter transfer orbits, with periods between four and six years.

Transfer orbit: Elliptical orbit tangent to two circular orbits, which the spacecraft follows when boosting from the first circular orbit to the second.

SSTO: Single-stage-to-orbit rocket.

Reciprocating mass driver: Electromagnetic catapult combined with an electromagnetic deceleration tube that recovers energy from incoming payloads.

Upgrade: Convert a planet or sun's mass into useful form (massive computers, space habitats, etc.)

Conclusion

From Konstantin Tsiolkovsky to Freeman Dyson and beyond, visions of space have fired our imagination. Space offers a vast field of future boundless expansion. Space is not a dire necessity; we can obtain the resource and liberties we need to be posthuman here on Earth. But in the long run we need not limit ourselves to one tiny nugget of the solar system. Space is useful in bits and pieces now, and becoming more so. Most of the technology needed for future space efforts is being developed now for use on Earth. Space colonization will emerge from the work we do now to make Earth a free and prosperous place, an extropian planet.

Bibliography:

Dyson, Freeman, *Infinite in All Directions*, Harper & Row 1988.

Finney, Ben and Jones, Eric eds. *Interstellar Migration and the Human Experience*, University of California Press 1985.

Hobbs, David, *An Illustrated Guide to Space Warfare*, Prentice Hall 1986.

Lewis, John & Ruth, *Space Resources: Breaking the Bonds of Earth*, Columbia University Press 1987.

McDougall, Walter, *The Heavens and the Earth: A Political History of the Space Age*, Basic Books 1985.

O'Neill, Gerard, *The High Frontier*, Bantam Books 1976.

Wertz, James and Larson, Wiley, *Space Mission Analysis and Design*, Kluwer Academic Publishers.

IF UPLOADS COME FIRST

The Crack of a Future Dawn

Robin Hanson
Hum. & Soc. Sci. Div. 228-77, Caltech, Pasadena CA 91125.
hanson@hss.caltech.edu

What if we someday learn how to model small brain units, and so can "upload" ourselves into new computer brains? What if this happens before we learn how to make human-level artificial intelligences? The result could be a sharp transition to an upload-dominated world, with many dramatic consequences. In particular, fast and cheap replication may once again make Darwinian evolution of human values a powerful force in human history. With evolved values, most uploads would value life even when life is hard or short, uploads would reproduce quickly, and wages would fall. But total wealth should rise, so we could all do better by accepting uploads, or at worse taxing them, rather than trying to delay or segregate them.

Introduction

The future is hard to predict. We may feel confident that eventually space will be colonized, or that eventually we'll make stuff by putting each atom just where we want it. But so many other changes may happen before and during those changes that it is hard to say with much confidence how space travel or nanotechnology may affect the ordinary person. Our vision seems to fade into a fog of possibilities.

The scenario I am about to describe excites me because it seems an exception to this general rule — more like a crack of dawn than a fog, like a sharp transition with sharp implications regardless of the night that went before. Or like a sight on the horizon much clearer than the terrain in-between. And, as scenarios go, this one seems rather likely. Here it is.

If A.I. Is Hard

The human brain is one of the most complex systems we know, and so progress in understanding the brain may be slow, relative to other forms of technological and scientific progress. What if artificial intelligence (A.I.), the problem of designing intelligent systems from scratch, turns out to be similarly hard, one of the hardest design tasks we confront?[1]

If so, it may well be that technological progress and economic growth give us computers with roughly the computational power of the human brain well before we know how to directly program such computers with human-equivalent intelligence. After all, we make progress in software as well as hardware; we could now make much better use of a thirty year old computer than folks could the day it was built, and similar progress should continue after we get human-equivalent hardware. We don't know just how good human brain software is, but it might well be beyond our abilities when we have good enough hardware.[2]

Not having human-level A.I. would not mean computers and robots couldn't do better than us on many specific tasks, or that computer-aided humans wouldn't be many times more productive than unaided humans. We might even realize extreme "cyborg" visions, with biological brains and bodies wrapped in lots of artificial extras — imagine heavy use of computer agents, visual pre-processors, local information banks, etc.

But not having human-level A.I. could mean that human intelligence continues to be very productive — that on average the amount of valued stuff that can be produced decreases by a substantial fraction when the amount of human labor used to produce that stuff decreases by a substantial fraction. Cyborg add-ons, without that brain inside, couldn't do nearly as much.

Thus, as today, and as standard economic models[3] predict, most folks would still spend much, perhaps most, of their time working. And most wealth would remain in the form of people's abilities to work, even if the median worker is incredibly wealthy by today's standards. We are, after all, incredibly wealthy by the standards of the ancients, yet we still work. In contrast, having loyal human-level A.I.s could be more like owning a hundred human slaves, each as skilled as yourself — in this case there is hardly any point in working, unless for the pleasure of it.

A limited understanding of the brain and biology in general would also suggest that humans would not be highly modified — that whatever we would have added on the outside, inside we would be basically the same sort of people with the same sort of motivations and cognitive abilities. And we would be likely still mortal as well. After all, even biology has evolved the brain largely by leaving old complex systems alone; new functionality is mainly added by wrapping old systems in new add-on modules.

Uploads

Imagine that before we figure out how to write human-level software, but after we have human-level hardware, our understanding of the brain progresses to the point where we have a reasonable model of local brain processes. That is, while still ignorant about larger brain organization, we learn to identify small brain units (such as synapses, brain cells, or clusters of cells) with limited interaction modes and internal states, and have a "good enough" model of how the state of each unit changes as a function of its interactions. The finiteness and locality of ordinary physics and biochemistry, and the stability of brain states against small perturbations, should ensure that such a model exists, though it may be hard to find.[4]

Imagine further that we learn how to take apart a real brain and to build a total model of

that brain — by identifying each unit, its internal state, and the connections between units.[5] A "good enough" model for each unit should induce in the total brain model the same general high-level external behavior as in the real brain, even if it doesn't reproduce every detail. That is, if we implement this model in some computer, that computer will "act" just like the original brain, responding to given brain inputs with the same sort of outputs.

That model would be what we call an "upload" — software with human-level intelligence, yet created using little understanding of how the brain works, on anything but the lowest levels of organization. In software terminology, this is like "porting" software to a new language or platform, rather than rewriting a new version from scratch (more the A.I. approach). One can port software without understanding it, if one understands the language it was written in.

Of course some will doubt that such a brain model would "feel" the same on the inside, or even feel anything at all. But it must act just as if it feels, since it must act like the original brain, and so many people will believe that it does so feel.

Now without some sort of connection to the world, such an upload would likely go crazy or attempt suicide, as would most proto-uploads, not-quite good-enough brain-models that fail on important details like hormonal regulation of emotions. But with even very crude fingers and eyes or ears, uploads might not only find life worth living but become productive workers in trades where crude interaction can be good enough, such as writing novels, doing math, etc. And with more advanced android bodies or virtual reality, uploads might eventually become productive in most trades, and miss their original bodies much less.

Thus some people should be willing to become uploads, even if their old brains were destroyed in the process. And since, without A.I., uploads should be productive workers, there should be big money to be made in funding the creation of such uploads. The day such money starts to flow, uploads should begin to be created in significant quantity. This day would be the "dawn" I referred to above, a sharp transition with clear and dramatic consequences.

Upload Consequences

The consequences for the uploads themselves are the most immediate. They would live in synthetic bodies and brains, which could vary much more from each other than ordinary bodies and brains. Upload brain models could be run at speeds many times that of ordinary human brains, and speed variations could induce great variations in upload's subjective ages and experience. And upload bodies could also vary in size, reliability, energy drain, maintenance costs, extra body features, etc. Strong social hierarchies might develop; some might even be "gods" in comparison to others.

To a fast (meaning accelerated) upload, the world would seem more sluggish. Computers would seem slower, and so fast uploads would find less value in them; computers would be used less, though still much used. Communication delays would make the Earth feel bigger, and space colonization would seem a slower and more forbidding prospect (all else equal). Interest rates would seem smaller, making investing in the future less attractive for a given set of values.

Fast uploads who want physical bodies that can keep up with their faster brains might use proportionally smaller bodies. For example, assume it takes 10^{15} instructions per second and 10^{15} fast memory bits to run a brain model at familiar speeds, and that upload brains could be built using nanomechanical computers and memory registers, as described in [Drexler]. If so, a 7mm. tall human-shaped body could have a brain that fits in its brain cavity, keeps up with its 260 times faster body motions, and consumes 16W of power. Such uploads would glow like Tinkerbell in air, or might live underwater to keep cool. Bigger slower bodies could run much cooler by using reversible computers [Hanson].

Some people should be willing to become uploads, even if their old brains were destroyed in the process. And, since uploads should be productive workers, there should be big money to be made in funding the creation of such uploads. The day such money starts to flow, uploads should begin to be created in significant quantity. This day would be the "dawn" I referred to above, a sharp transition with clear and dramatic consequences.

Billions of such uploads could live and work in a single high-rise building, with roomy accommodations for all, if enough power and cooling were available. To avoid alienation, many uploads might find comfort by living among tiny familiar-looking trees, houses, etc., and living under an artificial sun that rises and sets 260 times a day. Other uploads may reject the familiar and aggressively explore the new possibilities. For such tiny uploads, gravity would seem much weaker, higher sound pitches would be needed, and visual resolution of ordinary light might decline (in both angular and intensity terms).

Alternatively, uploads seeking familiarity might withdraw more into virtual realities, if such simulations were not overly expensive. For relaxing and having fun, virtual realities could be anything uploads wanted them to be. But for getting real work done, "virtual" realities could not be arbitrary; they would have to reflect the underlying realities of the physical, software, knowledge, or social worlds they represent. Since, compared with software we write, the human brain seems especially good at dealing with the physical world, and since dealing with physical objects and processes should remain a big part of useful work for a long time to come, many uploads should remain familiar with the physical world for a long time to come.

An intermediate approach between tiny bodies and virtual reality would be to separate brains from bodies. Brains might be relatively fixed in location, and use high-bandwidth connections to "tele-operate" remote bodies. Of course such separation would not be economical at distances where communications costs were too high relative to brain hardware costs.

Uploads might need to find better ways to trust each other. While ordinary humans can often find unconscious signs of deception in facial expressions, upload faces may be under more direct conscious control. And uploads' minds could be tortured without leaving any direct physical evidence of the event.

If, as seems reasonable, upload brains are given extra wiring to allow the current brain state to be cheaply "read out" and "written in", then uploads could change bodies or brains relatively often, and could be transported long distances by ordinary communication lines. "Backups" could be saved, allowing near immortality for those who could afford it; if your current brain and body is unexpectedly destroyed, your latest backup can be installed in a new brain and body.

The most dramatic consequences for both uploads and everyone one else come, I think, from the fact that uploads can be copied as well as backed-up. The state of one upload brain might be read out and written into a new upload brain, while that state still remained in the original brain. At the moment of creation, there would be two identical upload minds, minds which would then diverge with their differing experiences.

Uploads who copy themselves at many different times would produce a zoo of entities of varying degrees of similarity to each other. Richer concepts of identity would be needed to deal with this zoo, and social custom and law would face many new questions, ranging from

"Which copies do I send Christmas cards to?" to "Which copies should be punished for the crimes of any one of them?".[6]

New forms of social organization might be useful for families of copies of the same original mind; some families of copies might be very loyal, while others might fight constantly. Teams of people who work well together might even be copied together, creating "team families". Political institutions like "one person, one vote" might require substantial modification, though large copy families could find obvious candidates to represent them in legislatures.

A Population Explosion?

Perhaps the most dramatic consequence of upload copying is the potential for an huge population explosion. If copying is fast, cheap, and painless, and if enough uploads desire to, can afford to, and are allowed to make such copies, the upload population could grow at a rate far exceeding the rate at which their total wealth grows, triggering a rapid reduction in per-capita (meaning per-copy) wealth.

Would an upload population explode? For a little perspective, let's review ordinary human population growth. In the short term one might take people's values[7] as given. In that case reproduction rates depend on values and per-capita wealth, and per-capita wealth depends on values and reproduction rates.

People choose to have more or fewer babies depending on their values and culture, how much such babies will cost them, the wealth they have to give, how much payback they expect to get from their children later, and on how their children's lifestyle will depend on family size. Technology and wealth also influence contraception and the number of babies who survive to adulthood.

Changes in per capita wealth, on the other hand, depend not only on reproduction rates, but also on how much folks value current consumption over future consumption, and on the rates of growth possible in physical, human, and knowledge capital. And knowledge capital growth rates seem to grow with the size of the human population [Simon].

The net result of all these factors is not clear from theory, but since we have observed rising per-capita wealth for the last few centuries, we might suppose the net tradeoff, given current values, favors rising per-capita wealth.

A few centuries is only about a dozen generations, however. And Darwinian arguments suggest that if values can be inherited, then after enough generations the values in a species should evolve to favor the maximum sustainable population for any given technology, and the maximum sustainable growth rate as technology improves [Hansson&Stuart].

This Darwinian view holds that our familiar human values, for resources, health, comfort, leisure, adventure, friendship, etc., were well suited for promoting maximal population and growth in the sort of environments our ancestors faced long ago. And this view suggests that any current conflict between values and maximal growth, such as that suggested by declining populations in Europe, is a temporary aberration due to "recent" rapid changes in the human environment.

Thus, given enough generations, human values should evolve to promote maximal growth in our new sorts of environments — one may still worry, for example, that small minorities who value exceptionally large families[8] will eventually come to dominate the population.

To a fast (meaning accelerated) upload, the world would seem more sluggish. Computers would seem slower, and so fast uploads would find less value in them. Communication delays would make the Earth feel bigger.... Interest rates would seem smaller, making investing in the future less attractive for a given set of values.

Of course a complete story of how human values evolve must include the evolution of idea and value elements as "memes", entities in their own right and not just as properties passed from human parent to child through a combination of genetic and cultural evolution. But if our receptivity to accepting non-parental values can be genetically or culturally modulated, it is hard to see how human values could consistently resist human Darwinian pressures over the long term, even with memetic evolution. Overall, these Darwinian arguments suggesting maximal growth seem roughly right.

Fortunately, however, this Darwinian process seems slow, and if economic growth rates continue their historical acceleration, they should soon exceed the maximum rates at which ordinary humans can have babies. From then on, per-capita wealth would have to increase, at least until artificial wombs were created, or until raw materials or knowledge progress started to "run out", and could no longer expand exponentially with the population as they have so far. For now though, the world seems to be changing too fast for Darwinian evolution to catch up.

How do uploads change all this? An upload considering making a copy is much like a parent considering making a child. An upload would consider the cost to create a copy, the lifestyle that copy could afford, and how much they would value having another entity like themselves. Uploads may value having copies of themselves more or less than ordinary folks now value having children somewhat like them — this is hard to predict.

But what is clearer is that upload reproduction rates can be very fast — the upload population could grow as fast as factories could generate new upload brains and bodies, if funds could be found to pay these factories. Upload copies, after all, do not need to be raised to adulthood and then trained in some profession; they are immediately ready to become productive members of society. Thus the main limitations on reproduction, and hence on Darwinian evolution of values, would become economic and political. Who would want to pay how much to make upload copies? And who would try how hard to stop them?

Upload Economics

To separate some issues, let us first imagine an upload, a contract lawyer by trade, who is neutral on the subject of whether she would like more entities like herself around, but who is considering an offer from someone else to pay for the creation of a copy. For simplicity, imagine that the original would keep all unique possessions and exclusive associations, such a painting, spouse, or job, and that the copy will have to start from scratch.

Such an upload might plausibly agree to this copy if she decided such a copy would consider their life "worth living", better to have existed than not. And since this copy could earn wages as a contract lawyer, she might consider life worth living if those wages, plus interest on some initial wealth endowment, were enough to cover some minimum standard of living.

Note, however, that if an upload expects wages to be high enough above their minimum required income, they might agree to a copy even with a negative initial endowment. That is, if a copy were to be loaned enough money to buy their new brain and body, that copy might still find life worth living even under the burden of paying back this loan.[9]

If we now add in the original upload's values for having copies around, presumably positive for having more company but negative for the added wage competition, we should find that such an upload has some minimum expected income at which she would be willing to spin off copies. And given that this upload has decided to make a copy, she may or may not prefer to transfer some of the original's wealth to that copy.

Of course some uploads, perhaps even most, might not accept this line of reasoning.

But those that do would, if not forcibly prevented, keep making copies until their minimum income threshold is reached. Thus if there are even a few such uploads[10], wages for contract lawyers should quickly fall to near the lowest wage any one such upload contract lawyer is willing to work for. At this point many previous contract lawyers would find themselves displaced, even though the total number of contract lawyers has risen. And a large fraction of all contract lawyers should be copies of that one upload!

Of course abilities vary, and the lack of an ordinary body could be a disadvantage for early uploads competing with ordinary workers[11], limiting the number of ordinary workers uploads could initially displace. And reduced diversity of thought among a large family of copies may put them at a disadvantage in trades which place a premium on creativity. But in many trades, like contract law, a large number of standardized workers might have special advantages, especially in reputation-building.

It also takes time for a labor market to absorb new workers; each job is somewhat different, and it takes time for people to learn each new job. Uploads running faster than ordinary humans might quickly master the relevant book-learning, but for most jobs most learning comes from watching and working with co-workers. At first, most co-workers will not be uploads, and most physical processes being managed would be tuned for ordinary human speeds, so being very much faster than usual may not be worth the cost of the faster hardware.

But as uploads became a larger part of the economy, upload communities which standardize on faster speeds would become more economical. If the rate at which faster uploads can grow wealth increases to match their faster speeds, then market interest rates should grow with the speed of such uploads. Slower individuals would then be much more tempted to save instead of consuming their wealth.

Falling wages should mean that, on the margin, labor is substituted for other forms of capital. So lower wage uploads should use fewer computer and other productivity aids, and hence seem less "cyborgish".

What about professions where no upload has prior training? Even if the cost to upload people were very high, or the number of volunteers very low, upload workers should still displace other workers, though at a slower rate. If the wage in some trade were above an upload's minimum, even considering the costs of learning that trade, and if loans could be arranged, copies would be created intending to master that trade.

The economics of training uploads could be much like the current economics of software. For example, labor "products" might be sold at substantially above marginal cost in order to recoup a large initial training cost. To control prices, some families might want to formally centralize their decisions about how many copies they make, so that each copy is no longer free to make more copies. In other families, informal mechanisms might be sufficient.

As with other software, uploads might reach capacity limits; after a few hundred or thousand years of subjective experience, uploads might go crazy in some now unknown way, or simply be less and less able to learn new skills and information. If this happens, then investments in training might be limited to backups made and saved when uploads are below some critical subjective age.

Fast uploads who want physical bodies that can keep up with their faster brains might use proportionally smaller bodies. A 7mm. tall human-shaped body could have a brain that fits in its brain cavity, keeps up with its 260 times faster body motions, and consumes 16W of power. Such uploads would glow like Tinkerbell in air, or might live underwater to keep cool.

Also as with software now, illicit copying of uploads might be a big problem. An upload who loses even one copy to pirates might end up with millions of illicit copies tortured into working as slaves in various hidden corners. To prevent such a fate, uploads may be somewhat paranoid about security. They may prefer the added security of physical bodies, with "skulls" rigged to self-destruct on penetration or command. And without strong cryptography, they may be wary of traveling by just sending bits.

The Evolution of Values

The analysis above suggests that, at least at first, the upload population should expand as fast as people can arrange loans, build brains and bodies, learn new jobs and professions, and as fast as the economy can absorb these new workers. Per-capita wages seem likely to fall in this period, for ordinary humans as well as uploads, though total wealth should rise.

This population explosion should continue until it reaches limits, such as those of values or of subsistence. Values limits would be reached if almost no capable, versatile upload found copies worth making at the prevailing low wages. Subsistence limits would be reached if uploads simply couldn't make ends meet on a lower income; lowering their standard of living any more would lower their productivity, and hence wages, by so much that they could not afford even that lower standard.

Would values limit this explosion? Yes, of course, if typical values were held constant; few people now who would make productive uploads would be willing to work at subsistence levels. It seems, however, that values will not be held constant. With upload copying, the potential rate and selectivity of reproduction could once again be comparable to the rate at which the world changes; Darwinian evolution (this time asexual) would have caught up with a changing world, and be once again a powerful force in human history. And since the transmission of values from "parent" to "child" is so much more reliable with upload copying, the direct evolution of "memes" should have even less room to modify our basic Darwinian story.

As wages dropped, upload population growth would be highly selective, selecting capable people willing to work for low wages, who value life even when life is hard. Soon the dominant upload values would be those of the few initial uploads with the most extreme values, willing to work for the lowest wages[12]. From this point on, value evolution would be limited by the rate at which people's values could drift with age, or could adjust to extreme circumstances.

Investors with foresight should be able to make this evolution of upload values even faster than ordinary "blind" biological evolution. Investors seeking upload candidates, or upload copies, to whom to loan money, would likely seek out the few capable people with the most extreme and pliable values. After all, these candidates would, all else equal, have the best chances of repaying their loans.

Values might evolve even faster by combining crude modification techniques, like the equivalent of neuroactive drugs or even torture, with the ability to rerun experiments from identical starting points. Of course I do not advocate such experiments, but if they were effective, someone somewhere would likely use them. Fortunately, I suspect ordinary human values are varied and flexible enough to accommodate demand without resorting to such techniques. For example, identical twins who live together are much more different from each other than those reared apart. Similarly, an upload in a million-copy family should try all the harder to be different somehow, including in their values. Thus, given all these factors, the evolution of upload values might be very fast indeed.

What would values evolve to? Would wages hit subsistence level limits? I expect

that over many generations (i.e., times copied) Darwinian selection should favor maximum long-term generation of "wealth" that can be used to buy new copies. That is, since upload reproduction can be so directly bought, we expect evolution to favor uploads whose values induce them to take actions which give their copy lineage the maximum long-term financial return on their investments, including their investments in new copies, new skills, or in "leisure".

Uploads who are overly shy about copying would lose out, holding less of the total wealth (as a group), measured by market value of assets, and constituting less of the population. Similarly, uploads who go wild in copying, just because they like the idea of having lots of copies, would become more numerous in the short term but lose out in the long term, both in total wealth and population. Thus we don't expect uploads to become as poor as possible, though we do expect them to eliminate consumption of "frills" which don't proportionally contribute to maximum long term productivity.

We should also expect an evolution of values regarding death and risk.[13] Imagine situations in which making a copy might pay off big, but most likely the copy would fail, run out of money and have to be "evicted" from its brain and body. Many people might decline such opportunities, because they so dislike the prospect of such "death". Others might consider this not much bigger a deal than forgetting what happened at a party because they were too drunk; "they" would only lose their experiences since the last copy event. I expect evolution to prefer the later set of values over the former.

Perhaps the hardest values to change in uploads will be our deeply-ingrained values for having children. Early upload technology would likely not be able to create a baby's brain from scratch, or even to upload a child's brain and then correctly model brain development processes. And even when such technology is available, children would likely be a poor investment, from a long-term growth point of view. New children may offer new perspectives, but with enough adult uploads, these benefits should only rarely exceed their high costs. Adults can offer new perspectives as well, and can do so cheaply.

Eventually, human-level artificial intelligence may be achieved at competitive hardware costs, or we may learn enough about the high-level organization of our brains to modify them substantially, perhaps merging distinct copies or splitting off "partials" of minds. The upload era would have ended, and many of the consequences of uploads described above may no longer apply; it seems particularly hard to project beyond this point.

But before then the upload era may last a long time, at least subjectively to uploads running at the dominant upload speed. If many uploads are fast, history will be told from the fast uploads' point of view; history chronicles wars and revolutions, triumphs and disasters, innovations and discoveries, and cares little about how many times the earth spins.

Upload Politics

If voters and politicians lose their composure at the mere prospect of genetic modification of humans, or of wage competition by foreign workers, imagine the potential reaction against strong wage competition by "machine-people" with strange values. Uploading might be forbidden, or upload copying might be highly restricted or forbidden. Of course without world government or strong multi-lateral agreements, uploads would eventually be developed in some country, and the transition would just have been delayed. And even with world government, covert uploading and copying might happen, perhaps using cryptography to hide.

Even if you weren't one of the highly-copied uploads, your reduced wage-earning ability would be more than compensated for by your increased income from other sources. You could stop working, yet get richer and richer. By uploading and resisting copying, you could become effectively immortal.

If level heads can be found, however, they should be told that if uploading and copying are allowed, it is possible to make almost everyone better off. While an upload transition might reduce the market value of ordinary people's human capital, their training and ability to earn wages, it should increase total wealth, the total market value of all capital, including human capital of uploads and others, real estate, company stock, etc. Thus it can potentially make each person better off.

For example, if most non-uploads had about the same fraction of their wealth in each form of capital, including owning shares in firms that make loans to uploads, and if a large enough fraction of upload wages went to pay off such loans, then most non-uploads would get richer from the transition. Even if you weren't one of the highly-copied uploads, your reduced wage-earning ability would be more than compensated for by your increased income from other sources. You could stop working, yet get richer and richer. By uploading and resisting copying, you could become effectively immortal.

The per-capita wealth of highly-copied uploads might decline, but that would not be a bad thing from their point of view. Their choice would indicate that they prefer many poorer copies to a single richer copy, just as parents today prefer the expense of children to the rich life of leisure possible without them.

Could a big fraction of upload wages go to paying loans? Yes, if there is enough competition between uploads, and if investors are not overly restricted by law. For example, refusing to loan to an upload if any other copy in their family has purposely defaulted on a loan might discourage such behavior. Alternatively, loans might be made to a copy family as a whole. But these options would have to be allowed by law.

Could most non-uploads sufficiently diversify their assets? Yes, if we develop financial institutions which allow this, such as allowing people to trade fractions of their future wages for shares in mutual funds. But tax laws like those that now encourage highly undiversified real estate holdings could cause problems. And even if people are able to so diversify their assets, they may not choose to do so, yet later demand that politicians fix their mistake.

If forced to act by their constituents, politicians would do better to tax uploads and copies, rather than forbidding them, and give the proceeds to those who would otherwise lose out.[14] Total wealth would grow more slowly than it otherwise would, but grow faster than without uploads. Of course there remains the problem of identifying the losers; politicals systems have often failed to find such win-win deals in the past, and could well fail again.

What about those who have values and abilities compatible with becoming part of the few highly-copied uploads? Would there be great inequality here, with some lucky few beating out the just-as-qualified rest?

If the cost to create an upload brain model from an ordinary brain were very high relative to the cost of creating a copy of an upload, or if computer hardware were so cheap that even the earliest uploads were run very fast, the first few uploads might have a strong advantage over late-comers; early uploads may have lots more experience, lower costs, and may be a proven commodity relative to new uploads.[15] Billions of copies of the first few dozen uploads might then fill almost all the labor niches.

Computer technology should keep improving even if work on uploading is delayed by politics, lowering the cost of copying and the cost to run fast. Thus the early-adopter advantage would increase the longer uploading is delayed; delaying uploading should induce more, not less, inequality. So, if anything, one might prefer to speed up progress on

uploading technology, to help make an uploading transition more equitable.

Similar arguments suggest that a delayed transition might be more sudden, since supporting technologies should be more mature. Sudden transitions should risk inducing more military and other social instabilities. All of these points argue against trying to delay an upload transition.[16]

Contrary to some fears, however, there seem to be no clear military implications from an upload transition, beyond the issue of transition speed and general risks from change. Yes, recently backed-up upload soldiers needn't fear death, and their commanders need only fear the loss of their bodies and brains, not of their experience and skills. But this is really just the standard upload trend toward cheaper labor translated into the military domain. It says little about fundamental military issues such as the relative expense of offense vs. defense, or feasible military buildup speeds vs. economic growth rates.

What if uploads decide to take over by force, refusing to pay back their loans and grabbing other forms of capital? Well for comparison, consider the question: What if our children take over, refusing to pay back their student loans or to pay for Social Security? Or consider: What if short people revolt tonight, and kill all the tall people?

In general, most societies have many potential subgroups who could plausibly take over by force, if they could coordinate among themselves. But such revolt is rare in practice; short people know that if they kill all the tall folks tonight, all the blond people might go next week, and who knows where it would all end? And short people are highly integrated into society; some of their best friends are tall people.

In contrast, violence is more common between geographic and culturally separated subgroups. Neighboring nations have gone to war, ethnic minorities have revolted against governments run by other ethnicities, and slaves and other sharply segregated economic classes have rebelled.

Thus the best way to keep the peace with uploads would be to allow them as full as possible integration in with the rest of society. Let them live and work with ordinary people, and let them loan and sell to each other through the same institutions they use to deal with ordinary humans. Banning uploads into space, the seas, or the attic so as not to shock other folks might be ill-advised. Imposing especially heavy upload taxes, or treating uploads as property, as just software someone owns or as non-human slaves like dogs, might be especially unwise.[17]

The Bottom Line

Because understanding and designing intelligence is so hard, we may learn how to model small brain units before learn how to make human-level A.I. Much will have changed by that time, but an upload transition would be so fundamental that we can still forsee some clear consequences. Subjective lifespans could be longer, minds could run faster, and reproduction could be cheaper, faster, and more precise. With human labor still in demand, an upload population should explode, and Darwinian evolution of values should once again become a powerful force in human history. Most uploads should quickly come to value life even when life is hard or short, and wages should fall dramatically.

What does this all mean for you now? If you expect that you or people you care about might live to see an upload transition, you might want to start to teach yourself and your children some new habits. Learn to diversify your assets, so they are less at risk from a large drop in wages; invest in mutual funds, real estate, etc., and consider ways in which you might sell fractions of your future wages for other forms of wealth. If you can't so diversify, consider saving more.[18]

Those who might want to be one of the few highly copied uploads should carefully consider whether their values and skills are appropriate. How much do you value life when it is hard and alien? Can you quickly learn many new skills? Can you get along with people like yourself? And such people might consider how they might become one of the first uploads.[19] Those who don't want to be highly-copied uploads should get used to the idea of their descendants becoming a declining fraction of total wealth and population, of leaving a rich but marginalized lineage.

If you participate in political or social reform, you might consider sowing seeds of acceptance of an upload transition, and of the benefits of an integrated society, and might consider helping to develop institutions to make it a win-win outcome for everyone. And if you research or develop technology, consider helping to speed the development of upload technology, so that the transition is less sudden when it comes.

Footnotes

[1] This is my impression from 9 years of A.I. research, though of course many A.I. researchers disagree.

[2] We might well have good enough hardware now for a slow A.I. that doesn't deal much with the physical world — say an A.I. contract lawyer.

[3] Consider a model where utility is roughly a product of powers of leisure and consumption, and amount produced is roughly a product of powers of labor and other capital. Such a model can explain why leisure time has not changed much as per capita wealth has increased dramatically over the last few centuries, can explain high leisure among slave owners, and explains why leisure is higher in places and times with high income taxes. One can explain seasonal high leisure among foraging tribes as due to seasonal limits on foraging productivity.

[4] Roger Penrose, in *The Emperor's New Mind*, suggests that non-local corrections to quantum gravity may play an important role in the brain; I find this extremely unlikely.

[5] See [Merkle] for an exploration of the near-term feasibility of this, and [Platt] for a fictional account.

[6] A viable, though perhaps not optimal, alternative is to hold all copies responsible for the actions of any one of them. If punishment is by fine when possible, then copy families could use insurance to contract away this interdependence.

[7] By "values", I mean all preferences, desires, moral convictions, etc.

[8] The Hutterites, a U.S. religion group, has averaged 9 kids per family for a century.

[9] Such a loan might come from the original upload or any other source, and might involve more risk-sharing than a simple loan — more like a joint investment.

[10] Meaning enough so that they can't effectively conspire to keep their wages high.

[11] Thus janitorial jobs should be safer longer than programmer jobs.

[12] These wages are per product produced, not per time spent.

[13] It seems that evolution should favor values that are roughly risk-neutral over the long term, with utility linear up to near the point of total world wealth. This seems to imply values roughly logarithmic in returns to short independent periods.

[14] Note that such a tax would be a tax on the poor, paid to the relatively rich, if one counted per upload copy.

[15] Many initial uploads might well be cryonics patients, if legal permission to dissect and experiment with their brains were easier to obtain.

[16] Note that, in contrast, a delayed nanotechnology assembler transition seems likely to be less sudden, since pre-transition manufacturing abilities would not be as far behind the new nanotech abilities. Efforts to "design-ahead" nanotech devices, however, might make for a more sudden transition.

[17] A similar argument applies to A.I.s capable of wanting to revolt.

[18] This is, by the way, the same strategy that you should use to prepare for the possibility that A.I. is developed before uploads.

[19] Cryonics patients might want to grant explicit permission to become uploads.

References

K. Eric Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, John Wiley & Sons, Inc., New York, 1992.

Robin Hanson, "Reversible Agents: Need Robots Waste Bits to See, Talk, and Achieve?", *Proc. 2nd Physics of Computation Workshop*, 1992.

Ingemar Hansson, Charles Stuart, "Malthusian Selection of Preferences", *American Economic Review*, June 1990, V.80 No. 3. pp.529-544.

Ralph Merkle, "Large Scale Analysis of Neural Structures", *Tech Report* CSL-89-10, Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304, 1989.

Charles Platt, *The Silicon Man*, Tafford Publishing, Houston, 1991.

Julian Simon, *The Ultimate Resource*, Princeton University Press, 1981.

Acknowledgments

This paper is better for the thoughful comments on earlier drafts by Stuart Card, Hal Finney, Daniel Green, J. Storrs Hall, Nancy Lebovitz, Hans Moravec, Max More, Jay Prime Positive, Mike Price, Marc Ringuette, Nick Szabo, and Vernor Vinge, and because of prior discussions of related issues with David Friedman, Keith Henson, Richard Kennaway, David Krieger, Tim May, Ralph Merkle, Perry Metzger, Mark Miller, Ravi Pandya, and Steve Witham. Many of these discussions took place on the Extropians mailing list (extropians@extropy.org).

UTILITY FOG

Part One

J. Storrs Hall

What I want to be when I grow up, is a cloud

Less poetically, consider the kind of technology that can support uploading: enormous computational power; the ability to build billions of cell-sized machines; structures smaller and/or more powerful than those of our existing bodies.

In a pinch, i.e. if I were going to die otherwise, I'd upload onto just about any working platform. But absent that, I'd want some advantage to induce me to take that somewhat hazardous step. In the words of E.E. Smith, I want it to run faster, jump higher, dive deeper, and come up drier than my existing hardware.

If a whole community of people uploads, they can live together in an almost limitlessly rich and varied, shared virtual environment. Indeed, physical humans can participate in such an environment with VR equipment, at least to the extent that the phenomena of the environment are translatable to human senses. But for the upload to participate in the real physical environment, it needs a robotic body to sense and affect its surroundings.

The human body, when you get right down to it, is a pretty nifty gadget. It has some maddening limitations, most of which are due to its essential nature as a bag of seawater. It wouldn't be too hard, given nanotechnology, to design a human body that was stronger, lighter, with a faster brain and senses not limited to such a narrow range of electromagnetic and vibrational frequencies, nor so imprecise in their measurements of what they do perceive — a body able to operate comfortably in any natural environment on Earth or in outer space (excluding the Sun and a few other obvious places). But these would be basically extensions in degree, not kind, to the existing design.

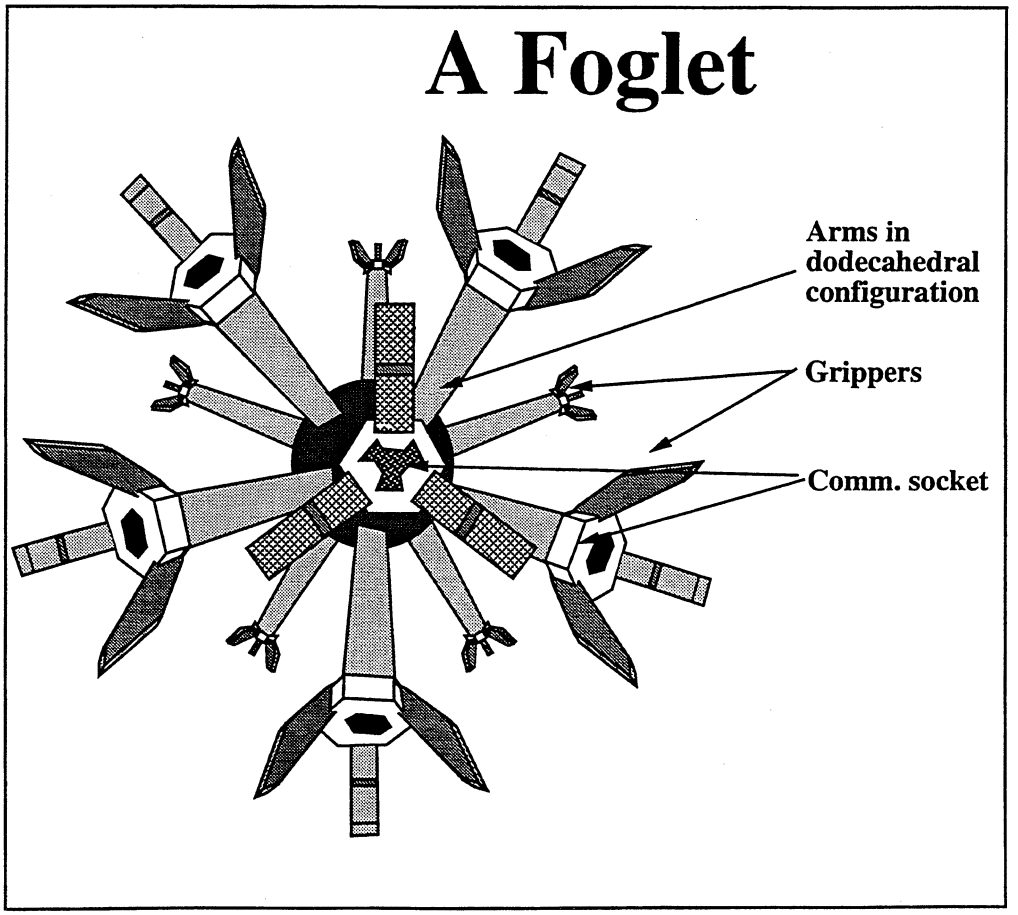
In the virtual environment of the uploads, not only can the environment be anything you like; you can be anything you like. You can be big or small; you can be lighter than air, and fly; you can teleport and walk through walls. You can be a lion or an antelope, a frog or a fly, a tree, a pool, the coat of paint on a ceiling. (Let's assume for the sake of the discussion that you can imagine retaining your mind and senses while taking on the physical form of a coat of paint!)

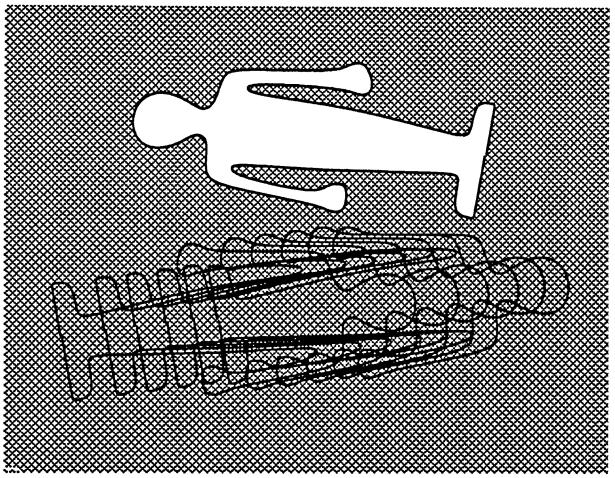
You can be these things in the real world, too, if your body is made of Utility Fog. Utility Fog is an intelligent substance, able to simulate the physical properties of most common substances, and having enough processing power that human-level processes could run in a handful or so of it.

Imagine a microscopic robot.

It has a body about the size of a human cell and 12 arms sticking out in all directions. A bucketful of such robots might form a "robot crystal" by linking their arms up into a lattice structure. Now take a room, with people, furniture, and other objects in it — it's still mostly empty air. Fill the air completely full of robots. The robots are called Foglets and the substance they form is Utility Fog.

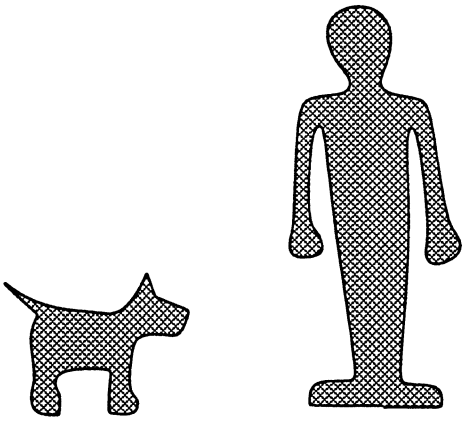
With the right programming, the robots can exert any force in any direction on the surface of any object. They can support the object, so that it apparently floats in the air. They can support a person, applying the same pressures to the seat of the pants that a chair would. They can exert the same resisting forces that elbows and fingertips would re-





Flying

The biological human can fly in a Fog environment the same way that any object can be supported and moved. The uploaded human not only flies but is inertialess.



Naive-mode objects

Here we have a person and a dog, both uploaded into Fog, existing as naive-mode objects. That is, each is a separate, distinct group of Foglets not unlike the cells in a biological body.

ceive from the arms and back of the chair. A program running in the Utility Fog — a very distributed processor indeed — can thus simulate the physical existence of an object.

The Utility Fog operates in two modes: First, the "naive" mode where the robots act much like cells, and each robot occupies a particular position and performs a particular function in a given object. The second, or "Fog" mode, has the robots acting more like the pixels on a TV screen. The object is then formed of a pattern of robots, which vary their properties according to which part of the object they are representing at the time. An object can then move across a cloud of robots without the individual robots moving, just as the pixels on a CRT remain stationary while pictures move around on the screen.

The Utility Fog which is simulating air needs to be impalpable. One would like to be able to walk through a Fog-filled room without the feeling of having been cast into a block of solid Lucite. Of course if one is a Fog-mode upload this is straightforward; but the whole point of having Fog instead of a purely virtual reality is to mix virtual and physical objects in a seamless way. To this end, the robots representing empty space can run a fluid-flow simulation of what the air would be doing if the robots weren't there. Then each robot moves where the air it displaces would move in its absence.

How can (physical) people breathe when the air is a solid mass of machines? Actually, it isn't really solid: the Foglets only occupy about 10% of the actual volume of the air (they need lots of "elbow room" to move around easily). There's plenty of air left to breathe. The Fog, as part of its air simulation, squeezes out appropriate quantities of real air for its occupants to breathe.

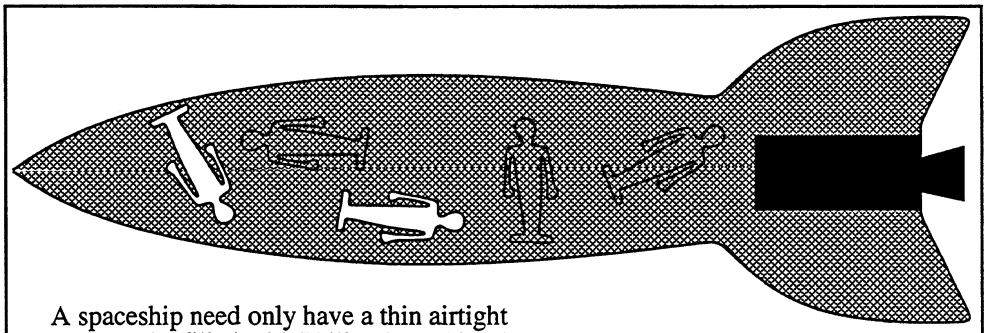
The other major functions the air performs, that humans notice, are transmitting sound and light. Both of these properties are obscured by the presence of Fog in the air, but both can be simulated at a level sufficient to fool the senses of humans and most animals by transmitting the information through the Fog by means we'll consider later, and reconstructing the physical wavefronts of the light or sound at the Fog/air surface.

To understand why we want to fill the air with microscopic robots only to go to so much trouble to make it seem as if they weren't there, consider the advantages of a TV or computer screen over an ordinary picture. Objects on the screen can appear and disappear at will; they are not constrained by the laws of physics (because of course they're not real "objects", but patterns of dots.) The whole scene can shift instantly from one apparent locale to another. Completely imaginary constructions, not possible to build in physical reality, could be commonplace. Virtually anything imaginable could be given tangible reality in a Utility Fog environment.

Remember, though, that the Fog is not virtual but real. An instantly appearing Fog ladder can be climbed; if it disappears, the climber will tumble to the ground (through solid Fog now simulating air). A physical human embedded into Fog who wants to pick up a hot potato has several choices. He can have a thin layer of Fog attach to his skin and simulate gloves (it would be a good thermal insulator). He can materialize a pair of tongs (of Fog, of course) and use them; or he can have the Fog surrounding the potato simply alter its air simulation enough so that it carries the potato wherever he wants, such that it appears to levitate.

General Properties and Uses

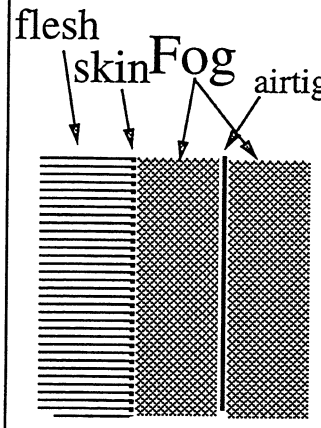
Clearly, if the Fog can simulate people, chairs, and ladders, it can simulate walls, roofs, and doors. "Fog City" need have no permanent buildings of concrete, no roads of asphalt, no cars, trucks, or busses. It can look like a park,



A spaceship need only have a thin airtight skin and be filled with Utility Fog. The Fog provides the structural strength and all interior items, including instruments, bulkheads, furniture, and storage compartments are fog-mode. Fog makes moving in zero-G easy and controlled -- it could even simulate gravity if desired.

Drive systems would need to be true nanotech engineering near the limits of its capabilities.

Utility Fog in a spacesuit

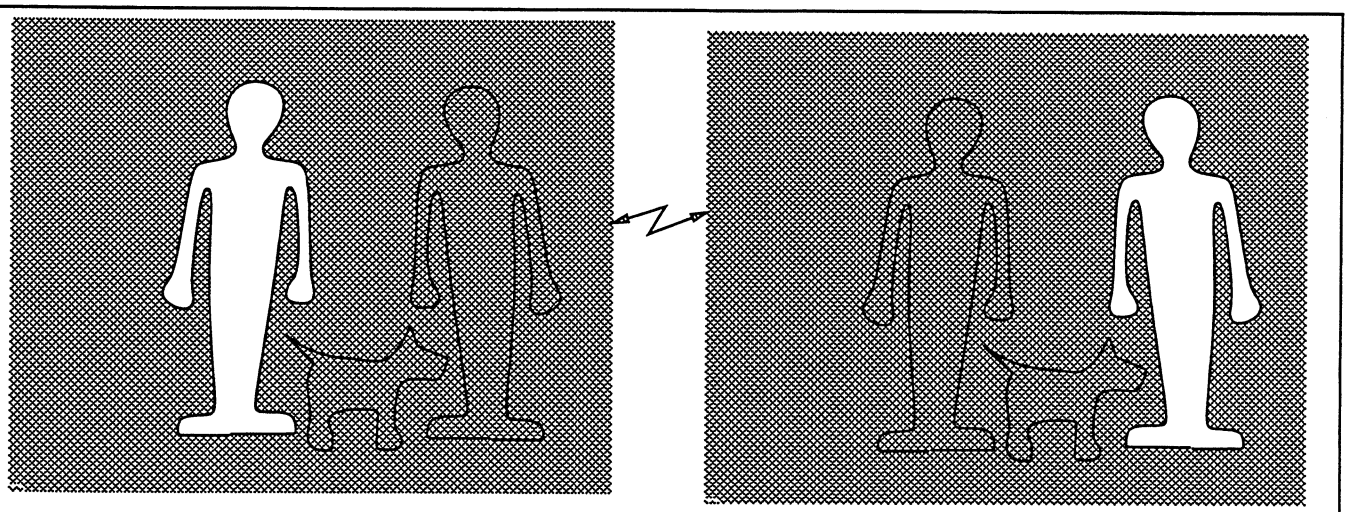


Use of Fog both inside and outside the airtight skin of a spacesuit is beneficial. Inside Fog provides air circulation and temperature control. Outside Fog protects the fabric and provides fine manipulating capability. All the Fog participates in volume and pressure control so the suit is effortless to move in.

or a forest, or if the population is sufficiently whimsical, ancient Rome one day and Emerald City the next.

It will be more efficient to build dedicated machines for long distance energy and information propagation, and physical transport. For local use, and interface to the worldwide networks, the Fog is ideal for all of these functions. It can act as shelter, clothing, telephone, computer, and automobile. It will be almost any common household object, appearing from nowhere when needed (and disappearing afterwards). It gains a certain efficiency from this extreme of polymorphism; consider the number of hardcopy photographs necessary to store all the images one sees on a television or computer screen. With Utility Fog we can have one "display" and keep all our physical possesions on disk.

Nanotechnology will allow us to build some really monster computers (see my previous *Extropy* article on the subject). Although each Foglet will possess a comparatively small nanoprocessor — which is to say the computing power of a current-day supercomputer — there are between a million and a billion Foglets to a cubic inch, depending on how big they are (more on that later). When those Foglets are not doing anything else, i.e. when they are simulating the interior of a solid object or air that nothing is passing through at the moment, they can be used as a computing resource.



Telerobots and virtual reality combined

The two biological humans are at different locations but experience the same virtual location, which also contains a virtual dog. Virtual locations can contain arbitrarily many actual objects which may be at any number of actual locations.

The Limits of Utility Fog Capability

When discussing something as far outside of everyday experience as the Utility Fog, it is a good idea to delineate both sides of the boundary. The Fog is capable of so many nearly incredible things, like nanotechnology in general, that one is tempted either to dismiss it entirely, or to suspend disbelief and regard it uncritically. Neither is appropriate; Fog capabilities do have limitations:

Hard and/or waterproof objects could not materialize instantly. The Foglets at their normal arm's length position form a substance that is about as hard as fingernail, and has air/water properties much like felt. To be really hard or waterproof, the Foglets must pack themselves down to a solid mass where all their bodies are touching; this would take as much as a tenth of a second for something the size of, say, a coffeecup.

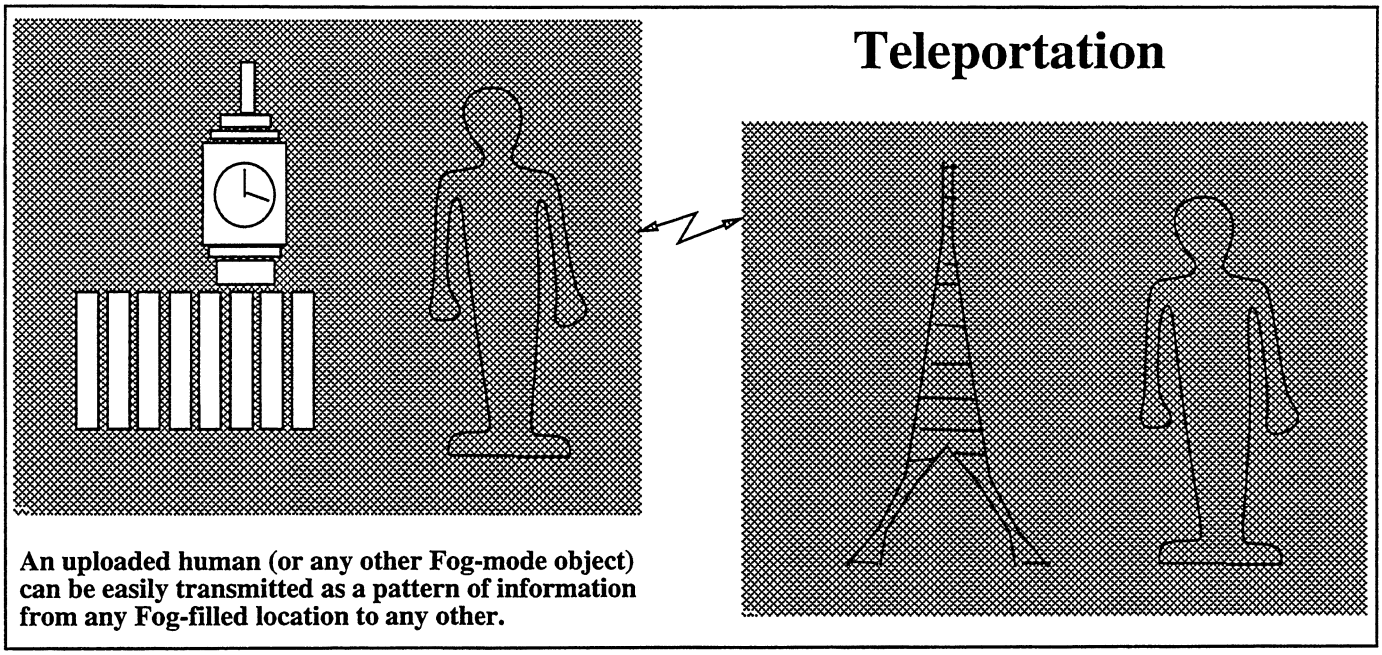
Modern materials having both high strength and low volume could not be simulated directly. Packed Fog would have about the strength and weight of aluminum, and would not be dynamically polymorphic. Fog could not substitute directly for Nylon, Kevlar, or indeed steel.

The internal mechanisms of Foglets would probably be designed to operate in a specific temperature range, particularly the nanocomputer controllers. More than likely this range would be optimized for everyday conditions; it would be possible to design Foglets that could operate at Fahrenheit 451 (or 1000) and thus simulate flame, but ones that could not would be a lot more efficient under normal conditions.

Foglets are not assemblers. They are on the wrong scale to manipulate atoms directly; it would be like working on a wristwatch with a construction crane. On the other hand, they can do chemical things like preparing food the same way a human cook does them — by mixing, stirring, and using special-purpose devices that were designed for them to use.

Fog cannot simulate food, or anything else that is destined to be broken down chemically. Eating it would be like eating the same amount of sand or sawdust.

Fog can simulate air to the touch but not



An uploaded human (or any other Fog-mode object) can be easily transmitted as a pattern of information from any Fog-filled location to any other.

to the eyes. The surface of raw Fog might look like mother-of-pearl or white sand, depending on the size of the Foglets. Individual Foglets can be made small enough that they are invisible; however, the mass of them would scatter light like the droplets in a cloud. Thus when the appearance of transparency is desired, Foglets on one side of a mass register the light that's coming in, transmit a description of what they see through the Fog on its regular data communications network, and the Foglets on the other side produce what is essentially a hologram on the Fog surface. The hologram cannot recreate each wavefront of the incoming light exactly; there isn't enough data bandwidth to transmit that much information in the network. However, there's enough to fool human senses.

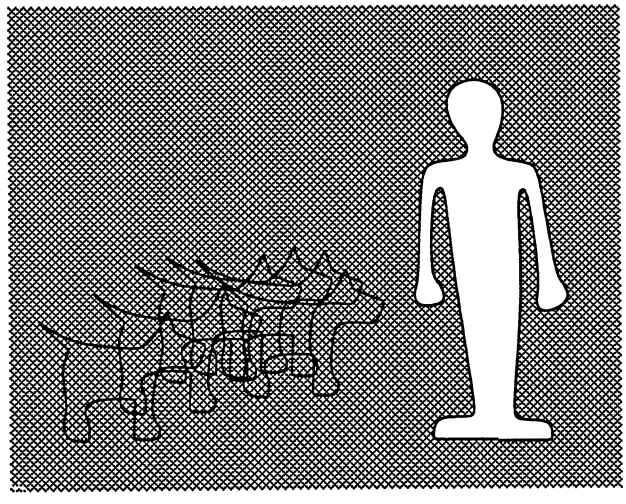
Living in Uploadia

For the person uploaded into the Fog, the whole tricky simulation of sight, sound, airflow, and so forth can be dispensed with. The region of Fog running the person's mind process simply takes the data inputs in native form; the upload can have this interpreted as human senses if it wants, or can have a sense of 3-D perception a la E. E. Smith's Rigellians, or "X-ray vision" like Superman, or however else you care to experience the knowledge of where everything is and what its surface properties are.

The flip side of this is that the upload doesn't really have any need to simulate the appearance of a body, either. You'd simply be a "presence", immanent in a given location so to speak, knowing everything that transpired there, but also able simply to make things happen by wanting them to. The Fog over a relatively wide area could be your "body", sensing and affecting things, materializing any desired physical manifestation at will.

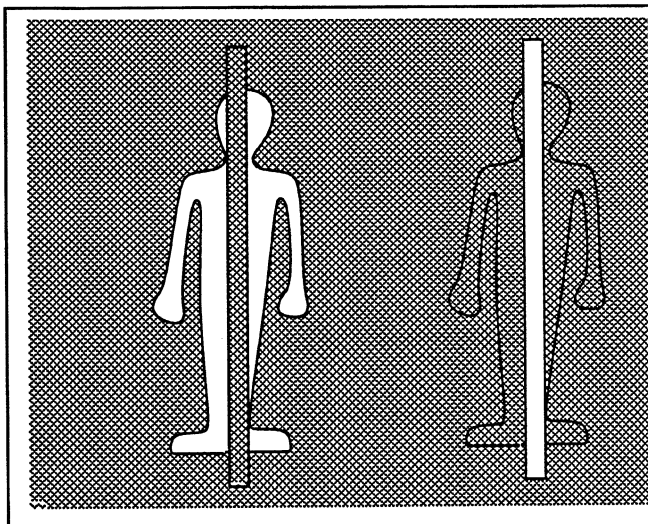
Of course what is really happening is that there is some group of Foglets that is running the program that is your mind, and instead of sending nerve impulses to muscles, it is sending data packets in the FogNet to regions of Fog that are desired to do something. Similarly, each bit of Fog would broadcast descriptions of its physical situation, so that all uploads in the vicinity would receive a constant update.

This isn't different in principle from each bit of surface reflecting light in all directions so that all eyes in the area can see it; the



Virtual Reality

Here a biological human is embedded into a roomful of Fog. The Foglets making up the dog do not move but the pattern of them changes position like a picture of a dog running across a television screen.



Walking through walls

A biological human can walk through Fog walls, and a Fog (uploaded) human can walk through dumb-matter walls. (Of course Fog people can walk through Fog walls, too.)

reception of the data in the upload would presumably be just as unconsciously integrated into a world model within the mind. It would simply be much more complete.

The best estimates I can come up with are that it would take somewhere in the neighborhood of a cubic inch of Fog to support a human-level upload. This depends on a lot of factors that are quite variable about the Fog itself, as well as uncertainties in the nature and quantity of processing power necessary for uploads. The kicker is of course that nobody is going to be satisfied with merely human-level mentality given the ability to do better. I sure won't!

There are several choices available. A roomful of Fog would have a truly astounding computing capacity. On the other hand, a cubic inch nanocomputer with serious power and cooling would also have astounding computing capacity. Either would be nice to live in; both would be better. Fog can easily support, move, protect, and interface with non-Fog computers that are the loci of people's minds. However, we give up some of the non-physical aspects of Fog existence; it's harder to teleport, for example.

Space Exploration

Once we are away from a domestic environment where there is already Fog everywhere, and most objects and people are Fog-mode virtuals, we need to worry about the problems of naive-mode and mixed-mode operation. Suppose you're an upload and you want to colonize the Moon. You'll look a lot like an enormous amoeba. You're a big batch of Fog with a bunch of embedded special-purpose devices: nanotech factories for making Foglets, processing plants for lunar material, propulsion units, power plants, etc.

Mining is easy. You have special purpose tools that can break rock into small chunks. You dig a hole and flow into it. Fog transports the rock chunks to the mineral processing units, moves the refined material and the waste to the appropriate places, and supports the walls of the tunnel, essentially filling it, with the special purpose tools and processing units floating around.

When you're tired of being a mine you can be a spaceship. The major systems of spaceships will need to be made with special-purpose nanotechnological mechanisms, and indeed with such mechanisms pushed much closer to their true capacities than anything we have talked about heretofore. You probably wouldn't normally carry them around; they might be something closer to an actual physical possession, a concept we'd almost forgotten.

Like the mine supports, your Fog body is the structural strength of the ship itself; the rest of the structure need be not much more than a balloon. Actually, unless you want to carry physical animals or other planetary surface items, you can dispense with the balloon and the air; Fog, as we assumed implicitly above, works fine in vacuum.

In the somewhat more distant future, we could turn whole planets — the Moon, for a start — into Fog, and be Fog-mode creatures roaming around inside. If you turned the Moon into Fog it would expand, Fog being a bit less dense than the Moon's current substance, and each existing human could upload into a cubic mile of Fog. And that's something to conjure with.

Next time: Technical details.

Uploading

"Uploading" means copying your mind from your brain into some artificial processor, hopefully one that is faster, more capacious, or more durable than the original.

There are several notions about how this might be done. First, we might build an artificial brain with a neuron corresponding to each neuron in your present brain, making sure that all the connections are the same and that the input/output function of each new neuron is the same as that of the corresponding old one.

Next, we might take some clump of neurons and build a machine that mimics the behavior of the clump as a whole (at every point on the interface between the clump and the rest of the brain) but not have to reproduce the internal structure exactly.

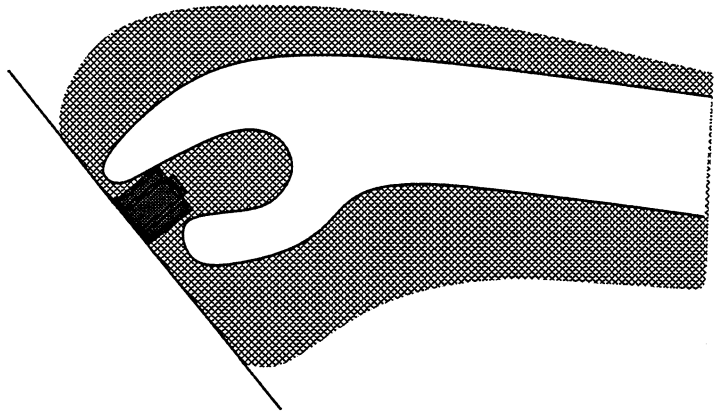
Given this freedom of internal mechanism, we can imagine simulating some clump of neurons on a sequential computer, if the computer were fast enough, in real time. Then a network of such computers, each simulating a little piece of the brain, could, if connected appropriately, simulate the whole brain.

This would have the advantage that the computer network, having its function defined by software, could simulate anybody's brain, given the right software. So the uploading problem comes down to reading your brain, adducing software to match its function, and loading the software onto the computers.

The Foglets of Utility Fog come equipped with substantial processing power, probably about a giga-op in current-day parlance. Estimates of the amount of computing power needed to upload a human brain range from 10 tera-ops to a million times that. Assuming 100-micron Foglets, we have a tera-op per cubic millimeter so it might take between ten cubic millimeters (half an inch of pencil lead) and ten cubic decimeters (two and a half gallons) of Fog to run your mind. The higher estimates involve simulating each neuron in excruciating detail, so my guess is that the the lower ones are fairly reasonable.

For more reading relevant to uploading, see *Mind Children* by Hans Moravec, *AI: The Tumultuous History of the Search for Artificial Intelligence by Daniel Crevier*, and "Reverse Engineering the Brain" by Ralph Merkle, (to be found, among other places, in *Proc. AIAA Comp. in Aerosp. IV*).

Fog Spacesuit -- glove closeup



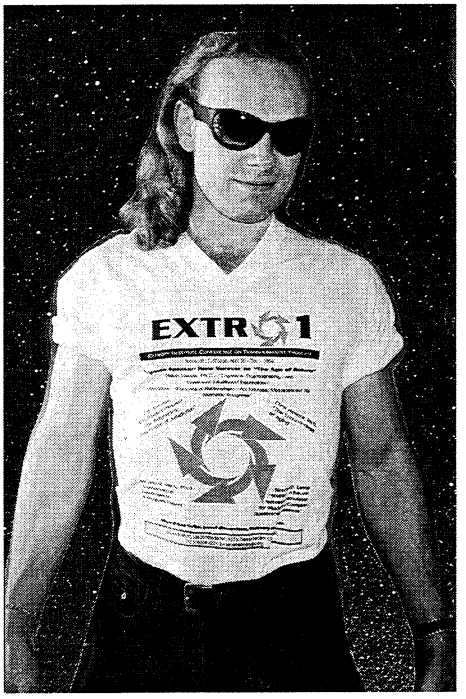
Gloves are tight-fitting and the Fog conforms to the manipulation from outside, making fine work easy. Just as on Earth it can amplify force, making wrenches unnecessary.

Utility Fog in New Book

There will be a chapter about Utility Fog in *Nanotechnology and the Culture of Abundance*, B.C. Crandall, ed., forthcoming from MIT Press. The chapter deals with Utility Fog completely from the perspective of interfacing organic humans to a Fog environment; no uploading.

EXTRO 1 T-shirt

T-shirts cost $15 each ($14 for ExI members), postage included.



To order your EXTRO 1 T-shirt, send a check or cash to:

Department S, Extropy Institute
13428 Maxella Avenue, #273
Marina Del Rey, CA 90292

Two Questions for Extropians

by Charles Platt

1. The Non-Word-Driven Mind

Various attempts have been made to subdivide mental activity. In Freudian terms we have the ego and the id; in physiological terms we have lower and higher brain function; and so on. Most of these schemes seem an attempt to articulate something which we vaguely sense about ourselves: the difference between our animal inheritance of needs and fears, and our evolved, verbalized, conscious human intelligence. To draw the line as clearly as possible and minimize semantic quibbling, I will distinguish these functions by referring to them here as word-driven and non-word-driven brain activity.

I make a living as a writer. When I am working, my thoughts exist primarily as streams of words. Perhaps because of this, I always used to assume that a lot of mental activity in everyday life falls into this same category. However, some brief introspection showed me that I was wrong. Everyday tasks such as measuring ingredients for a meal, sealing envelopes, or dialing phone numbers seem to require no silent mental verbalization at all. (In fact, they can take place while the brain is running some other word-script entirely.) If my experience is typical, most of the thinking that we do, in order to function on a daily level, is not word-driven.

This seems to apply even in tasks that are highly evolved or complex, such as driving a car, composing music, doing carpentry, cleaning the house, or painting a picture.

Well-known signing experiments with chimpanzees have confirmed (reluctantly, in some cases) that language is probably the decisive attribute dividing human beings from animals. Since animals do not have language as we understand the term, by definition, none of their actions are word-driven.

However, rats can still learn mazes, and in my own home, I have seen our three cats "figuring out" problems such as opening a door or perfecting a hunting strategy.

From this I conclude:

1. Everyday human activity is similar to animal activity in that it is not word driven, even though it may be complex and demanding.

2. Learning a task may be faster if we use words to define and control it; but even here, language is inessential, and deductive reasoning can occur without words.

How does this affect the Extropian desire for transcendence?

I suggest the desire for "pure intellect" may be a contradiction in terms, and may derive much more from wish-fulfilment than from a rational understanding of what it means to be alive.

Suppose we contemplate an idealized Extropian model: a human brain replicated as data. This appeals to me personally (provided the replication is complete and accurate) because it implies unlimited life and an unlimited range of experiences. When I wrote my novel *The Silicon Man*, I presented it as an optimistic vision; a form of liberation.

Other writers have extended this scenario. In a chapter that I have seen from a forthcoming book by Hans Moravec, he points out that since the mind does not easily tolerate being deprived of sensory input, information entities (or "infomorphs," as I like to call them) will need the active, sustained illusion of sights, sounds, and other sensation. Maintaining this illusion will require a heavy overhead of processing power; and therefore, if there is such a thing as a non-human infomorph operating as "pure" intellect without need for the sensory-simulation overhead, Moravec suggests this entity will have a competitive evolutionary advantage relative to ourselves.

This concept of "pure intellect" as being somehow "superior" recurs implicitly in many Extropian writings. It seems that Extropians are so conscious of physical human limitations, they are eager to jettison not only the unreliable biological support system of the human body, but also the "messier" features of the human mind. As Max More remarked to me recently, when he was asked what he was going to be doing later in the evening, "I have to go to sleep because, unfortunately, I am human."

I tend to share this instinctive bias. However, as I have suggested above, word-driven, "higher" brain activity takes a less active role even in logically deductive tasks than we might like to imagine. Non-word-driven or "intuitive" functions may be much more than background noise interfering with a nice clean thought-signal. This noise may, in fact, be an integral part of the signal. In other words, non-word-driven processes may constitute the primary activity of the brain and may be the key aspect of being human, without which we would be unable to function at all.

I suggest the desire for "pure intellect" may be a contradiction in terms, and may derive much more from wish-fulfilment than from a rational understanding of what it means to be alive.

However, as I mentioned at the beginning, I earn a living as a writer. I trust word-driven thought more than non-word-driven thought. Therefore, I hope my conclusion is wrong.

2. First, Do No Harm

The libertarian ideology, which recurs frequently in visions of an Extropian future, argues that individuals should be unconstrained so long as their actions cause no harm to others. I propose that this ideal is unrealistic, since some degree of harm is inevitable; and ethical individuals should acknowledge this and accept limits to their behavior, especially if we acquire greater personal power.

Consider the specific issue of free speech. Libertarians are understandably discontented when courts rule that some speech is unprotected by the First Amendment. I share their discontent, because as a writer, I want to write whatever I feel like writing. (In my British homeland, one of my raunchier books was seized by the police, and I narrowly escaped a fine or imprisonment under the Obscene Publications Act. This kind of experience makes a durable impression.)

The principle that some speech is not constitutionally protected goes back to World War I, when agitators were jailed for distributing leaflets advocating pacifism (which supposedly endangered the nation by threatening to undercut the war effort). As patriotic fever declined and rationality reasserted itself, the Supreme Court reached a compromise articulated by justice Oliver Wendell Holmes, who argued that speech should only be unprotected if it posed a "clear and present danger." This principle has survived to the present day.

The commonly understood implication is that you should be free to shout "Impeach the president!" in a crowded theater, but if you shout "Fire!" (another test invented by Holmes), that speech creates a clear and present danger; therefore, it is not protected by the

first amendment, and you may be sued for damages by people who are injured in the stampede. At first glance this seems to fit the libertarian ideal, since speech is unconstrained so long as it does no harm to others. But as soon as we allow the underlying principle that speech has the power to do harm, logic leads us to some difficult conclusions. At one extreme, Dworkin and McKinnon have proposed that men may be encouraged to assault women if they view a lot of erotic magazines in which women are depicted as sexual victims; therefore, women should be able to claim damages from people who produce the magazines.

Is this argument as extreme as it seems? Do magazines depicting the torture of women create the "clear and present danger" of violence against women? The ACLU refuses to accept that this linkage can occur. However, undeniably, art has power, and there are certainly cases of "copycat crimes" based on fiction. I myself have written at least one book which I believe could tempt an unbalanced person to murder or rape. Do we blame the person for being unbalanced, or do we blame the writer for indulging his fantasies with no concern for potential consequences? If I shout "Fire!" in a crowded theater and an "unbalanced" person has a heart attack, isn't that a similar situation? Just what is a "clear and present danger," anyway?

Having thought about this extensively, I conclude that creative artists have a moral obligation to consider the effects of their work; but their position is not so different from that of any citizen who interacts with other people. After all, almost *any* speech has the potential for harm. I can go out around town and be surly, obnoxious, and insulting, angering people and raising their blood pressure, which will be bad for their health and may encourage them to take out their anger on others in turn. Or, I can spread good karma wherever I go.

The same principle applies to other social freedoms. We can use them, or we can abuse them. But the potential for harm is there, and it's all just a matter of degree.

How does this apply to Extropian thought?

Extropians seem to accept the libertarian principle that it is possible to enjoy greater freedom without doing any harm at all. However, as I have argued above, in the case of at least one freedom (freedom of speech), the potential for harm is always present. A writer has more power than average, to create that harm. By extension, in an Extropian future where technology has enhanced personal power much farther, the potential for causing harm will be greater still.

I sense that this is an unpopular train of thought in Extropian circles because it involves community considerations and threatens to inhibit us. As individualists who seek to transcend most forms of inhibitions, the last thing we want to hear is, "You shouldn't do that, because other people might not like it."

However, in the words of an old Frank Zappa lyric, we are the other people, and you're the other people too. Few of us can genuinely claim that we look forward to living out an unrestricted lifespan in total isolation. Unless we somehow rewire our brains at a very fundamental level (which may be desirable, but presents some practical problems), we tend to need social interaction.

I would like to see the development of an "Extropian Ethic" addressing the issue of interpersonal harm on a more realistic level. This will be all the more urgent if individuals turn into infomorphs, since an information entity would be more vulnerable in some ways than a physical human being.

I suggest two guiding principles for those who want to transcend limits and indulge their freedoms:

Response to "Two Questions"

by Max More

1. The Non-Word-Driven Mind

Charles Platt states that a concept of "pure intellect" as superior recurs in many Extropian writings. I can see some ground for this assessment, especially in the writings of Hans Moravec and a few others who have discussed the idea of uploading their selves to better hardware (becoming uploads, or "infomorphs", to use Charles handy term). However, my experience doesn't suggest to me that most Extropian-minded thinkers take this view. I find it difficult to assess Charles claim since he cites only Moravec's writings and my half-joking reply implying frustration with my human biological and neurological limits. Though I doubt that many Extropians assume this view of pure intellect in the way Charles suggests, I thank him for bringing this issue out into the light where we can examine it explicitly. I welcome others to respond to Charles' concerns from their own perspective; my own comments will be brief.

True, we Extropians dislike the "messier features of the human mind" (and body), but that isn't a rejection of the unconscious or less conscious aspects of mind. In my view, this dislike is the flipside of a desire to fine-tune the mind or self – to eliminate unnecessary aspects and to replace inefficient processes with more efficient means. Eliminating sleep (or compressing its useful functions into a shorter duration) is one thing, but this doesn't imply a desire to eliminate all the machinery underlying and supporting our conscious mentation. We need to distinguish the goal of transcending our human biological limitations from a rejection of the unconscious or less-conscious.

Related to Charles' concern, I worry that some uploading enthusiasts will leave behind important parts of themselves if they attempt to simulate only the top level of their thinking – just enough to produce plausibly similar behavior to their previous human selves. Only further research into the nature of cognition and sensation will tell us how deep down we need to go in simulating the brain (and endocrine system) if we are to conserve all that matters to us about our selves and our experiences. Just how deep we need to go has been debated on the Extropians e-mail list a number of times; there is no space here to get into the topic.

To illustrate the difference between abandoning the richness of our human experience and expanding and refining it, let us look at the case of emotions. Although emotions can (according to the cognitive psychologists) be modified by conscious thought, generally our emotional responses emerge from unconscious mental activity. Below the level of awareness, or at a dim and nonverbal level, we evaluate perceived events as harmful or beneficial, appropriate or inappropriate, rewarding or frustrating. I – in common with most Extropians, I believe – do *not* share the Mr. Spock-style dichotomy and antagonism of reason and emotion. Rather than eliminating emotions in favor of an imaginary (and perhaps inconceivable) pure intellect, we seek to bring them under more control. We wish to be able to intervene more effectively in our emotional responses, by means both psychological and technological, to overcome unhelpful urges and responses built into us by evolution or absorbed unwittingly from our environment. We wish to open ourselves fully to positive emotional experience, while moderating, eliminating, or transforming negative emotional experiences when they serve no useful purpose.

Again, although I don't think *most* Extropians are pure intellectualists, Charles' concern may be supported by noting that creativity is largely unconscious, and perhaps has to remain that way. It may be that being fully conscious of creative processes may stifle them by interfering with the flow of thoughts. Creativity seems to be one of those non-word-driven aspects of mentation, but I don't recall any Extropian writer rejecting unconscious mechanisms of creativity in favor of purely conscious, linguistically-based creativity.

So I agree that the desire for pure intellect is unrealistic, but I don't believe that many Extropians have proposed this. Most Extropians of my acquaintance clearly enjoy the less-conscious and the non-rational (not *irrational*), sensual

Platt cont.

1. Admit that any action may cause harm, and assess this harm before taking the action, using the old "Golden Rule" principle.

2. Never intentionally corrupt data.

At first glance it may seem that the second principle has nothing to do with the first, or with any of my preceding argument. But in its broader, most general terms, the second principle actually subsumes the first principle.

Chemical states in the brain constitute data. Suppose we measure human behavior in terms of efficient operation, just as we would measure the performance of a computer. A contented individual who does not feel threatened will almost always be more mentally productive and will live a longer average natural lifespan as a result of reduced stress. Child abuse, threats by authority figures, frightening events of all kinds, disappointments, abandonment, loss – all these experiences degrade mental performance by disturbing the electrochemical balance of the brain. In this sense, then, such harmful experiences can be seen as corrupting mental data. True, hardware, software, and data are not cleanly separable in the brain; but if we envisage human beings converted to infomorph status, the principle becomes a lot clearer.

Using this model, the "harm scale" can now be defined as follows:

Maximum harm: total loss or corruption of data, as occurs in biological death.

Minimum harm: Alteration of one bit of data in such a way that mental performance is likely to be degraded.

We can now attempt to construct a more detailed ethical code based on the quantity and the quality of data corruption. (Not all data bits are equal; some are more vital than others.) Erasing an infomorph – destroying all data – would be tantamount to murder. Rewriting a few bits in an emotion register would be perhaps similar to shouting at someone and making him flinch. Other forms of harm will be located between these extremes.

I find it interesting to wonder how a society of infomorphs might implement these ethical principles as laws. I also wonder how "sociopath" infomorphs would be dealt with.

Here in the everyday world, convicted criminals face punishment, restitution, or rehabilitation.

Punishment supposedly serves as a disincentive, though many criminologists doubt that it works, and it tends to dehumanize or corrupt the people who administer it. A strict disciplinarian outlook, being fundamentally repressive, seems at odds with the Extropian ideal. I would hope that an infomorph society would find more civilized ways of coping with sociopathic behavior.

Restitution seems to make better sense — except that it forces juries to place a money value on human suffering, with unpredictable results. Also, in an information universe, the concept of wealth becomes rather nebulous. Voltage, processing power, and storage will be cheap, and physical wealth, being located outside the system, may seem irrelevant.

As for rehabilitation, it hasn't worked very well in the physical world; but if the brain processes of infomorphs became be readily accessible and adjustable, there would be obvious opportunities for behavior modification. This, however, seems the most frightening scenario of all: a golden opportunity for thought control implemented by info-police.

I conclude that the popular vision of a transcendent libertarian society populated by creative entities moving freely and doing no harm to each other is unrealistic, since some degree of harm is inevitable, and it will become a greater danger as personal power increases. A society of information entities would also be more vulnerable to damage, and would have a difficult time controlling sociopathic behavior without quickly degenerating into a fascist dystopia.

I think it would be useful to devote some more thought to these problems.

More cont.

aspects of life. A few may look forward to existing as purely conscious intellects, devoid of emotions, occupying themselves throughout eternity with mathematical and logical explorations. Most of us anticipate enjoying life more richly than ever, supplementing our logical and rational cognitive activities with enhanced and refined physical, emotional, sensual, and aesthetic experiences. We will cautiously trim away some of the unnecessary biological and neurological features of our human inheritance, but we will see that other, treasured aspects of our nonverbal, non-rational selves blossom into a glorious posthuman panoply.

2. First, Do No Harm

I'll be even briefer in this part of my response, leaving the issue open to others to respond. While I agree with Charles in terms of overall sentiment and in his view that an artist, writer, or other communicator has a responsibility to consider the effects of their work, I disagree on several specifics.

Stating the libertarian thesis as recommending that "individuals should be unconstrained so long as their actions cause no harm to others" strikes me as incomplete at best. The notoriously vague term "harm" is given a specific interpretation in libertarian theory. Most libertarians hold that the only forms of harm that may be legally restricted are nonconsensual physical harm and its threat, fraud, and damage to property. Disregarding this private property rights-based understanding of harm gets us into trouble with issues like free speech.

The Supreme Court's view that speech is unprotected if it poses a "clear and present danger" will be only a secondary heuristic to libertarians. The reason why you can't shout "fire!" falsely in a crowded theater is because, unless the owner has stated explictly otherwise, such an ejaculation amounts to violating the rights both of the theatre-owner and her customers. Each person buying a ticket is agreeing to generally accepted conditions (again these may be overridden by a statement from the property owner). The clear and present danger may help us know when implicit conditions have been contravened, but they are secondary to the property rights involved. If the owner placed a big sign at the entrance saying that patrons were free to shout "fire!", no one would have a legitimate complaint, even if such behavior did pose a clear and present danger.

As stated at the outset, I agree with Charles' underlying sentiment that creative artists should pay attention to the effects of their expression on others. Taking refuge in "artistic freedom" is simply an evasion of responsibility. I had occasion to think about this recently, while watching a documentary on legendary film-maker Leni Reifenstahl. Reifenstahl, who made excellent movies for the Nazis, adamantly held to the view that she was only doing her job and had no interest in politics. While I believe it mistaken to condemn as just like the Nazis, she can reasonably be criticized for reckless pursuit of her career.

I worry that Charles does not make a clear distinction between our ethical responsibilities and the laws we should seek to enforce (or to buy in a polycentric legal system). For instance, he wonders how, in the future, we "might implement these ethical principles as laws". As I see it, no matter what you write or say, if you do not directly cause harm (in the above sense) and do not organize those who cause the harm, then you may be subject to ethical sanction but should not be subject to legal sanction. Everyone can choose how they respond to the utterances of others; one person can be another's tool only by choice.

Though I cannot tackle the question here, I hope to see more discussion of the problem Charles raises of: what are we to do with sociopathic infomorphs or other posthumans in the future? I loathe the idea of giving any agency, governmental or private, the power of life and death or that of involuntary personality alteration. Exile seems a possible option; another might be offering the offender a choice between indefinite incarceration and personality modification.

Souls, Cyberspace, Sins, & Singularity
A Conversation With Dave Ross
Part 2
by David Krieger with Gayle Pergamit

Dave Ross founds companies for a living. He co-founded Palantir Corporation, which became Calera Recognition Systems, the longtime leader in optical character recognition software. He also founded Arkenstone, a non-profit corporation which makes reading machines for the blind, and his current venture, RAF Technologies. Previously, at NASA's Jet Propulsion Laboratory, he formulated the 'prime rib' technique of selecting orbital rendezvous trajectories, still the standard method for planning solar system missions. I interviewed Dave at the home of Gayle Pergamit, co-author of *Unbounding the Future* and co-founder, with her husband, economist Phil Salin, of the American Information Exchange online information marketplace. In Part One of this interview, we discussed Dave's early career and the paths that brought him to both Christianity and Extropianism. In Part Two, we discuss Dave's 1992 Eris Conference talk "Seven Paths to Immortality," in which he talked about longevity technologies, uploading, cryonics, and how he views these technologies from a Christian perspective. Along the way we discover why Extropians change jobs so frequently, and who gets the "essence" when three philosophers share an artichoke.

There's another direction in this, which was of course the origin of my talk at Eris: looking more at what a human being is, and tying together the Extropian things, from life extension through cryonics through human uploading and all of that, in the Christian viewpoint.

I'd like you to elaborate more on that.

This actually came about from Jim Bennett; we were sitting around last spring and he asked me if I had a topic for Eris this year, and I said, "No, not really." He said, "Would you let me suggest one?" I said sure, and he said, "Well, you're a Christian, which makes you unusual in our group of friends." He said that he had been thinking, more and more, "Y'know, the more I've learned about information theory, and the more I've read stuff like Ralph Merkle's on information-theoretic death[1], the more and more the information content that makes up a human looks like a soul to me," and he said, "Could you do something with that?" I said, "Okay!" So I've started thinking about it more seriously, or rather, more directly, and he's correct, there are phenomenally interesting parallels.

My argument, or my discussion, is not meant of course as a mathematical proof of anything, but it's more to suggest and show how things seem to work. If we start at the very basic level of, "What is aging? Why do we get older? Why do we die? Why do things fall apart?" to a tremendous amount it seems to be such things as imperfect copying, imperfect replication of DNA, cells that don't quite get copied correctly, in which the protein balance is wrong and so they do weird things, or cells that just go wacko and reproduce without any consistency—they generate cancers and tumors and other things like that. In all of these cases, if you could take one of those cells and restore it to the way it had been—in other words, recover the information as it should have been, before it got distorted—you'd be back where you'd started from. Tautologically.

I say, "Well, that's interesting." So you look at taking antioxidants which stop the breaking of some of the bonds by free radicals which scramble information. So you're intending, at a very gross level, to stop or slow down the degradation of the information content. Interesting. So then I went a step farther, and I said "What are we talking about doing with nanotechnology?" The whole idea of nanotechnology is, we want to take small systems and program them to do specific tasks, and then have them go and do those tasks. Among other things we want them to do is to build other machines, replicating machines—we want to build useful machines, but we also want them to go and do cell repair.

Well, what is cell repair, what does "cell repair" mean? It means putting it back the way it was supposed to have been, or maybe even improving it "Putting it back the way it was supposed to have been" means reconstructing the information that was there, perhaps by voting from neighboring cells, perhaps by understanding the degradation process so in some ways you can reverse it[2], perhaps by combinations of all these, perhaps by sampling now and correcting later—lots of different ways of doing it. But what we're trying to do is restore the information content. I said "Well, that's interesting."

Now, what happens if we go to "Well, we want to improve on the human." What does "improving" mean? It means making them better able to do the things that we want to do. That means building on the knowledge base that we have, adding to that knowledge base, and then going out and instantiating it in physical objects. Okay, what about cryonics? What's cryonics trying to do? Cryonics is the idea that, particularly from Ralph Merkle's paper[3], that a human is really dead when there does not exist sufficient information to reconstruct them, i.e., when the information content is sort of dissipated in entropy, and, if you look at it from an information-theoretic point of view, information is in some sense the opposite of entropy. Mathematically it *is* the opposite of entropy, but that's a little fortuitous rather than real, but still, there's something to the concept that information and entropy are sort of enemies of each other—if not opposites, they're enemies.

So if you have taken a person, and the organization that makes up the brain, the individual cells and all that, is allowed to dissipate to nothingness, or degenerate into chaos, with the idea being that there's no way to get the person back, the person is gone. The converse is also thought of in there, that if it isn't, if the information still survives, then it ought to be possible, if by no other means than simply putting every atom back where it was supposed to be, to get the person back. While I don't believe that if you simply took all of the atoms that are *here* and put them back *here*,

that you'd be where you were, because there's such things as motion and temperature and so forth, it's at least a good start, and if you extend the information content concept to mean not just positional but all sorts of other information, then, yeah, that starts making a good deal of sense.

Then I got to thinking further. I had given a talk about two years earlier on the "Age of Magic:" what happens after the Information Age. That was the talk where I met Max, when I gave it later at the Nock Forum. Reading from *Mind Children* and thinking about it some more beyond that, was the idea that humans are going to have about four choices as machine intelligence—there's four things that may happen as machine intelligence continues to rise. First of all, it may be that intelligent machines are for some reason impossible—intelligent algorithms are for some reason impossible, in which case the question doesn't come up. I don't happen to believe that's true, but that's one of the four possibilities. Second possibility, it may be that there's a limit to the size, to the complexity, of an intelligent program or intelligent system, somewhere near the human level, and that evolution sort of stops at that level. Then what happens is, humans are sort of left behind because of speed considerations rather than because of intelligence considerations, but still, interesting things take place on a level that's faster and faster and faster than humans can deal with, and we sort of get left behind out of boredom because all of the excitement's going by a million times faster.

The third option is that, no, there is no limit, or the limit for an intelligent system is far more complicated than a human and far more complex, and that we get left behind not in speed but in intelligence, and that's more ominous, because it's not a case of being ignored, it's a case of being supplanted. That's even more dangerous, from our perspective. Then there's the fourth, which is if you can't beat 'em, join 'em, and that is the straightforward path; I've done work in adaptive technologies. One of the companies that I work for does reading machines for the blind. The idea of transposing senses—in other words, you're taking the material that's on a sheet of paper that's normally visual, and having a machine translate it into auditory, i.e., it reads it aloud, so that you've done synesthesia. You've intentionally switched senses.

If you can give artificial senses to people who are missing them, and there's been work done on giving vision to the blind and so forth—If you can do those things, then you probably then can at least do some movement in the direction of giving senses that don't exist, to people who are not deprived of normal senses. So you can do sensory enhancement, so you can improve the brain. And then I look at it, from the article I wrote in *Extropy*, okay, let's go one step farther than that, let's look at the idea of running a brain emulator. Not that I think, as I explained in some detail in there, I don't think that the mind-brain software program/hardware platform is the correct analogy; but I don't think that that's relevant. You can still look at it as the idea of running a brain emulator at some level, and then running the brain, or the brain-mind-whatever-it-is as a program on top of that. If you can do that, why can you do that? You can do that because, the mind, whatever it is, is an information structure, an information system.

"Oh, well, that's interesting." So that means that if cryonics can work, because the information-theoretic death concept is stopped by freezing, and nanotechnology enables you to reverse what loss is there (or some other technology; it doesn't have to be nanotechnology), and if I think that a human can be uploaded onto a computer system, or cross-loaded, or whatever you want to call it, into another biological system or into a physical system of some other kind, then I have to say that what makes up the human is nothing material at all. The organization of a material system is not itself material. That's interesting.

Then I started thinking, okay, let's look at it from a slightly different perspective; let's look at some of the other directions I've gone in the "Age of Magic" talk. Let's look at telepresence and virtual reality concepts—the idea of this sort of cyberspace world, that we're sort of vaguely in when we're on Internet, and we're a little bit more in when we're working in things like the American Information Exchange, where you're doing something that's physical there. You're really in it when you're running a teleoperator, for example. What's going on there? You're in a world in some sense that, if I combine that with the idea of uploading, you can actually move into that world; that world is in some sense separate from this one. That is the first place, in a materialist sort of sense, that I had come upon the idea that there could be two worlds coexisting, and you couldn't point the direction from one to the other. They do not exist in physical relationship to each other. Cyberspace isn't *that* way, it isn't *that* way; it's different in kind, and yet it's real.

Although you can *point to this physical box, which is running the program.*

Sure, but that program may not be running in a particular physical box. It may be running bits and pieces scattered all over the place. Yes, you can point to all of them, but by the time you've pointed to everything, you haven't pointed to anything.

It's coexisting in the same space, but it's not part of the same world, in the sense that looking at the box that's running the program, you don't know anything about the program that it's running.

And, to a certain extent, it really is wrong to say that the cyberspace world that I'm in is inside my computer, because when I go inside my computer, I don't see it. It really isn't inside my computer. The computer is sort of the way it's manifested—

A gateway to an information space.

Right. And in some very odd mathematical way, its relationship to our world is very much a transform rather than a direction. It's a very strange sort of thing. That struck me as particularly interesting. Then it struck me that—and this is where I went into my talk—if I am manipulating objects in that world, I'm changing objects in this world, in that I'm moving electrons around in transistors, but I'm not aware that I'm moving electrons around in transistors; I'm only moving objects around in this world. So I came up with the idea that what you have here is two different worlds, one of which influences the other and vice versa, but not in any particularly obvious way.

For example, if we're both in cyberspace, I talk to you. How do I talk? I actuate my vocal cords, it creates vibrations in the air and on your ears and you hear them, but of course that's not what happens at all. What happens is, I, sitting in my office, talk into a microphone, which produces electrical impulses which you perceive, in your non-existent ears, as having come through the air, which isn't there either. What I perceive, or what I *can* perceive, is that we're interacting just as we would interact in here. Okay, now assume we've both uploaded. Now, we really think we're just sitting here talking, but we're still just—"just"—electrons in transistors somewhere. But of course we aren't, any more than we are "just" atoms that make up our bodies now, but still we are, in some sense, those impulses going around in a computer somewhere; maybe in two different computers thousands of miles apart.

If that's the case, then how do I talk, as an entity in cyberspace, to somebody outside of cyberspace? Well, there's got to be some sort of transducer between the two. It can be as simple as, they have a microphone which converts into electrical impulses which goes into the transistors which I hear in my nonexistent artificial ears, and vice versa—I say things that create patterns in the transistors that they read out that become electrical impulses that then go to a speaker, and they hear. So you can talk back and forth—they can talk to me and you can't hear it, but I can talk to you back and forth and they can't hear it. This becomes very interesting. I can make actuators that work. I can see through a window here, in cyberspace, and look into the external world; they can put a window there and look into this world; we can look at each other through this window—and yet there's no window. There's nothing that I can knock on and,

if I could reach around somehow, be outside of.

So we have two worlds that can interact, in which the way in which they interact is non-physical from their own perspective when they interact with each other. You and I interact by moving molecules around—there aren't any real molecules but we don't know that, and there might as well be—we have no way of knowing there aren't molecules; we don't have the senses for sensing the transistors that we are. The relationship that I have to the person outside seems to parallel—now I began to see the Christian parallels in all of this.

I can pray; I pray by speaking but the communication mechanism is not through the air. I have an analogue in cyberspace. I'm *not* saying this is how it works, I'm saying I have an analogue here that seems to mimic the same sort of phenomenon. By communicating in one realm, I'm actually communicating in another realm undetectably.

In the sense that, if you have an observer who's outside the electronic system, outside the computer—if you speak into the "air" within cyberspace, they can be monitoring it.

Correct. And no one inside cyberspace could detect that, if they didn't want to be detected. In fact, no one inside cyberspace could detect anybody outside at all unless they wanted to be detected. One can at least imagine a cyberspace so constructed that that's the case.

What happens now if, while I'm being run here, a backup program is running, copying everything I am at all times, and I die. Well, of course, I don't die at all. The backup program is still running. All that it does is sever the telephone connection. If I'm sitting here with goggles and a microphone, or I've been uploaded, and I'm talking over wires, here, but my presence nonetheless feels like it's here, and the wires are cut, you think I'm dead. I'm not dead; the telephone line's down. Hmmm. That looks an awful lot like the standard Christian view of the soul. "To be absent from the body is to be present with the Lord." "We shall not all sleep, we shall all be changed, in the twinkling of an eye." All of those things, all of the scriptural view, suddenly seemed to have cyberspace, physical, information- theoretic analogues—not in the sense that I thought that that's what was really going on, exactly, but that, it's again a case of "Gee, this makes sense," but it makes sense in a perspective that, certainly no one 2000 years ago was thinking of, and in fact almost no one is thinking of today, that they're seeing the way it's working. We are working towards a world that looks increasingly like the world I've always known from another perspective. "It's eerie," is sort of the way I would describe it. Not disturbing, exactly, but eerie.

Are you familiar with the phrase "immanentizing the Eschaton."[4]

Yes. [laughs] Making real the end of the world, or Last Things. I'm also reminded of

Windows to the Soul

by Dean Tribble

"Where do the Windows go when my Computer Dies?"

I firmly believe in the soul, the mind, the self, in the same way that I firmly believe in the windows on my screen. They don't exist physically, but they certainly exist as abstractions of emergent processes embedded in the physical universe.

The windows on my screen certainly exist: I move them, open them, close them, describe them to people. At different levels of reality, they are patterns of data in memory, patterns of glowing phosphors, structured records in C, or ideas in my mind about how to change my screen; yet they exist.

Likewise, souls exist: a pattern of reactions, cares, emotions, ideas, thoughts, intentions, memories, personality, bad puns, intellectual heuristics, and so on. To me, a soul is the gestalt of a person, it is what I care for. Many people intuitively accept this as well. It is real, but just like windows, doesn't exist separate from the complex elements out of which it arises — at one level, all the ideas, emotions, etc.; at another all the computational processes currently embedded in neurons. When you turn off the computer, the medium in which the windows exists (the underlying computation) terminates, and takes the windows with them. When you turn off the brain, the medium in which the soul operates goes away, taking the soul with it.

IMPORTANT DIFFERENCE: the window is represented in transient memory; when the computer turns off, the memory is erased, along with the representation of the process that we called a window. When the brain is turned off, the interactive process is terminated, but the memory for it, the representation, remains in the patterns of neurons. The brain represents most information persistently! This suggests that technology could be developed to preserve the patterns of intelligence — the soul, mind,etc. — and to restart the process out of which our perception of the soul emerges. Oh that's what cryonics is for!

Many people find such a view of reality sterile, devoid of warmth, devoid of magic. I find the emergence of life, love, intelligence, and all the magic of humanity from such pedestrian components to be a tremendous source of awe and beauty. I find it far more inspiring that such wondrous things can be realized from the world at our fingertips than the cop-out mumbo-jumbo of religions in which such patterns are external to the world. Why, you can find the same sense of wonder by looking at the patterns of veins in plants, or pictures generated from fractals, or patterns of commerce that create airplanes that let us fly! It inspires me to create some of these rich creative patterns myself: AMIX, or Xanadu, or *Extropy*, or nanotech, or any of the things I might build to help the world be a little bit more wonderful of a place.

Why don't people get it? Our intuitions about reality start with a naive view that only physical things really exist. We add to that the assumption that our mind/self/consciousness really exists. It's the only obvious exception to the rule that only physical things exist, so it must be dramatically different and magical. With the advent of computers, we can observe and use lots of abstractions that clearly exist, so we can start to see a smooth spectrum from physical objects (or rather our perception of them at the macroscopic level!) and things that we consider real but can only experience internally.

With the advent of programming, and the direct exposure to manipulation of abstractions and the creation of processes with a static meta-level representation, more people are figuring out for themselves the answers to some of these questions that some philosophers still get themselves confused about (not to belittle philosophers). I hope those same people also realize the deeper sense of wonder that those answers can inspire.

Where do the Windows go when my Computer Dies? Same place as my soul: back to the abstraction closet out of which we manufactured it. While they live, though, the beauty remains in the soul of the beholder, the gentleness remains in the lover, the depth of understanding grows in the people who care about the future.

Cultivate the sense of wonder.

the scripture "A thousand years with the Lord is as a day, and a day is a thousand years," for the difference in time rates that take place between cyberspace and the physical world. Which is true. Today, if you uploaded me onto a computer, it would seem to take years to do anything, to a person in the physical situation, it would seem that I was taking years, but give that ten or fifteen years and it'll be the other direction. It all depends on how fast the processor speed is running.

Finally, I looked at it and I said, well, how much would it take to emulate the entire universe? If the entire universe was a program running, how much would it take? I looked at it, and I started thinking it through, and I came to the conclusion that if you were really clever, it would take exactly as big as the universe is, so you would take the universe to emulate the universe. I've been reading *The Anthropic Cosmological Principle*[5], and it comes to the conclusion that the entire universe is a program running, not on a real computer but on a hypothetical one. Since that takes about 600 pages to come to that conclusion, I won't propose to go into it, but it also mentions and discusses in some detail an idea that, again, I don't support, but that is curious, from Teilhard de Chardin[6], who had the bizarre idea that, if you have a finite world, and you have humanity evolving and changing and multiplying and eventually filling the world, and improving and building and so forth, that eventually, at the end of time, or at some time in the future, you have this sort of perfect humanity that is in complete control of its environment.

Among other parts of the complete control of the environment is the ability to work backwards in time and grab everybody from back there. This sort of Omega Point, he calls it, where everything sort of converges, is also an Anthropic Cosmological Principle, except that it's the entire universe instead of the Earth, but so what? (It just takes longer.) The idea, to de Chardin, that that Omega Point, that convergence of all of humanity in this sort of infinitely powerful entity, looks a lot like, particularly when viewed backward in time, looks a lot like God. I don't subscribe to that, but I find it interesting that de Chardin, from essentially Christian principles, worked to this point of view that Tipler and Barrow, in *The Anthropic Cosmological Principle*, from purely materialist and computer science ideas, worked forward to roughly the same conclusions, looking on these things and saying "There's another parallel."

There's parallel after parallel after parallel, and since, as a scientist I see the one, the sort of Extropian growth and progress, as real; and as a Christian, I see the other one as real; and I see them converging in many ways—or let's put it, the possibility of convergence—I see it there. I don't see convergence between Extropianism and strict materialism, because it throws out the information-theoretic content. Materialism—standard materialism—sort of goes away, and this sort of information-plus-materialism universe, and the spiritual universe that I'm familiar with from that point of view, start looking more and more like images of the same thing, or things that are very similar in some important ways. Then again, that's a long way of getting around to the final support for why I am a Christian is, again, the consistency. It fits. It fits with so many different things, so many things that, from the beginning, I had no business to figure would fit, particularly Extropianism. I realize that most Extropians wouldn't agree with me on that, but I think I can make arguments to show where there's a lot of overlap and a lot of fit.

Have you read Dean Tribble's essay about "Where do the windows go when I shut off my workstation?" What's your reaction to that?

Yes. I thought it was a very, very bright point of view. It hasn't taken the position to its logical conclusion, but I think he's correct. Now, do I think we're being backed up on a great Hard Disk in the Sky? Well, perhaps—that's a trivial way of looking at what's going on. I don't tend to think it's a backup; I think it's the real copy—

And you're telepresent at the moment.

Insofar as the concept means anything. Being telepresent, even in cyberspace, doesn't really mean anything; there's no "tele-". "Tele-" implies distance, but there's no real distance.

When you're interacting with a simulation, rather than with a robot operating elsewhere.

The vocabulary fails, but still, in that sense, yeah. To put together my talk for Eris this past year, one of the things I did was to kick up the anthill on the Extropians list, with my reply to some of these comments about fundamentalist memes. One of the things that came out of that was Dean sending me his "Where do the windows go?" among many other odd things that I borrowed from in my talk.

What's your definition of a sin?

[Long pause.] That's interesting. [Long pause.] A sin is that which does harm intentionally to an entity that's capable of sensing, not necessarily meaning a sentient entity. You can sin against a dog; to torture a dog is sin. So that which intentionally inflicts harm—no, that which intentionally causes harm; the intent may not be to inflict harm. That which causes harm to a sentient or semi-sentient entity is a sin. The entity may be yourself; in fact, it usually is. It can't be a sin if it does no harm; it's hard to be a sin if it's unintended—the action is unintended, not the consequence.

There are no sins against God, exactly. Clearly, I believe that we do things that God does not approve of and doesn't want us to do, but it's not because we hurt God. It is impossible to harm God; it's possible to hurt God, not to harm God. You can hurt my feelings; you can't harm them. But the damage isn't done to God; the damage is done to yourself. The ultimate damage done to yourself, of course, is Hell. That's precisely what Hell is, is perpetually self-inflicted harm.

Could you elaborate on that?

Gayle: Living in Los Angeles.

[Laughs.] Intentionally living in Los Angeles is Hell. Living in Los Angeles and knowing you put yourself there.

Because I believe that we are immortal, and because I believe that every action that we take — "is recorded" is the wrong word—has impact on the backup copy that isn't the backup copy at all, but is the primary entity, our souls—everything we do, everything we intend, in particular, and do, has an impact on that. That it is possible to be such that you get used to willing evil, wrong, sin, harm, until you are not capable of willing anything else. At that point, because you are eternal, you are locked in your own mind and your own self, forever, and if there is no one left to harm, you will harm yourself. And that is Hell. It is separation from God; it is separation from everything except yourself. That's not that I think that selves are necessarily evil, but self left alone, to feed only on itself, will degenerate into monstrousness, and that is what I believe, at least, is an element of what Hell is. And it's always self-inflicted.

The idea of there being a primal copy elsewhere, of which the apparent original is merely a local representation, is somewhat reminiscent of the shadows in Plato's cave; that notion that there's a Platonic ideal. So would you consider yourself to be a Platonist?

I remember I took a course called "Medieval Christian and Renaissance Philosophy" at Stanford, and I was making some argument, but I forget quite what it was. The professor looked at me and said, "You're a Platonist! You're a raging Platonist!"

And you unapologetically said, "Why, yes!"

Not quite. I wrote a paper for that course that described the following game performed between an Aristotelian, a Platonist, and a Cartesian. They're sitting around a table, and on the table is an artichoke. The idea is that each one of them take a leaf of the artichoke and eat it; and the winner of the contest is the one that consumes the essence of the artichoke. Of course, the Platonist would say you can't

because the essence is elsewhere; the Aristotelian and the Cartesian say the essence is in the artichoke—the Aristotelian, in some sort of actual sense; the Cartesian in some sort of categorical sense—but the erroneousness of both of those points of view is that you're making gradual transitions and finally you've got something else than what you started with, so there can't have been an essence inside it. The Platonist has no trouble with that philosophy: you've changed the manifestation; you haven't done anything to the essence. In that sense, I suppose you could say I'm a Platonist, and I certainly have sympathies running in Platonistic directions.

I don't believe, as Plato does, that there exists some kind of Platonic ideal of a chair. I think a chair is a convenient name we give to things that are sort of alike. The world consists not so much of categories—Platonism is sort of the extrapolation of that to some ideal—as imperfectly replicating patterns, that it is convenient for us to group them together. Wittgenstein made the mistake, in one direction, in which they mistake the fact that we're assigning categories—names—to things, that therefore that's all that's going on, it's semantic; categorization is purely internal to us. Well, that isn't true. The imperfectly replicating patterns actually do exist in the world, and they really are similar in some effective sense.

On the other hand, the Platonists go too far, in saying you can draw absolute boundaries around things. Things don't work that way; there are such things as clines, things that shade from one thing into another, and at no point can you say, "This is one thing; this is another." A person growing from boyhood to manhood shades; you can point at one and say, "This is a boy"; you can point at another and say, "This is a man," but there's no point in between where you can say where one leaves off. The Platonist is wrong in that sense. There exist things that don't fit. We see that problem all the time in species. Speciation is a sort of Platonic idea; you have "This is this and not that," but that isn't what's really there. What's there is an infinite cline, of which we've lost most of the pieces, so we call them species; but we also have places like oak trees in California that are clines from one end of the state to the other. They can't reproduce from either end; but anywhere along the line, they can reproduce.

Have you read *Bionomics*, by Michael Rothschild[7]?

No.

He talks about clines in that sense. It's actually a good book because it tells you all about Austrian economics without indicating any prior knowledge of being connected; he seems in a sense to have independently arrived at the lessons of Austrian economics through study of biology.

Interesting.

Teilhard de Chardin....had the bizarre idea that, if you have a finite world, and you have humanity evolving and changing and multiplying and eventually filling the world, and improving and building and so forth, that eventually, at the end of time, or at some time in the future, you have this sort of perfect humanity that is in complete control of its environment.

The book was reviewed for Extropy. *He does things like, shows you the profit-and-loss statement for a beehive.*

Gayle: Of course, you know that Hayek's basis was, was in fact biology? When I was reading Hayek, again and again the biological nature of things kept hitting me and hitting me. When Phil and I were in Freiburg and visited him, I asked him if there was in fact a biological basis for his work, and he said that both of his parents were biologists, and he used to spend immense amounts of time out with them on field studies. He would sketch, study communities. So there is, in fact, a biological basis. Phil and I ran into Rothschild, when the book first came out, at one of Esther Dyson's conferences, and he was in fact not, at that time, really aware of the existence of the Austrian school.

You said that in your undergraduacy you took things that had previously been merely felt, and intellectualized them. You did that with your liberarianism. At what point did you do that with your economics?

I think that was probably almost entirely through Phil. I had always been free-market, not because I thought that was a sensible economic system, but because I thought it was a sensible political one. I had no idea whether or not it worked as an economic system; the only economics I had ever had in formal classes was standard Keynesian nonsense. It simply made little if any sense, and as a result, it sort of went by the boards.

I met Gayle, and then through Gayle I met Phil, and we had [clears throat] *occasional* discussions [chuckles].

Gayle: Brief.

Occasional, brief discussions. Sometimes for most of the night. That was really my intellectual introduction from the economic point of view, and understanding what free political systems led to—If you pursued that, you would see free consequences of actions; if you had free consequences of actions, what develops, how do people exchange goods, how do people develop things? You wind up viewing things from a Hayekian point of view more generally. I would say that was probably last of the pieces to fit in.

What does "the Singularity" mean to you? Everyone has their own idiosyncratic definition, I've found.

In stuff that I've written, I've referred to it as "the wall," rather than as the Singularity, because a singularity is something that information goes into and doesn't come out of, and I don't happen to believe that that's exactly what's going on. I think it's more like a barrier we can't see through. The other problem I have is that the Singularity is generally modeled as the natural consequence of exponential growth, and I don't happen to believe that's true. I think that, in general, we can understand new things at about the proportion to what we already understand, and that is the formula for exponential growth, and that we can maintain almost indefinitely.

Nonetheless, I believe that in the relatively short term we will have something that you can call the Singularity or the wall take place, but it will be as a result of super-exponential, not of exponential growth. Intelligent machines, for example, is a place where super-exponential growth can occur.

How is that analysis affected by the fact that when you say "we," humanity collectively is increasing their understanding with exponential growth. Do you think that individuals can expand their individual intelligence, leaving aside future intelligence increase technologies? The problem that I perceive as leading to a wall or Singularity is that society collectively is increasing its knowledge at a rate that is much faster than an individual can keep up with.

I don't think that's true. First of all, I think there are evolutionary reasons why that isn't true. It may be instantaneously true at any particular time, but it's not a sustainable phenomenon. Nor do I happen to think it's taking

place now. I don't think that the average person is any more out of sync with the advances of technology today than they were 100 years ago. Roughly the same percentage of people at roughly the same average age group can't handle advances in technology, and for roughly the same reasons, and that is inflexibility of mindset, or whatever.

Actually, I tend to see the opposite phenomenon taking place, and this may be a phenomenon just of our circle, but I see a lot of people whose minds I don't believe are going to "gel" and turn into concrete, and reach a point where they cannot handle the changes in technology. So actually it seems to say that, because we expect rapid technological change, we can cope with it. It's almost the exact opposite of future shock. The most interesting thing about Toffler's book *Future Shock*[8] is that it was wrong. There's no such thing as future shock. The reason there's no such thing as future shock is, I think, what's wrong with the standard interpretation of the Singularity, is that it's the product of exponential growth. The error it makes is that, well, if the amount of information that I have to store in my head goes up exponentially, and my head is finite, I'm going to reach a capacity. But it's not the amount of information you have to store, it's merely the amount of information you have to deal with. It doesn't matter how much water the firehose is putting out if you can control the firehose; and we are, at about an exponential rate, capable of controlling the flow of information, and that's a steady-state phenomenon; you can continue doing that indefinitely.

However, I believe there are things about to occur, or in the process of occurring already, that make the discussion moot, because I don't think it is going to be exponential growth; I think we're going through a super-exponential growth period in the relatively short term. There are lots of ways this can occur; the one I think is most likely to occur is if you look at the fact that the processing power of computers is growing exponentially, you assume that it goes on at exactly the same exponential rate of growth that it's gone on since electromechanical computers in the '40's, and you wind up sometime around 2025, another thirty-some-odd years from now, they go through a point of having human intelligence—a human brainpower processing equivalent, which is not the same thing as intelligence, because you have to have the growth in algorithms, too. In fact, as I think I said in my *Extropy* paper, an intelligent system can run on a computer made out of Tinkertoys; it just runs very slowly.

Nonetheless, the same sort of exponential growth *is* taking place in machine intelligence, the algorithmic part; it's just lower on the curve. At some point, that's going to reach a human level of intelligence, too. If you then have those human-level intelligent programs working toward improving the speed of computers, that's a formula for super-exponential growth. That's one of the many, many ways in which the wall or the Singularity can be reached in some finite amount of time.

Some people differ on the definition of the wall or the Singularity in saying that it's a singularity because it cannot be reached, in the sense that their picture of the Singularity is a horizon beyond which we can't see, but which we'll never reach. It's receding in front of us, and we're getting nearer and nearer to it asymptotically.

The... Singularity is generally modeled as the natural consequence of exponential growth... I believe that in the relatively short term we will have something that you can call the Singularity or the wall take place, but it will be as a result of super-exponential, not of exponential growth. Intelligent machines, for example, is a place where super-exponential growth can occur.

No. I don't believe that that's right. That's not what I mean by it, and I agree that people do that; Vinge seems to have some vision of it being that way too. Yet, clearly, they reached it, in his book[9], though we never find out what it is. It's a cheat; we never find out what it is. We sort of get a guess at it.

Nor do I believe that it's something that all of humanity's going to go through one bright day in April. I don't think it's any of those things. I think that what it is, is a period in which, in order for humans to continue to exist at all, their very nature must change radically. There are many reasons why that must come about, but the sort of global reason why I believe that will come about is that our creations—our machines, things we don't even think of as intelligent machines—will get to the point of rivaling us, at about the same point we have the ability to modify ourselves to stop that from taking place. Not in any adversarial way; we'll just form alliances with them, but by forming alliances with machines, by using them to enhance our intelligence, our memory, our capability, our senses, we will become something different from what we are.

At the point where humans modify themselves, that's another form of super-exponential growth, and at that point we go through something you might as well call the wall, because you can't predict what's beyond it. I call it the wall because we can't see across it—not because it's retreating; not because you can't get there, because I believe you can get there, and that we'll get there in a finite amount of time, but because you simply cannot predict what's beyond it. For a lot of reasons, not the least of which is, because humans will change—normally you can predict things either because you can extrapolate current trends, or because you can know where you want to be and assume you can do everything. The problem is, in this particular case it's almost impossible to say where "they" will want to be, when "they" can modify what they want as freely as they can achieve what they want. That's what makes it a super-exponential change, and impossible to see much beyond.

Other descriptions of the wall or the Singularity have been in terms of a phase change; that sounds very compatible with what you're saying.

It's much more of a phase change than it's a singularity, in my view. In a lot of ways. If you look at the so-called inflationary universe theory, what happened with the universe was that our region of natural law sort of crystallized out in a sort of a phase change that proceeded almost infinitely rapidly. Literally, a change of physical law. If we're living half in cyberspace, physical law is what we make it. Things in the physical world are so constrained, but less and less of us is in the physical world—less and less of each individual is in the physical world. We have shared consensual realities that have nothing to do with standard physical law. In reality, physical law crystallizes out in a very different way; there's a phase change in many different dimensions of how you look at it.

When you look at the entities that we or our descendants will become, these vaster machine entities, do the questions of identity trouble you?

They don't trouble me; they certainly fascinate me. From such simple questions as: If you have two indistinguishable copies, and neither knows which is the original—and in fact I believe there are ways in which you can make copies in which the question "Which is the original?" is meaningless—then what happens to property rights? I agree that the question is somewhat ameliorated by the fact that information is one of those properties that's duplicable without loss, and so many things are solved at the same moment that they come about, but many things are not. If you have a patent, for example—we can argue patent law—but if you have a patent, who has it? You've just divided its value in half, whatever its value was. So there's lots of

questions from a simple case like that.

We have the ability to undergo Lamarckian evolution-driven intentionally-driven evolution for the first time. We have the ability to form children of many parents, and they needn't be children, they may be bits of ourselves. We have the ability, perhaps, to partition and recombine our consciousness. In those senses, the concept of identity does become mutable. Does it become disturbing? People will do things they oughtn't to do, because people always do things that they oughtn't to do; people will do things that are disastrous, because they always do things that are disastrous; and people will do things that are wonderful and beautiful, because people always do things that are wonderful and beautiful, as well. Since I think we will do these things rather than get supplanted by our machines, among other reasons, the alternative seems worse, and the potential seems truly amazing.

Because I believe that human nature is fallen, and therefore corrupt, there will always be people who do bad things, but I also believe that there are very good people, and there are people who will fight them, and the universe is so constructed, I think, that the hand is tilted it slightly in favor of the good. Because I believe that to be the case, giving more power to all people, and particularly to individuals, tilts that balance better. It's when power is centralized—if my ability to hurt you is amplified by a government, or is amplified by my restrictions that I place on you, then if I'm evil, then I'm likely to do more harm than if I have less power relative to you. That's one of the many reasons that I'd like to see that these things take place—the movement to cyberspace, the increasing computing power in each individual's desk—because I've always believed, and I continue to believe, that free minds are always going to be able to maneuver faster than governments can.

We've talked about uploading; do you see any reason why individual consciousnesses can't be, first, duplicated, and second, backup copied?

The answer to the question is harder than I think a lot of people understand. The short answer is, no, I don't see any reason why not, in both cases. I believe that a human consciousness can be uploaded. I think that it is very difficult to make sure that what you are making is a transfer rather than a copy. It's very hard to do, and I go into that a lot in my *Extropy* paper[10]: not only *why* I think that's hard, but also how to get around it. But hard is not impossible, and I think it is doable. The key, in my view, seems to be gradual replacement, and replacement at a low enough level that whatever makes up the consciousness is not operating at that level. It's basically the emulation idea.

So copying is easy, transfer is hard.

Yes, copying is easy, transfer's hard, and if you want to do the transfer, you have to do the copying at a low enough level, and you must be destroying the original as you go, and the parts must always remain interconnected while you're doing it, in some real sense, and so on.

It's when power is centralized—if my ability to hurt you is amplified by a government..., then if I'm evil, then I'm likely to do more harm than if I have less power relative to you. That's one of the many reasons that I'd like to see... the movement to cyberspace, the increasing computing power in each individual's desk—because I've always believed, and I continue to believe, that free minds are always going to be able to maneuver faster than governments can.

That may not be required, and it certainly isn't provable afterwards, but for my comfort level—if that's what's happening; if it's *me* we're talking about—that's what I want to do.

It's more aesthetically pleasing to have the dissolution of the original be gradual during the process, rather than, "Okay, that's *now you over in the machine, and now we're going to take this biological body and get rid of it."*

Right, no thank you, no, un-uh, that's not acceptable. This is the problem with backup copying. If it's just a copy, that's of almost no interest to me.

If there's an original elsewhere—

That's right; it's not me. It may be like me in many many ways; it may be identical to me in most ways, but it isn't me. However, I believe it is possible to use the same transference process, because, after all, the transference process says nothing about how long it takes; the speed of the process has nothing to do with it; it's the method of the process that matters, and in that sense, you may be able to have a link between two things, the copy and the original, that is intimate enough, that when the so-called original ceases to exist, the copy continues and there's no break.

The telepresence is merely switched from one—
Yes, and again, here's where it looks a lot like the soul and the body, to me; and in fact you could operate it in some sort of way in which the copy is the real piece and this is the telepresence, as sort of a crude way of looking at it. That's a possibility. But I believe that those are both possible things to do, though a good deal harder than the process of copying. After all, we know how to copy things much earlier than we know how to duplicate them. I mean, you can copy a chip, you go to Korea, or Singapore, or these days Thailand, with a chip, and they strip it off layer by layer and produce masks and copy the chip, but they don't know how to build them in the first place. Copying is always easier; you just do the same thing again. The problem is to make the first item, or doing the variant of it, or whatever, and there the transferences are harder because it's a harder problem, but even it's not the same as being able to create from whole cloth. That's the hard part. They're all hard, but that's the hardest part.

We were talking earlier about the fact that people in our community seem to be seeking novelty in doing something different every couple of years.

Mm-hmm. With consistency to it, though. The things tend to tie in and tie together. Yeah, that's true; we do tend to that. I'm not sure it's exactly seeking novelty; I think it's more seeking interesting things. Novelty is part of it, but depth of knowledge is also interesting, which is a novelty of all. I think we tend to play with things with far more effort than most people work at things, as a group.

Play at things with such an intensity as to be able to make a living from them.

That's exactly right. So much the better, in my opinion.

It has never occurred to us that there is—Somebody made the comment the other day that your life is divided into two pieces: Your life is divided into what you do for a living, and what you do that interests you. That concept is utterly alien to most of the people that are my close friends—the idea that what you do for a living and what you do that's interesting are separate. Why should they be separate? You spend eight hours a day doing your work; oughtn't you to enjoy it? If you spend ten or fifteen hours a day doing your work, you really ought to enjoy it. But not because it's your work, but because it's your play. That doesn't make it trivial; that just makes it fun. It doesn't always make it fun, either, but it works out better that way, I think. That's what's almost unique about this group of people—the

sense of play. Very serious play.

Gayle: Do you intend to grow up?

No. Why should I? One of the moments of understanding that comes upon everyone several times, if they're very lucky, during their lives, came upon me watching Mary Martin playing Peter Pan, as a child, singing "I won't grow up," and I took this to heart. [chuckles] I remember the wonderful line from C. S. Lewis that says that the most mature adults are childlike and most childish adults are the ones that try to act the most mature. It's absolutely true. My father worked 40-some-odd years for the same company. I am in almost no danger of ever doing that.

Why don't you talk a bit about what you're doing now, and how it relates to your long-term goals?

In a personal letter some weeks after this interview, Dave added:

"... Please don't confuse the (admittedly unusual and extropian) arguments I make in "Seven Paths to Immortality" with my beliefs. I consider my belief system to be orthodox Christian and I call myself a fundamentalist. This does not mean that I subscribe to all the (peripheral) beliefs of all those who also call themselves fundamentalists — that would be impossible because they are not completely consistent within themselves. Some, for example, take the Genesis creation account as literal while others (both me and St. Augustine, for example) do not. What I mean by "fundamentalist" is what most who call themselves that mean: I subscribe to the fundamental beliefs of the Christian religion. These include the belief that the universe is created by God, Who takes an active interest in its current existence; that Jesus of Nazareth is God incarnate; that He was crucified and rose from the dead, and that because of that act all people have open to them freedom from the evil in their lives and reconciliation with God. I believe the Bible is God's word to mankind for the purpose of salvation; I do not believe it is either a biology textbook or a cookbook."

What I do for a living is start companies and try to bow out soon enough that they don't get tiresome. Right now I'm working on what I would call, at a sort of meta-level, machine intelligence or algorithmic intelligence, although that's a little bit too grandiose. Ten years ago I founded Calera, to do optical character recognition, and that was sort of the first area that I got into that was doing something that looked like artificial intelligence or machine intelligence. Basically, how do we teach a system to recognize a pattern that appears to us to be similar to, but not identical with, another pattern it has seen? How do we convey that information to a program; to be able to say: "Yes, you haven't seen this before, but you've seen something like it. Categorize it the same. Probabilistically categorize it like other things you've seen that are like it." That whole sort of meta-problem is one I've been working on now for quite a while. It has lots of different manifestations, but the character recognition is the more important one. I left Calera; I worked on the reading machines for the blind; I've used some of the same techniques that have been used for character recognition to make a text-to-speech system that works just about as well as DecTalk but has no rules at all.

I've developed the three paradigms of artificial intelligence; this is sort of my idea. There's the expert system one, which I call the totalitarian paradigm, which is, "Here are the rules, don't bother me with the details or the facts. I'm going to impose the rules from the top down, and we will live according to them." This works every bit as well as central planning in an economy does. This is of course why everything we have works on an expert system, right? This is the wonderful comment which I think the head of AI at Yale said, that if computers worked as well as the expert systems branch of artificial intelligence, had made as much progress in the same thirty years, they'd still be made with vacuum tubes.

The second one is the sort of neural network approach, which I call the libertarian paradigm of artificial intelligence, which is, "I don't know how this thing works, but if I throw enough things together and stir vigorously enough, something useful will come out of it." This is the bottom-up approach. The problem with the totalitarian one is worse: that it is impossible to know all the rules operating in anything other than trivial systems. There may not be rules; there may be statistical, probabilistic things working. The world is not, as I've said before, categories; it's imperfectly replicating patterns that replicate with some degree of probability. The problem with that is, unless you have in infinite number of rules, at which point it ceases to be tractable. If you have an infinite number of probabilistically applied rules, you haven't got rules at all. The problem with the libertarian paradigm is the fact that, although the basic approach is right, there is global information; even though you don't know all the rules, you do know some global information, and it fails to apply it.

So I came up with what used to be called, when I was a Republican, the Republican paradigm—and I'm currently looking for a better name for it—which is a combination of the two. An example of that would be, the obvious one, from character recognition: I have a series of images, I know that these images are likely to be *e*'s; they may be an *a*; and I don't know, though I know probabilistically what they are; but I can go to a contextual system, in which I say, "*This* is a word that exists in English, and *this* is not," and that's an absolute, imposed rule. Therefore, I will choose the *e* over the *a*, regardless of their relative probabilities. That's imposing a rule, but it's a very sensible rule to impose.

The systems that I build work that way, and they've been applied to speech synthesis, to coin grading, to character recognition, to all sorts of different things. I'm looking at doing it in speech recognition. Right now, I have a new company, called RAF Technology, that does pattern recognition to order. Right now I'm working with the Post Office, reading addresses on mail pieces for automatic routing and improving the recognition rate so it can be routed automatically by machine.

And, thinking increasingly about how machines learn. My longer- term goal in that is, of course, "If we can't beat 'em, join 'em," and how do we do that. We're now making some weak steps in that direction.

NOTES:

[1]See, for example, Merkle, Ralph. "The Technical Feasibility of Cryonics." Available from the author, merkle@parc.xerox.com.

[2]See Merkle's paper "Cryonics, Cryptanalysis, and Maximum Likelihood Estimation" in the forthcoming Proceedings of EXTRO 1: The First Extropy Institute Conference on Transhumanist Thought.

[3]Merkle "The Technical Feasibility of Cryonics."

[4]Coined by Eric Vogelin. The history of the phrase is described in *The Illuminati Papers* by Robert Anton Wilson (London: Sphere, 1982).

[5]Barrow, John and Tipler, Frank. *The Anthropic Cosmological Principle.* Oxford: Oxford University Press, 1990.

[6]Teilhard de Chardin, Pierre. *The Phenomenon of Man.* New York: Harper & Row, 1959; *The Divine Milieu.* New York: Harper & Row, 1968; *The Future of Man.* New York: Harper & Row, 1974.

[7]Rothschild, Michael. *Bionomics: The Inevitability of Capitalism.* New York: Henry Holt, 1990.

[8]Toffler, Alvin. *Future Shock.* New York: Random House, 1970.

[9]Vinge, Vernor. *Marooned in Realtime.* New York: Baen Books, 1987.

[10]Ross, David J. "Persons, Programs, and Uploading Consciousness," in *Extropy* #9 (vol.4, no.1, Summer 1992).

Neurocomputing 7: Sequential Networks and their Dynamics

by Simon! D. Levy

In the last two installments of this series, I departed a bit from the neural networks theme to give readers a taste of some exciting developments in related fields. In this issue of *Extropy* I return to neural networks in order to describe a relatively new type of net that is the focus of a great deal of current research in connectionist modeling, namely, sequential (a.k.a. recurrent) neural networks.

Recall that the basic task of a neural network is to perform a *mapping* between two sets of data patterns, usually referred to as the input and output patterns. In the exclusive-OR (XOR) network that I have used as an example in this series, the patterns consist of binary digits (ones and zeros); the input patterns consist of pairs of these digits, and the output patterns of a single digit. The goal in the XOR problem is to "train" the network to produce an output of one when the members of the input pair are different from each other, and a zero when they are the same.

The most general thing that can be said about such a network is that it will always produce the same output pattern every time you present it with a given input pattern. If we present a well-trained XOR network five times with the pattern (0,1), it will spit out five 1's.

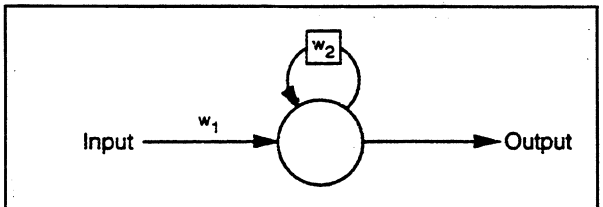


Figure 1. A one-node net with a recurrent connection

This consistency occurs because the connections in the net are unidirectional: the input nodes feed into the hidden nodes, which feed into the output nodes[1]. In other words, the network has no *feedback* mechanism to allow its state at a given time to be influenced by its state at an earlier time; it has no *state dynamics*.

By putting *recurrent*, or "backwards" connections into a neural net, we can provide this type of feedback mechanism. One nice result is that the network can now learn not just static patterns, but patterns made of *sequences*, which is why these sorts of networks are usually referred to as "sequential" or "recurrent" nets.

To get an idea of how sequential nets work, think of the simplest possible network of this type: It would consist of a single node with an input and output and one recurrent connection. This connection would feed the node's output from a previous time back into the node at the current time; in other words, the connection would provide a *delay*. Like any network connection, this connection will have a *weight*, a factor by which the value passing along the connection is multiplied before it reaches its destination. Since we are keeping things simple, we'll assume a delay of one time unit, or "tick".[2] In other words, the value fed back to the node is the node's output at the previous tick, multiplied by the weight. Such a network is illustrated in Figure 1, in which I have adopted the usual convention of indicating a delay connection by means of a small box.

As a further simplification, we assume that the mapping done by the node is the *identity* mapping; that is, the node's output is the same as its input at a given time. Finally, we assume that the output of the network is zero as an initial condition (i.e., at the zeroth tick). To run the network, we perform the following steps:

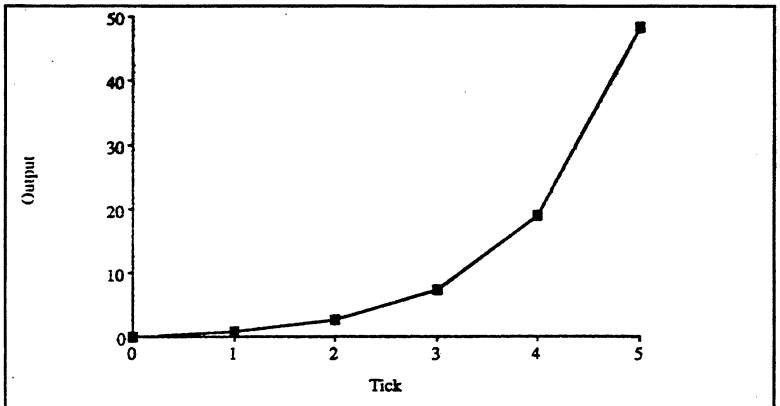


Figure 2. Output of the network in Figure 1, with Input = 1.5, w1 = 0.5, w2 = 2.5

(1) Multiply the input I by the weight w1 and record the output O = I * w1.
(2) Multiply O by the weight w2.
(3) Feed a new input I * w1 plus the result of (2) back into the net and
record the new output O.
(4) Go to (2).

We note that our little network is only *stable* – i.e., that its output doesn't fly out toward positive or negative infinity – when the value of the recurrent weight w2 is between negative one and positive one. If the weight were any greater (less) than positive (negative) one, the output value would keep getting farther and farther from zero because of the amplifying effect of w2, in a way reminiscent of what happens to the normally inaudible hiss from a speaker when you hook up a microphone to it and place the mike in front of the speaker. Figure 2 shows an example of such feedback for I = 1.5, w1 = 0.5, and w2 = 2.5:

For values of w2 between negative one and positive one, the output of the net will *converge* to a single value given by I*w1/(1-w2), as time approaches infinity. This is illustrated in Figure 3 for w2 = 0.5 and the

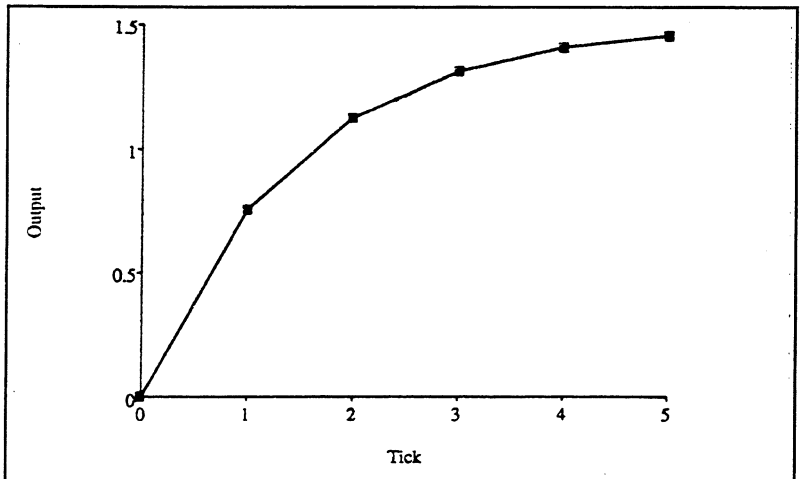


Figure 3. Output of the network in Figure 1, with Input = 1.5, w1 = 0.5, w2 = 0.5

values of I and w1 used in the figure above.[3]

Now, such a simple net isn't useful for very much in the real world (whatever that is!), but we can still get some pedagogic mileage out of it by using it to introduce the concept of *attractors*, and their less-frequently-discussed opposites, *repellors*, and the notion of a *dynamical system* in general.

First, let's simplify the net even more by eliminating its input (i.e., by setting its input weight w1 to zero) and allowing ourselves to set the value of its output – also known as its activation – at any given time. Having done this, we observe that the output still tends toward infinity when the value of the recurrent weight w2 is greater than one; but, since the input weight w1 is now equal to zero, the asymptote (convergence value) is zero when the absolute value of w2 is less than one; i.e., 0/(1-w2) = 0.

To put it in other terms, this network affords us with two distinct simple dynamical systems, "dynamical system" being the general term for any system (such as a sequential neural network) whose behavior changes over time. We have one system when the absolute value of w2 is less than one, and another system when it is greater than one.

In the first system, no matter what value we choose for the activation of our single node, the ultimate value will have to be zero. We say that zero is the "point attractor" for this system, which is shown in Figure 4a. The arrows pointing toward zero represent the trajectories taken by the system for a given starting point; that is, the sequence of node activations that will be observed after we set our initial node activation. In the figure, I have chosen the arbitrary initial activations -3 and +2 as an illustration.

The second system (in which the absolute value of the recurrent weight w2 is greater than 1) is a bit more complicated. If w2 is positive, the system will have a single point repellor at zero: If we start with an activation less than zero, we'll shoot out to negative infinity, and if we start with an activation greater than zero, we'll shoot out to positive infinity, as shown in 4b.

If, however, w2 is negative, we'll get an unstable oscillation (back-and- forth trajectory) whose limits will be positive and negative infinity, as shown in Figure 5.

So, as you can see, a maximally simple neural network can have a surprisingly rich set of dynamical behaviors, depending on the values chosen for its weights. Still, point attractors and unstable oscillations aren't of much benefit in describing the kinds of systems (biological, social, and economic) that interest most scientists. For those kinds of systems, we need a network that can give us stable oscillations.

Figure 6 illustrates one such network. Again, since we are more concerned with the dynamical properties of the network than with having it perform a mapping, we don't bother to give it any input. If we plot the activation of node X against that of node Y, we get a trajectory in the form of a circle (or an ellipse, depending on how we set the weights). If we think of X as representing the position of something and Y as representing its velocity, then the trajectory becomes a phase plot of motion in one dimension. In fact, the network in Figure 6 models the dynamics of a frictionless pendulum, which swings back and forth forever through the same path without winding down.

Of course, no one has even seen, or is ever likely to see, a frictionless pendulum, and the behavior of such a system is pretty boring, anyway: No matter where you start the thing, it always swings back to that same point (and back out again), so that the set of trajectories you get is just a bunch of concentric circles. Much more interesting is a system that has stable oscillations (like the frictionless pendulum) and an attractor (like the system in Figure 4a). In other words, we want a system whose attractor is a cycle, not a point. One such system, whose equations are given by Norton (to appear), is shown in Figure 7. Though it contains only two nodes, this system has a

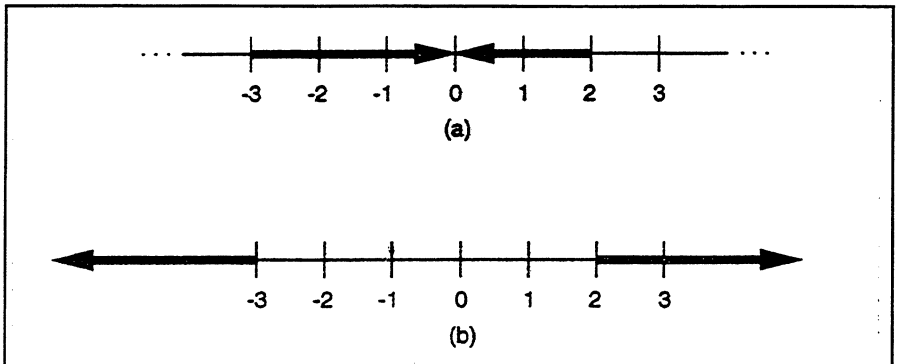


Figure 4. Sample trajectories for a system with a point attractor (a) or point repellor (b) at 0

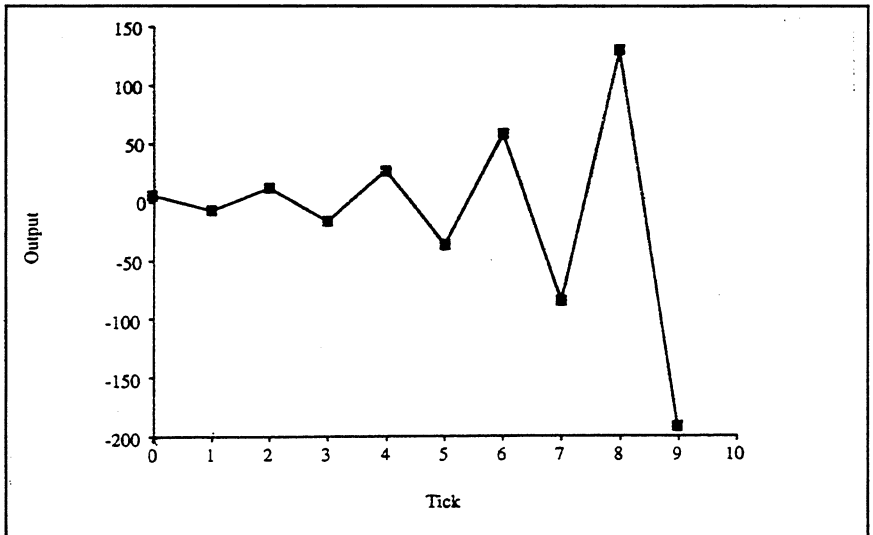


Figure 5. Output of the network in Figure 1, with no input, w2 = -1.5, initial output = 5

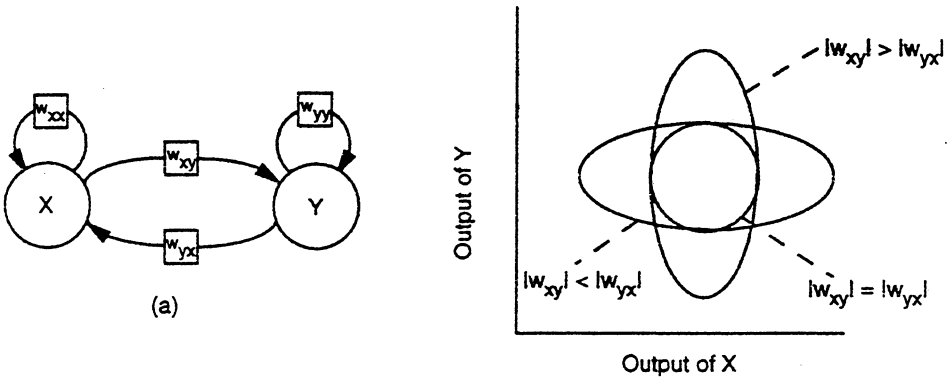


Figure 6. Network (a) and sample phase plots (b) for a stable oscillator with no attractor

complicated set of symmetrical connections to support its stable oscillatory behavior – we have come a long way from our simple single node-network! But, as you can see in the phase plot in Figure 7, the network does what it's supposed to: The circle represents the attractor, or limit cycle, into which all trajectories (curvy lines) flow, from both inside and out. No matter where we start, we end up on the cycle.

Now we have something useful! The network in Figure 7 describes a general property of oscillatory movements observed in a wide variety of human and animal activities, from interlimb coordination to speaking, viz., their resistance to perturbation: (Kelso et al 1981). If you, or your cat, are walking along the sidewalk and your gait is suddenly interrupted by some small obstacle – say, a rock that you step on – you will rapidly (and unconsciously) re-establish your walking pattern without further ado. It seems likely that our nervous systems are organized to provide these kinds of stable oscillations, in order to help us cope with unforeseen events. Partly for that reason, a number of researchers (e.g., Jordan & Rumelhart 1992) have been using recurrent nets as a way of modeling limb movement.

In general, the shift in AI research has been away from devices that use goals, plans, and an internal representation of their environment (e.g., Minsky 1975) and toward devices that use non-representational dynamical systems, like the one in Figure 7, as a way of getting around in the world (e.g., Beer, to appear; see my article "Neurocomputing 5: Artificial Life", in *Extropy* #8 for further information). Such robots are less likely to collapse in hopeless confusion when they encounter an unexpected obstacle than are robots built on more top-down principles.

Those who are familiar with neural nets will notice that I have given no attention yet to the question of how to train sequential nets; that is, how to make the nets learn the right weights to emulate a particular dynamical system. The answer to this question is that, since non-sequential nets like the XOR net are really a special case of sequential nets, but with no recurrent connections and a single time step per pattern, the algorithm for training sequential nets is a more general case of the back-propagation algorithm used for non-sequential nets (which I described in "Neurcomputing 3" in *Extropy* #6). This more general algorithm, called "Back-Propagation-In- Time" (BPIT) or "Back-Propagation-Through-Time" (BPTT), requires that, for each input/output pattern, we accumulate all the errors made by the network over time, and use the average of these accumulated errors to modify the network weights. Because this procedure necessitates the storage of the entire "history" of node activations for a given pattern, it can be costly in terms of computer memory. Nevertheless, it is possible to explore simple examples of such networks on your home computer. I have written a C program that I call BPIT ("beep-it"), based on a network architecture described by Jordan & Rumelhart (1992), that allows you to train such a network to do some rudimentary things, like moving around in a circle clockwise or counter-clockwise, depending on the input.

In fact, the ability to train a (sequential) network opens up an entirely new and interesting realm, called parameter dynamics, where the concern is not with the evolution of states (node activations) but rather with the evolution of parameters (connection weights), which have attractors and other properties all their own. Then there is the "final frontier" of graph dynamics, in which the nodes and connections themselves can evolve. Those who wish to learn more about such issues should read my paper in the Proceedings of the Extro 1 conference.

REFERENCES

Beer, R. (to appear) Computational and Dynamical Languages for Autonomous Agents. In T. vanGelder and R. Port.

T. vanGelder and R. Port, Eds., (to appear) *Mind as Motion.* Cambridge, MA: MIT Press.

Jordan, M. and D. Rumelhart (1992) Forward Models: Supervised Learning with A Distal Teacher. *Cognitive Science* 16, 307-354.

Kelso, J.A.S., K.G. Holt, P. Rubin, & P. Kugler (1981) Patterns of Interlimb Coordination Emerge from the Properties of Non-Linear, Limit Cycle Oscillatory Processes: Theory and Data. *Journal of Motor Behavior,* 13, 226-261.

Minsky, M. (1975) A Framework for Representing Knowledge, in P. Winston, ed., *The Psychology of Computer Vision.* New York: McGraw Hill. 221-279.

Norton, A. (to appear) Dynamics: An Introduction. In vanGelder and Port.

Rumelhart, D.E. and J.L. McClelland (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.

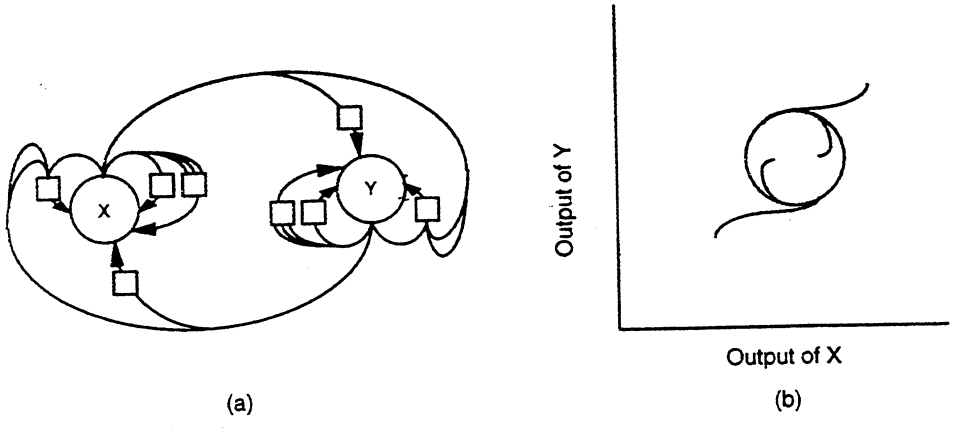


Figure 7. Network (a) and phase plot (b) for a stable oscillator with a periodic attractor. In (a), multiple lines entering a box indicate that values along the connections are multiplied together, as well as being multiplied by the weight. In (b), trajectories inside the circle spiral out; trajectories outside the circle spiral in.

Cont. on p.36

EXTRO²

The 2nd Extropy Institute Conference on Transhumanist Thought

CALL FOR PAPERS

Southern California, June 16-18 1995

PURPOSE: Extro² will be a rich, intellectually invigorating gathering designed to help push outward the boundaries of progress and possibility. It will be both a serious study and a spirited celebration of humanity's limitless potential and how it will be achieved. Besides presentations of accepted papers, the conference will feature lectures by leading thinkers, panel discussions, and the Extropy Awards banquet. The event will begin on Friday afternoon with a hike, followed by an evening reception.

Submitted papers should as much as possible exploit interdisciplinary connections, rather than presenting results in a particular narrow subfield. They should be aimed at an intelligent, educated, and interested audience that is not necessarily familiar with the detailed background of any field. Accuracy, rigor, and rationality are of course expected, but breadth of vision is also important.

Papers must be written in English, be 5,000-12,000 words in length, and should begin with an abstract of not more than 400 words. Papers must include a separate cover page containing the title, author, postal address, and email address if available. Submissions should not have been previously published or submitted to any journals or refereed conferences or workshops. Accepted papers are to be presented at the conference.

DEADLINES: Submission deadline is December 1 1994. Authors will be notified of review decisions by January 15 1995. Final versions of papers are due back by February 15 1995 for inclusion in the conference Proceedings.

Send papers to:
Extropy Institute
Extro 2 conference, 13428 Maxella Avenue, #273
Marina Del Rey, CA 90292.
310-988-0375 info@extropy.org

(See p.46 for audio tapes and the Proceedings volume for EXTRO 1.)

Neurocomputing, Continued from p.35

FURTHER READING

Abraham, R.H. & C.D. Shaw (1985) *Dynamics: The Geometry of Behavior.* Santa Cruz: Aerial Press.

A four-volume comic-book-style introduction to dynamical systems. Even with illustrations, I found it rough-going after Volume 1.

Farmer, J.D. A Rosetta Stone for Connectionism. *Physica* 42, 153-187.

The first attempt I am aware of to describe different kinds of networks (neural, autocatalytic, immune) in a single framework. Much of the paper was over my head, though readers with a solid background in the physical sciences may have an easier time.

Saltzman, E.L. & K.G. Munhall (1992) Skill Acquisition and Development: The Role of State-, Parameter-, and Graph-Dynamics. *Journal of Motor Behavior*, 24, 49-57.

A good introduction to the three different types of dynamics a system can exhibit, described within the context of motor skills. Short and easy to read.

NOTES

[1]It is possible to connect the input units directly into the output units as well, as in Rumelhart and McClelland (1986) p. 321, but the net remains unidirectional in this case too.

[2]A tick is an arbitrary discrete unit of time; it could be a year, a minute, a nanosecond, or any other unit relevant to the real-world system we are interested in modeling.

[3]Mathematically oriented readers will note the not-accidental resemblance of these figures to plots of the exponential and logarithmic functions, respectively.

Extropy Institute

The December '93 literary supplement to the *Village Voice* described Extropians as "radical humanist technophiles", and referred to "the movement's combustible mix of fringe academics, overeducated computer programmers, and renegade philosophers". A narrow description, but one that nevertheless hints at some of our concerns and enthusiasms. For those of you for whom this issue is your first real contact with Extropian ideas, the short version of The Extropian Principles to the right will help clarify our shared values and goals. (The full text appeared in *Extropy* #11.)

Extropy Institute (ExI) was incorporated in 1992 as an educational, tax-exempt organization. Like the Extropians e-mail list, ExI was an outgrowth of *Extropy* (founded 1988). We created ExI in order to provide a structure and network that would facilitate the spread and evolution of extropic ideas, values, and culture.

This organizational mission encompasses two aspects which together explain all our activities: (a) Within our existing Extropian culture refining and developing our ideas, working together to transform ourselves into transhumans and to evolve a radically new culture free of the irrationalities and limitations of the past. (b) To clearly and persuasively communicate our philosophy of life even to those who are not already attuned to the same ideas and attitudes, in order to influence the broader culture in more extropic directions.

In pursuit of these goals Extropy Institute — though yet limited by a relatively small (but rapidly growing) membership, and tightness of funds — continually seeks new outlets for its members' energies, abilities, and intellects. Our primary publication, *Extropy: The Journal of Transhumanist Thought* is supplemented by our members' newsletter, *Exponent*, edited by frequent *Extropy* contributor Simon! D. Levy. *Exponent* carries shorter articles, membership information such as forthcoming meetings, reports on progress of projects and new media attention, and reviews of relevant books, software, and other media.

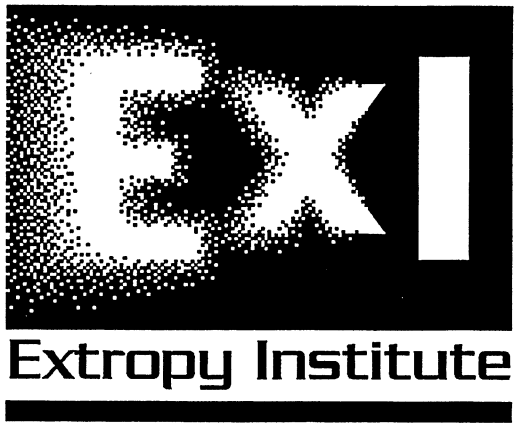
A variety of meetings take place, such as last summer's *Extropy* 5th birthday party, weekly lunch meetings in the N. California Bay Area, and monthly Idea Forum discussion meetings in the Los Angeles area. In addition, impromptu get-togethers take place all over the country. May of this year saw an important new development: EXTRO 1 heralded the start of a series of annual conferences where ideas can be explored in depth, and bounced off persons of many different specialities and perspectives. (A report on the conference can be found in the June '94 issue of *Exponent*, and the main talks from each session found in the Proceedings volume.)

Supplementing printed publications and physical meetings is the online Extropian virtual community. The

Continued on page 38

EXTROPY INSTITUTE
13428 MAXELLA AVENUE, #273
MARINA DEL REY, CA 90292
310-398-0375 INFO@EXTROPY.ORG

For details of membership rates, see p.2, lower right.



EXTROPIAN PRINCIPLES v.2.5

(Full version in *Extropy* #11)

1. Boundless Expansion — Seeking more intelligence, wisdom, and effectiveness, an unlimited lifespan, and the removal of political, cultural, biological, and psychological limits to self-actualization and self-realization. Perpetually overcoming constraints on our progress and possibilities. Expanding into the universe and advancing without end.

2. Self-Transformation — Affirming continual moral, intellectual, and physical self-improvement, through reason and critical thinking, personal responsibility, and experimentation. Seeking biological and neurological augmentation.

3. Dynamic Optimism — Fueling dynamic action with positive expectations. Adopting a rational, action-based optimism, shunning both blind faith and stagnant pessimism.

4. Intelligent Technology — Applying science and technology creatively to transcend "natural" limits imposed by our biological heritage, culture, and environment.

5. Spontaneous Order — Supporting decentralized, voluntaristic social coordination processes. Fostering tolerance, diversity, long-term thinking, personal responsibility, and individual liberty.

BACK ISSUES

#1, 2, 4, 5, 6, 7, 9, 10, 11: $5 each.
Available from Extropy Institute (address, p.2)

12, Vol.6 No.1 (1st Quarter '94): A Practical Look at Ocean Colonization, by Bill Eichman; The Last Free Place on Earth, by T.O. Morrow; Logical Languages: A Path to Posthuman Rationality? by Simon! D. Levy; The Open Society and Its Media, by Mark Miller, et al; God and Man at Yale: A Conversation with David Ross, pt.1, by Dave Krieger; Forum: Nanarchy (automated police and defense systems) by Drexler, Hanson, Finney, Szabo, Dinkelacker. Wormhole Warfare, by Robin Hanson; Reviews of *Fuzzy Thinking: The New Science of Fuzzy Logic*, and *The Children's Machine*.

11, Vol.5 No.1 (2nd Half '93): Uploading Consciousness, by Ralph Merkle; Extropian Principles v.2.5, by Max More; Traversable Wormholes: Some Implications or Contact! A Post-Singularity Phase Change, by Michael Price; A Conversation with Mark Miller, Part 2: The Day the Universe Stood Still, by David Krieger; "Bunkrapt": The Abstractions that Lead to Scares About Population and Resources, by Julian L. Simon; Reviews of *Theories of Everything, In Our Own Image: Building an Articial Person, Mirror Worlds*.

#10, Vol.4 No.2 (Winter/Spring '93): Pigs in Cyberspace, by Hans Moravec; Protecting Privacy with Electronic Cash, by Hal Finney; Technological Self-Transformation, by Max More; Mark Miller interview, by David Krieger, Pt.1: Creole Physics & the Credit Theory of Identity; Nanocomputers: 21st Century Hypercomputing, by J. Storrs Hall; The Transhuman Taste (Reviews): Two books on Ayn Rand & Objectivism; *Nanosystems; Genius.*

#9, Vol.4 No.1 (Summer 1992): The Extropian Principles, 2.0, by Max More; Extropy Institute Launches, by Max More; Persons, Programs, and Uploading Consciousness, by David Ross; Nanotechnology and Faith, by J. Storrs Hall; The Making of a Small World (fiction), by R. Michael Perry; Genetic Algorithms, by Simon! D. Levy; Time Travel and Computing, by Hans Moravec; Futique Neologisms 3; Exercise and Longevity, by Fran Finney; The Transhuman Taste (Reviews): *The Anthropic Cosmological Principle, The Blind Watchmaker, The Ultimate Resource, Population Matters, The Resourceful Earth, Bionomics.*

#8 Vol.3 No.2 (Winter 1991-92): Out of print.

#7 Vol.3 No.1 (Spring 1991): A Memetic Approach to 'Selling' Cryonics, H. Keith Henson & Arel Lucas; Privately Produced Law, Tom Morrow; Order Without Orderers, Max More; Futique Neologisms; Neurocomputing 4: Self-Organization in Artificial Neural Networks, by Simon! D. Levy; Forum on Transhumanism; Reviews of *Smart Pills, Surely You're Joking Mr Feynman, Great Mambo Chicken and the Transhuman Condition*; and more...

#6 (Summer 1990): Transhumanism: Towards a Futurist Philosophy, by Max More; The Thermodynamics of Death, Michael C. Price; The Opening of the Transhuman Mind, by Mark Plus; The Extropian Principles, by Max More; Neurocomputing Part 3, by Simon! D. Levy; Forum on Arch-Anarchy and Deep Anarchy; Reviews: *Order Out of Chaos, The Emperor's New Mind, A Neurocomputational Perspective, Loompanics Greatest Hits, The Machinery of Freedom*; Extropian Resources, and more.

#5 (Winter 1990): Forum: Art and Communication; Leaping the Abyss, by Gregory Benford; Arch-Anarchy, by A; Deep Anarchy, by Max O'Connor; I am a Child, by Fred Chamberlain; Perceptrons (Neurocomputing 2), by Simon D. Levy; On Competition and Species Loss, by Max O'Connor; A Review of Intoxication, by Rob Michels; Intelligence at Work, by Max O'Connor and Simon D. Levy; Extropian Resources, by Max O'Connor and Tom W. Bell; The Extropian Declaration, by Tom W. Bell and Max O'Connor; Our Enemy, 'The State,' by Max O'Connor and Tom W. Bell.

#4 (Summer 1989): Forum; In Praise of the Devil, by Max O'Connor; Neurocomputing, by Simon D. Levy; Why Monogamy? by Tom. W. Bell; What's Wrong With Death? by Max O'Connor; Reviews: Are You a Transhuman? Postscript to "Morality or Reality" by Max O'Connor; Efficient Aesthetics, by Tom. W. Bell; Intelligence at Work: Advances in Science by Max O'Connor.

#3 (Spring 1989) Out of print.

#2 (Winter 1989): Review of Mind Children, by Max O'Connor; Darwin's Difficulty, by H. Keith Henson and Arel Lucas; A Truly Instant Breakfast, by Steven B. Harris M.D.; Wisdomism, by Tom W. Bell; Nanotechnology News, by Max O'Connor; Weirdness Watch, by Mark E. Potts.

#1 (Fall 1988): A brief overview of extropian philosophy and an introduction to some of the topics we plan to address: AI, Intelligence Increase Technologies, Immortalism, Nanotechnology, Spontaneous Orders, Psychochemicals, Extropic Psychology, Morality, Mindfucking, Space Colonization, Libertarian Economics and Politics, Memetics, and Aesthetics; "Morality or Reality," by Max O'Connor.

Continued from page 37

Extropian cybercommunity continues to expand, now encompassing the main Extropians e-mail list (now three years old), the Exl Essay list, five local e-mail lists for arranging meetings, parties, and other joint activities, and now an Extropian presence in the Metaverse (initiated by Geoff Dale). (See the back cover for information on most of these cyberfora.)

For outside perspectives on Extropian activities, take a look Charles Platt's "Taking the N Out of Entropy" in the current issue of *Science Fiction Eye*, and the expected coverage of the conference in *Wired* (probably the September issue).

THE FUTURE: This year's successful EXTRO conference (whose keynote speaker was roboticist Hans Moravec) will be followed by the bigger and better EXTRO 2 in June 1995. Watch out for the list of speakers in upcoming issues. We will be fostering the growth of more local discussion groups and international chapters of Exl, and we will publish a new Members' Handbook. We look forward to the continued development of the Extropians cyberculture. As finances allow, Exl will expand the range of tapes, books, and other items for sale; we will build cooperation with other organization for shared goals and make contact with more scientists, technologists, philosophers, and artists to strengthen our network.

As we grow larger we will offer seminars and classes, publish and publicize public policy papers on aspects of technology, start discussion groups in more areas, supplement the general conferences with special-purpose conferences and seminars. Other ways of disseminating extropic ideas include producing extropic teaching materials for schools (e.g., critical thinking, thinking about the wise use of technology), the production of truly extropic TV documentaries, science fiction shows, and big-screen movies portraying the positive possibilities of the future.

We hope you will join us as an active participant in the Extropian movement. (See p.2 for membership information.) Help shape the future!

Not an advertisement!

Isn't it about time you attained

ULTIMATE POWER?

With the Galactomatic-1000 (TM) Basement Universe from General Cosmic, you can do everything God can do, in the comfort of your own home! This handy appliance can open up a negative-matter stabilized wormhole into a newly created universe at the push of a button.

The Galactomatic-1000 has a 27 inch wormhole opening (28 inches in Canada). It provides a full-color view of your new universes, everything from radio to gamma rays. You can climb into your new universe whenever you feel like it! Dispose of trash safely and easily by dropping it into your Galactomatic-1000 and then turning the unit off, thereby pinching the trash off from our space-time continuum forever. Great for yard waste and pesky used plutonium!

Tired of waiting for your universe to cool enough to permit the condensation of atoms? Wait no more! With the handy speed control knob, you can run your new universe at anywhere from real time to 10^{22} times faster than your own time stream. Set it to 10^5 and brew beer in seconds! Set it to 10^9 and breed unlimited quantities of valuable chinchillas in minutes! Set it to 10^{14} and watch galaxy-spanning alien empires rise and wither in an afternoon! (*)

And Basement Universes aren't just for basements any more! The Galactomatic-1000 comes with an attractive imitation wood-grain negative-matter case that makes it perfectly at home in your den or family room. The case reduces its total mass to zero, so you won't have to worry about imploding your house into a black hole, or discoloring the walls with unattractive gravitational redshifts. (**) And you don't have to worry about unlimited unidirectional acceleration due to placing large negative and positive masses next to each other. The case of the galactomatic-1000 (TM) is elegantly symmetrical, and the centers of gravity of the negative and positive matter are pre-set at the General Cosmic factory to exactly the same place, minimizing the pesky tendency toward unlimited acceleration that you may experience with other brands of wormhole. Any residual accelerations are prevented by the sturdy no-skid rubber feet, which are guaranteed not to smudge your floor.

Our patented Flexi-Law (TM) feature lets you make universes with a variety of physical laws. Make the strong force a little stronger and let primordial nucleosynthesis provide you with plenty of pure helium for the balloons at your next party!

Do you want to accomplish mighty industrial or scientific operations, or just do a few chores around the house? For a small additional charge, you can become the proud owner of a starter culture of Little Green Guys (TM). Just drop a breeding pair of these helpful aliens into your universe after it cools, set the speed control to whiz past a few millennia, and they'll create a population of trillions genetically programmed to serve you. Have their planet's wisest scholars help the kids with their homework! Send millions of coolies to weed the yard! And an army of billions armed with antimatter artillery can provide the ultimate in home security systems!

Feeling swamped by work? No problem! By having your Little Green Guys build their own Galactomatic-1000s, you can create a hierarchy of Sub-basement Universes (TM) and accomplish infinite amounts of work in a finite amount of time. (***)

Need the advantages of a Basement Universe while on the go or at work? Try our Galactomatic-50 (TM) Pocket Universe, a 3-inch wormhole in an attractive snap-shut case. This wormhole connects into the same universe as your home Galactomatic-1000, so you can have your Little Green Guys fetch things from your house, or take things home for you (as long as they're less than three inches across, of course!). You'll never worry about forgetting your keys again!

All your Galactomatic-1000 needs is a few square feet of floor space and a few quadrillion watts of electricity (available in 110V and European 220V models.)

Why wait? Call 1-800-GODLIKE for the location of a General Cosmic retailer near you.

(*) Caution: do not leave the Galactomatic-1000 unattended when evolution of alien life is possible. General Cosmic is not responsible for alien invasion or conquest of your home and possessions.

(**) Although the Galactomatic-1000 has no mass, it still has volume, so a shipping and handling charge will apply.

(***) Sorry, only countable infinities of work can be performed by Sub-basement Universes. This offer void where prohibited by the Axiom of Choice.

The Galactomatic-1000 is guaranteed against defects in materials and workmanship until the end of time, or for the life of the owner, whichever is longer. Our liability is limited to the cost of replacement of the Galactomatic-1000 itself. General Cosmic is not liable for damage to the person or property of the user, or to the contents of the universes created by the Galactomatic-1000, including damage due to black-hole creation or antimatter spills. General Cosmic is not responsible for normal wear of the case or any other damage due to the second law of thermodynamics.

Little Green Guys (TM) are guaranteed for ten thousand generations or one million years, whichever comes first. After this time, genetic drift may cause undesirable features to arise. We advise destroying their civilization when the warranty expires. By keeping a few spare Little Green Guys in the freezer, you can always start again!

The Origins of Order:
Self-Organization and Selection in Evolution

by Stuart Kauffman
Oxford University Press, New York, 1993.
709 pages, ISBN 0-19-505811-9

Reviewed by Reilly Jones

This is a landmark book, encompassing daring new holistic ideas about living systems. Stuart Kauffman is Professor of Biochemistry and Biophysics at the School of Medicine, Univ. of Pennsylvania, and External Professor at the Santa Fe Institute. The application of the mathematics of complexity theory to specialized branches of natural science is progressing rapidly. Kauffman's is the first comprehensive effort to apply complexity to the theory of evolution by placing evolution within a larger biophysical framework of potential universal laws.

By looking at evolution in such a new, over-arching way, Kauffman has placed himself at a considerable distance from established reductionists in developmental biology. He will doubtless be proven wrong in many of the details he covers, but by being bold, general and asking questions arising from new graphical computer modeling techniques, he points in many different directions for fruitful research.

The book begins with an introduction containing the contemporary theory of evolution and some of the peripheral challenges to it, along with Kauffman's effort to place Darwinism within a larger framework of biophysics. **Part I** is entitled "Adaptation on the Edge of Chaos" and deals with fitness landscapes and adaptation in sequence spaces (protein and DNA). Some bold hypotheses on construction requirements of complex evolving ecosystems are presented based on novel modeling techniques and the technological promise of Adaptive Molecular Evolution is outlined. **Part II** is entitled "The Crystallization of Life" and deals with the origins of life, metabolism and coding. Chapter 10 on random grammar models is a gem all by itself; it contains open-system analog modeling techniques of biological, economic, technological and cultural systems. **Part III** is entitled "Order and Ontogeny" and deals with cell differentiation and morphology. This section reads more like a textbook than the earlier sections, it contains much collected research material but no conclusions on the relative influence of spontaneous order in within-cell versus between-cell genetic regulation.

The first two sections of the book are written for general science readership although some familiarity with complexity theory would be of help. The last section is more of interest to the developmental biologists. The combination of complexity and evolution brings forth new concepts such as ecosystem attractors, extinction and speciation power laws, and frequent definitions of 'spaces' (what we used naively to call "systems"). In fact, there is such an abundance of 'spaces' throughout the book that Kauffman could be characterized as 'space'-happy.[1] He frequently uses 'If/then statements' typical of much of biology (as opposed to empirical laws typical of physics). For example, "If it is the case that systems poised between order and chaos are indeed the natural culmination of selective evolution, we shall have found deep laws indeed."

He clearly is aware of the new ground he is breaking with his holistic point of view, but throughout the book shows deep respect for the body of knowledge containing reductionist microphysics at the cellular level. In his own words: "The theories presented are merely the beginnings of a new area of thought and investigation in biology, chemistry, and physics — perhaps even in economics and other areas of social sciences. The spirit of all the ideas discussed... is a kind of unrepentant holism and a sense of synthetic biology rather than the familiar reductionistic analytic mold."

His search for universals, or what he terms the "physics of biology" leads him to conclude that, "Biology is surely harder than physics." He proposes some very broad potential universal laws that have direct relevance to Extropian principles of directed self-transformation and boundless expansion. The broadest law of all is of great interest towards development of more complex selection systems than mere survival; it involves the possibility of evolutionary feedback. "...The capacity to evolve is itself subject to evolution and may have its own lawful properties. The construction principles permitting adaptation, too, may emerge as universals."

These construction principles will be profound. What theory of morphology would

enable us to predict features of organisms that would evolve on any planet, in any environment? What forms of life are highly unlikely to evolve and how does selection work to achieve new families of forms? How does such a universal theory of forms fit the empirical facts of our own past and what part does random drift play in the speed at which forms evolve? These questions are of vital interest to potential development of new biological ecosystems or desired alterations of pre-existing systems.

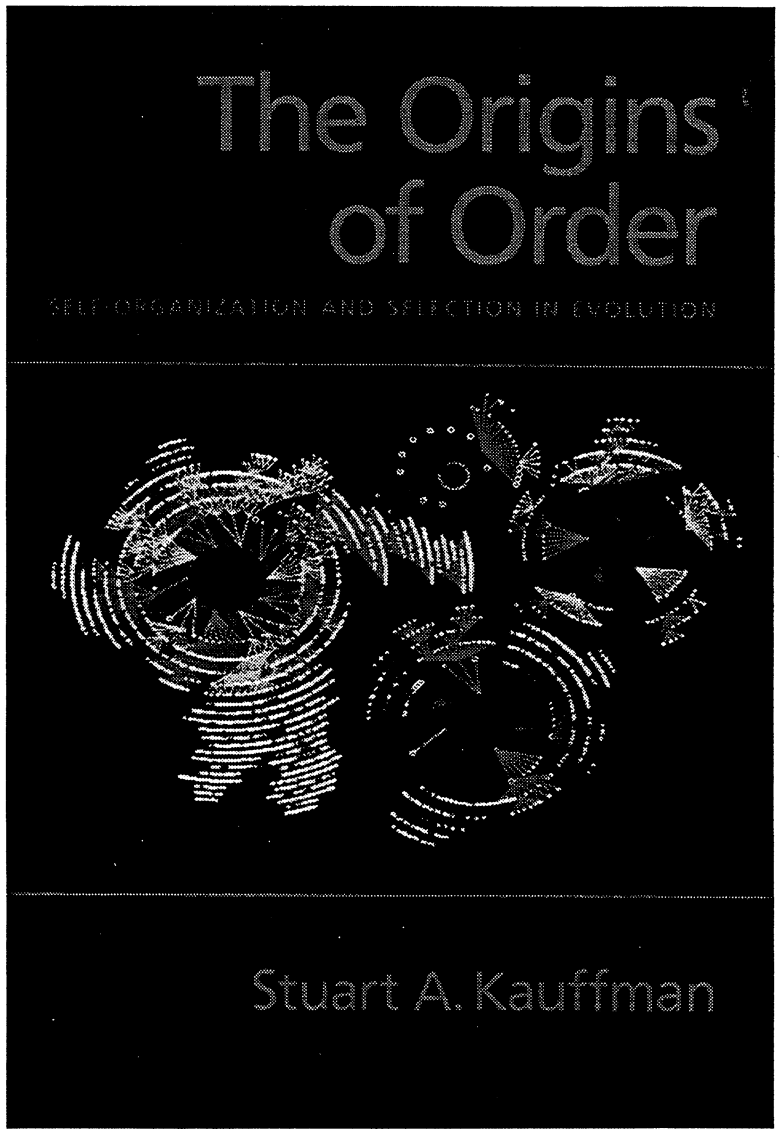
An important hypothesis that Kauffman reaches strikes me as a description of self-interested (myopic) individuals interacting in a free market. "In coevolution, organisms adapt under natural selection via a *metadynamics* where each organism myopically alters the structure of its fitness landscape and the extent to which that landscape is deformed by the adaptive moves of other organisms, such that, as if by an **invisible hand**, the entire ecosystem coevolves to a poised state at the edge of chaos." He even treads the dangerous ground of social science when he examines "What is a functional whole and how does it transform when its components are altered?" He finds features in technological, economic and cultural systems that are phase transitions between finite and potentially infinite growth.

We can relate self-transformation to his concept of "evolvability" and boundless expansion to his concept of "sustained fitness." There are potential biophysical laws that govern what paths our future evolution can take such that we can choose our destiny and transform ourselves faster in more complex ways than we safely could have without these models. "...Proper evolutionary tuning of mutation rate, population size, and landscape structure might simultaneously optimize both evolvability and sustained fitness."

Kauffman explores the possibility of genetic design rules. One design consideration is the amount of DNA needed to generate novel cell types. For example, if we wanted to add cell types to boost the complexity of our consciousness or to produce regenerative neurons to increase longevity, we would need a hefty increase in the amount of DNA in our chromosomes. 'Junk' DNA may support the complexity of cell types in a functionally whole way that will not reduce down to function codon by codon. It is, in fact, possible that much of large-scale order in genetic design "is a direct reflection of fundamental features of polymer chemistry."

Kauffman discusses an epistemological boundary we should keep in mind when working with complex design considerations in genomic or immune regulatory systems. He points out that these systems are so fluid that they "are dancing away from us faster than we may ever be able to grasp them. ...We may never to be able to carry out the reductionistic dream of complete analysis but will want nevertheless to understand how these systems work."

While he doesn't use the terminology of memetic evolution, he does interpret results of his models as showing how meaning and learning arise in complex organisms. Meaning does not arise in his digital, Boolean models but does arise in his random grammar models through modular interactions exhibiting their functional couplings within an evolving system. The appearance of meaning in this model is structurally similar to theories of human meaning arising from embodiment of the mind and social interactions.[2] We should be able to model how meaning will change in the future with accelerating self-transformation and more complex social interactions. Learning is characterized as "a walk in synaptic weight space seeking good attractors. Learning itself may be the fundamental mechanism which converts chaotic attractors to orderly ones." This is similar to current theories of memory formation through nitrous oxide cellular diffusion within statistical ensembles of neurons.[3] The unit of selection is the individual cell but appears to be a group selection because of the mechanisms of the attractor.



Kauffman identifies two major limitations to selection, what he terms "complexity catastrophes." In one scenario, as the complexity of the species increases, the fitness landscape it is operating in deforms to lower the overall possible fitness level. In the other scenario, as the complexity of the species increases, the population is unable to hold to the fitness peaks and falls back to a lower average fitness. Much of the discussion in Part I of the book discusses strategies to increase complexity in species while avoiding either of these catastrophes. A very promising possibility is the mapping of complex cost surfaces with the goal of optimizing energy flow to allow for increasing levels of civilization.

While he makes productive use of his NK fitness and Boolean models in many areas, he is careful to point out the inadequacy of digital models to really approximate the analog biological world. He does not hide his excitement over the potential of random grammar models to unite the natural sciences with the social sciences. He hopes to find universal classes of behavior in functionally whole systems through exploration of "grammar space." "Thereby we may obtain models of functional couplings among biochemical, technological, or ideational elements without first requiring detailed understanding of the physics or true laws governing the couplings."

In very strong theoretical support for boundless expansion, a sequence is traced from the open thermodynamic system on earth prior to the origin of life, to cascades of catalyzing organic molecules, to the explosion in organic diversity we see today. He says "open chemical systems can be self-extending. The fact that the biosphere as a whole is supracritical serves, I believe, as a fundamental wellspring for a

persistent increase in molecular diversity." As an aside, I could not help but reflect that recent pictures of the large-scale structure of galaxies in the universe look remarkably similar to what Kauffman calls "filagreed fog" random grammar end-states.[4] He then makes a random grammar model connection between bounded physical systems such as thermodynamic constraints in chemistry and budget constraints in economics, and "the worlds of ideas, myths, scientific creations, cultural transformations, and so on" that are unbounded.

Two very interesting results from these models are of particular note. The first is that we model each other's potential behavior (analogous to trust) in such a way that society tends towards a poised state at the edge of chaos. In essence, high degrees of trust (the most complex, discriminating models) will lead to decreased trust while low degrees of trust (the most brute, simple 'tit-for-tat' models) will lead to increased trust. While everyone knows that familiarity breeds contempt, it is also true that contempt breeds familiarity. The second result is that: "The extent to which the planner looks into the future governs whether the economy grows at all, slowly, or rapidly. ...Technological growth is strongly correlated with the capacity to see its implications.... If the consumer places little value on the future, diversity of goods and services remains small." The clear implication of this is that the Extropian principles, if adopted, will by themselves be self-fulfilling. There is good reason for dynamic optimism: It works!

Spontaneous order in the absence of outside work is found throughout biology in the form of small attractors. These attractors represent cell types, immune responses, etc. and are easily attained by natural selection to produce stable structures. However, in a strongly counter-intuitive finding, as the complexity of systems increases, natural "selection cannot avoid the order exhibited by most members of the ensemble. Therefore, such order is present not because of selection but despite it."

This finding of such an inseparable relationship between self-organization and selection that varies with the scale of the parts and the whole is typical of the holism found throughout this book. Other major examples of functional wholes include the idea of autocatalytic polymers being 'chicken-and-egg.' There is a lengthy outline of 'knower-and-known' systems where representation of and interaction between entities in their environment depends on stability of both the entities and the environment. "In a phrase, organisms have internal models of their worlds which compress information and allow action." Also, proper growth of organisms depends on a control system of 'map-and-interpretation.' "...The entire genomic system is, in reality, a single coupled system whose attractors constitute both map and interpretation at once." This holism of Kauffman's seems akin to the 'undivided universe' ontological interpretation of quantum mechanics by physicist David Bohm based on experimental results of non-locality.[5] I am also reminded of the position of the English philosopher Frances Bradley: "And what I repudiate is the separation of feeling from the felt, or of the desired from desire, or of what is thought from thinking, or the division of anything from anything else. For judgment is the differentiation of a complex whole, and hence always is analysis and synthesis in one."[6]

Technology is being developed that will allow experimentation in areas that have the potential of transforming society. Kauffman proposes that we create life anew. He notes that function must be extremely redundant in DNA and protein sequence space, and therefore, life is created far more easily than we have previously thought. "Life is an expected, collectively self-organized property of catalytic polymers. ...Self-reproduction and homeostasis, basic features of organisms, are natural collective expressions of polymer chemistry." The experiments must risk a complexity sufficient to achieve catalytic closure, but once accomplished, the path is open to make empirical tests on coding mechanisms to help understand why DNA coding is so prevalent today.

Kauffman proposes that we create life anew. He notes that function must be extremely redundant in DNA and protein sequence space, and therefore, life is created far more easily than we had previously thought.

Kauffman is seeking patents in a field of molecular nanotechnology that he calls "Applied Molecular Evolution." There is a finite number of enzymes that will catalyze all polymer reactions. The drive is on to explore DNA, RNA and protein sequence space to custom design for any function desired at all. The potential for individualized drug treatment is promising in immunology and cancer research. "...Using antibodies from an infected individual, it becomes possible, in principle, to find vaccines for diseases where the pathogen is not yet even known!" Such a powerful technology must be available on the private market for commercial uses and consumer benefits. The military potential for rapid development of customized, biological offensive weapons and equally rapid defenses against them is enormous. Biological warfare may not be plague versus plague, but highly accurate, targeted strikes and selected defense responses.

Finally, Kauffman uses artificial life researcher Thomas Ray's Tierra model ecosystem to show how closely artificial extinction patterns obey the same power law that has been recorded in Earth's fossil record. The artificially-generated graph is a close match with the actual graph. I bring this up because Ray's latest paper references Kauffman and discusses "ecological attractors" at length.[7] Ray also confirms the superiority of analog models to digital models for realism and even references Hans Moravecs' article "Pigs in Cyberspace" in *Extropy* #10.

Kauffman's deepest insight is a direct challenge to the current view of our lives as being merely the result of a series of frozen accidents. "I have made bold to suggest that much of the order seen in organisms is precisely the spontaneous order in the systems of which we are composed. Such order has beauty and elegance, casting an image of permanence and underlying law over biology. Evolution is not just 'chance caught on the wing.' It is not just a tinkering of the ad hoc, of bricolage, of contraption. It is emergent order honored and honed by selection."

This book is a challenging read for those interested in shaping spontaneously ordered living systems towards increased complexity and meaning. The search for a "physics of biology" to help minimize tragic and time-consuming trial-and-error methods of human-directed evolution is brought to the forefront of scientific priorities by Kauffman's bold thinking.

Notes:

1: Phenotypic space, morphospace, protein space, sequence space, trait spaces, genotype space, complex fitness spaces, RNA space, catalytic task space, shape space, space of biological systems, state space, synaptic weight space, local strategy space, action space, space of symbol strings, peptide space, space of possible polymers, open state space, fixed state space, grammar space, composition space, parameter space.

2: Lakoff, G. Women, *Fire, and Dangerous Things: What Categories Reveal About the Mind.* Chicago: Univ. of Chicago Press, 1987. Johnson, M. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: Univ. of Chicago Press, 1987.

3: Edelman, G. *Bright Air, Brilliant Fire: On the Matter of the Mind.* New York: BasicBooks, 1992. Schuman, E. & Madison, D. "Locally Distributed Synaptic Potentiation in the Hippocampus." *Science* 28 January 1994: 532.

4: Travis, J. "Cosmic Structures Fill Southern Sky." *Science* 25 March 1994: 1684.

5: Bohm, D. & Hiley, B.J. *The Undivided Universe: An Ontological Interpretation of Quantum Theory.* London: Routledge, 1993.

6: Bradley, F.H. *Appearance and Reality.* 2nd ed., Oxford, 1897.

7: Ray, T. In press. "An evolutionary approach to synthetic biology, Zen and the art of creating life." *Artificial Life* 1(1): xx–xx. MIT Press. I found this paper in the AI Expert Forum Library on Compuserve, dated 21 October 1993.

GOOD MOOD:

The New Psychology of Overcoming Depression

by Julian L. Simon

Open Court, La Salle, Illinois, 1993.
311 pages; ISBN: 0-8126-9097-4

Reviewed by Max More

"Life is hard," someone said to Voltaire. "Compared to what?" Voltaire answered.

Good Mood will appeal to and benefit many persons who are not and have not been genuinely depressed, as well as proving invaluable to depressives. Typically, those of us who think of ourselves as extropian in our values and goals, like to choose our emotional states so as to remove blockages to effective thinking and action. Being deeply attracted to positive self-transformation and the persistent sculpting of a new and superior self, we seek enhanced self-awareness and new tools for self-control and self-definition. Simon's approach to the understanding and treatment of depression (easily extended to anger, anxiety, dread, and apathy) — Self-Comparisons Analysis — builds on the foundation provided by modern cognitive psychology and its applications to therapy. According to cognitive therapy, whether it be Aaron's Beck's, Albert Ellis's Rational-Emotive Therapy, or aspects of Branden's self-esteem psychology, our emotions are largely determined by our thinking and are not completely inexplicable forces fully distinct from our cognition. By changing habitual thinking patterns we can leave behind depression, anxiety, apathy, and inappropriate anger, choosing more effective emotional states.

Apart from its value in teaching us how to become more self-aware and self-understanding, *Good Mood* addresses problems common among persons with exceptionally high standards for achievement, intellect, and behavior. The more we wish to achieve, the more intelligent and wise we wish to be, the grander our goals for ourselves, the stronger the emotional problem that can result when such ambition is combined with certain self-defeating cognitive patterns. The more we want, expect, or demand of ourselves the worse we potentially can feel if we believe we are failing our ideal standards. How *do* we make ourselves sad, depressed, anxious, or angry, and how can we avoid such self-frustration, whether it be occasional or chronic, mild or acute?

Simon presents the essence of his approach in the form of a "mood ratio", according to which:

$$\text{Mood} = \frac{\text{(perceived state of oneself)}}{\text{(Hypothetical benchmark state)}}$$

Negative self-comparisons (or "neg-comps") result in a Rotten Ratio where we conceive of our actual situation to be worse than some benchmark standard. Your benchmark state of affairs may be one that you were accustomed to and enjoyed, but which has gone; it may be something you expected to happen, such a promotion, a marriage, or getting a book published, but which never came about; it may be something that you had hoped for; it may be something that you feel obligated to do but are not doing; or it may be the achievement of a significant goal that you sought but failed to accomplish.

Negative self-comparisons cannot, alone, cause depression, or anxiety, anger, or dread, Simon explains. Neg-comps lead to depression, for instance, when they are accompanied by a sense of helplessness to change your situation. Sadness results from this assessment of helplessness. If a neg-comp combines with something other than helplessness, the result may be anger or determination. If the sadness produced by a neg comp plus helplessness persists, it becomes depression.

The healthy, effective response to negative self-comparisons and the emotional distress they engender is to launch into changing the circumstances involved in the neg-comp. Responding with anger not only masks the pain (so long as the anger continues to be experienced) but can galvanize you into an attempt to change the situation. Whereas, in depression, a person has lost hope and assumes they can do nothing to improve matters, anger arises where they feel frustrated in their efforts to remove the source of distress. A third possible response involves lying to yourself so as to banish the emotional pain. When lying about one's personal situation takes over from other responses the result can be schizophrenia or paranoia.

As we can see from this, Simon's cognitive model of depression allows us to understand the mechanisms producing many moods other than depression. We can see anxiety as produced by a neg-comp where the numerator of the mood ratio is an anticipated or feared outcome. Whereas the depressed person feels the feared outcome to be unavoidable, the anxious person feels uncertain of the outcome, and may feel a little less helpless about the situation. Depression, being past or present-oriented, drains energy with sadness, but anxiety causes a higher level of arousal as the sufferer dwells on the lack of certainty of the future state of affairs.

Mania can be seen as "the condition in which the comparison between actual and benchmark states seems to be very large and *positive*, and often it is a condition in which the person believes that he or she is able to control the situation." [54] *Apathy* takes hold if a person responds to their painful neg-comps by giving up goals. Without goals there can be no failure; but neither can there be the thrill of the chase and the joy of achievement. Finally, in terms of this approach, various "positive feelings arise when the person is hopeful about improving the situation — changing the neg-comp into a more positive comparison — and is actively striving to do so." [55]

The theoretical approach of Simon's self-comparisons analysis and values therapy yields a diverse array of practical means of tackling depression (or milder forms of sadness) or, appropriate modifications being made, apathy, anger, or anxiety. Leaving aside detail and example, these routes to good mood can be summarized as follows:

"These are the possible tactics: 1. Improve the numerator in your Mood Ratio, by getting rid of misconceptions about yourself, or by learning that your capacities to influence events in a desirable direction are greater than you thought. 2. Alter your denominator to make it less formidable, by changing the benchmarks against which you compare your actual state of affairs. 3. Change the dimensions on which you habitually compare yourself. 4. Retrain yourself so that you seem to yourself more competent and less helpless. 5. Reduce the number of comparisons you make each day, by immersing yourself in work or altruistic activity, or by recourse to meditation or related devices. 6. Examine your basic values to learn what is important to you that may influence your wanting to be depressed or not wanting to be depressed." [244]

This last tactic, that of examining your values and reordering them, differs from strategies involving changing your numerator and denominator. Simon devotes several chapters to this especially philosophical approach to mood control, including discussion of Victor Frankl's logotherapy and the effects of religious conversion on depression. Cognitive therapies, including Simon's variant, naturally see a place for philosophical analysis in psychology: Since emotional responses involve judgments (usually largely unconscious), the clarification of values and adjustment of

perspective can alter our feelings and dispositions.

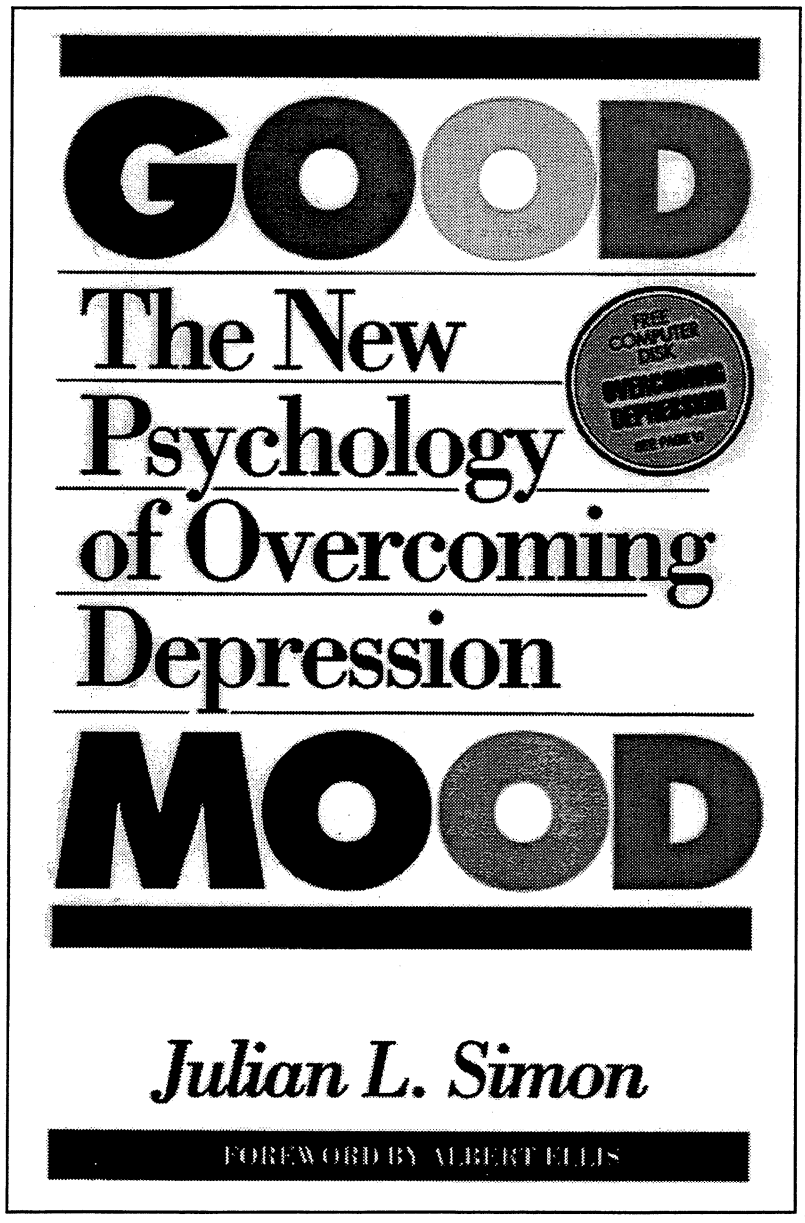
> Does this anti-depression tactic seem more like philosophy than psychology? Choose the label that you like. But more and more, the wisest psychologists have come to view many (though not all) depressions as philosophical in origin, and therefore as requiring a change of philosophy for a cure; some philosophers have known this for thousands of years....Ellis and Harper put the matter bluntly: "For effecting *permanent* and *deep-seated* emotional changes, philosophic changes appear virtually necessary." [157]

The method of improving your denominator by counting your blessings and adjusting your focus ties in neatly with some elements of the extropian attitude of dynamic optimism.[1] In my explication of dynamic optimism, the first of eight characteristics listed was "selective focus". This recommends ignoring unpleasant, painful, frustrating things unless attention to them is of the right kind (active problem solving rather than gloomy reflection) and can yield results. D.O. also suggests both content and context reframing in which you place a more positive, empowering interpretation on an event (content reframe) or you alter the context in which the event is seen. An example from Simon of what amounts to a content reframe would be where "Instead of comparing your minor arthritis with perfect and painless freedom of movement, you shift to comparing yourself with a paralytic." [157] A dynamically optimistic change of focus would involve something more active than comparing your situation to a worse possible situation; it would involve selectively focusing on lessons to be learned from a difficulty, and ways of moving forward from where you are.

Of course there's more to values therapy than changing your denominator. In common with life management approaches like that of Stephen Covey or Alan Lakein[2], values therapy requires the participant to systematically analyze their desires and goals so that they can be coherently ordered. Such a procedure involves 1. asking yourself what you want in life; 2. ranking these desires in accordance to their personal importance; 3. asking yourself if you have missed any truly important desires; 4. looking for conflicts in your list of wants; 5. taking "steps to resolve the conflicts between higher-order and lower-order values in such manner that higher-order values requiring you not to be depressed are put in control". [220] In common with the rest of Simon's approach, values therapy offers not a quick fix but an effective method that requires attention and perseverance. Discovering the structure of your wants is harder than it sounds, requiring contemplation to come up with a comprehensive list of wants and analysis to determine their relative importance to you, all things considered.

We can best discover our wants by looking inside ourselves, suggests Simon, and not by searching for something universal in humanity. Unlike Maslow or Selye, who agree that our basic values are based in our biology (but who disagree over what those values are!), Simon thinks it more plausible that there is a wide range of basic values. This individualistic approach, though unsatisfying to formalists, seems to me to encourage genuine self-discovery over constructing dubious theories about human nature. Even if there *are* basic, biologically-grounded wants common to all humans, we might best uncover them by each examining our individual natures.

Given the importance of self-transformation[3] to Extropians, we may be especially interested in Simon's view of how good mood relates to a focus on yourself and your wants or on wants or values outside the self. Simon emphasizes the value of contributing to the good of others, a recommendation that we find in practically all writers on the subject. Personal fulfilment, it is said, comes from looking beyond your personal interests. Consider this passage from Frankl, quoted by Simon:



> "If the meaning that is waiting to be fulfilled by man were really nothing but a mere expression of self, or no more than a projection of his wishful thinking, it would immediately lose its demanding and challenging character, it could no longer call man forth or summon him.... Human existence is essentially self-transcendence rather than self-actualization. Self-actualization is not a possible aim at all, for the simple reason that the more a man would strive for it, the more he would miss it. For only to the extent to which man commits himself to the fulfillment of his life's meaning, to this extent he also actualizes himself. In other words, self-actualization cannot be attained if it is made an end in itself, but only as a side effect of self-transcendence." [Frankl, *Man's Search for Meaning*, 156, 175]

If we find the meaning or purpose of our lives to lie within us, our primary goal being to actualize our potential, to strive to move towards an evolving ideal image of ourselves, will our lives lack the demanding character referred to by Frankl? I think not. If self-actualization is taken to mean that our central purpose is the enrichment of ourselves as we find them, Frankl's point is convincing. We would soon find ourselves lacking inspiration — there would be no lofty ideal to draw us forth. On the other hand, an extropian interpretation of self-

actualization as self-overcoming, as self-augmentation, clearly has this summoning power.

Terminology can be confusing here: Some Extropians have suggested that the principle of Self-Transformation be renamed "Self-Transcendence". Self-transcendence has two meanings: (a) Going outside ourselves to find purpose in contributing to others and to abstract goals; (b) Seeking to develop beyond the limits of our current self-stage towards an image of our ideal self. In the second sense, the principle of Self-Transformation embodies a commitment to self-transcendence but doesn't *require* service to others or to causes beyond ourselves as a primary goal.

Frankl does have an important point though. If we spend most of our time focusing on ourselves, we will sharply narrow our opportunities for self-improvement. An outward-looking attention to the well-being and growth of other persons and a deeply involved contribution to causes that go beyond our immediate self-interest are necessary for us to fully develop and to exercise our talents and virtues. Benevolence, generosity, and lack of excessive preoccupation with self are certainly healthy and good for us. However, putting self-transcendence ahead of self-actualization, in the senses given them by Frankl and endorsed by Simon, subjects us to the danger of self-sacrifice and manipulation by others who want to use as tools to promote *their* ends.

As Simon notes, "Values Therapy is especially appropriate when a person complains that life has lost its meaning — the most philosophical of depressions." [217] "Values therapy may be thought of as a systematic and understandable form of what used to be called 'changing one's philosophy of life'. It operates directly on the person's view of the world and himself." [229] This suggests to me that the development and promulgation of Extropian ideas may be seen as a kind of cultural value therapy, preparing people for the coming time when aging will be a thing of the past. Most people, confronted with the prospect of extreme longevity and possible physical immortality, express the fear that life will lose its meaning. Such persons are used to conceiving of their lives in the well-established pattern of infanthood, childhood, adolescence, adulthood, old age, and death. They ask how anyone could keep working at the same job for millennia, they expect to run out of stimulating things to do and learn, and they imaginatively project their current human forms into their limitless future. Perhaps we can understand part of our task as Extropians to be taking the initiative in acting as cultural psychotherapists, preparing the world for the tremendous changes ahead.

Good Mood, as I suggested at the beginning, is not just for individuals with clinical depression. Apart from supplementing our arsenal as cultural psychotherapists, most of us, some of the time, and to some degree, suffer from the self-defeating thinking patterns dragged into the light by this book. Why do many non-depressives have trouble getting down to the tasks "that they know they 'should' undertake for their own good? You know the answers: a. They prefer leisure to exertion. [...] b. They convince themselves that not doing the task really won't be very harmful, and doing it will not be very beneficial. c. They fool themselves that they are just postponing the task for a short time, and keep repeating the procrastination. d. They start the task then give up because they lose patience." [131] These patterns are shared by the depressive, just in an exaggerated, darker form. The techniques and cognitive discipline recommended in *Good Mood* therefore can be of benefit to all of us who are not perfectly motivated.

Julian Simon's new book may come as a surprise to readers familiar with his fine works in population and resource economics[4] or his more recent work developing resampling in statistics. Simon exemplifies the kind of intellectual that readers of this journal usually appreciate: One who is able to apply his intellectual acuity not only in his native field of economics, but also in the quite distinct areas of statistics and psychology (not to mention business administration). Furthermore, in each of these fields his thinking has challenged conventional thinking and helped clear the way to boundless expansion, improved thinking, and greater effectiveness. It may be that Simon's advance in the cognitive treatment of mood disorders can be attributed to his multidisciplinary mind, especially his economics background. Economists typically think about values and wants and their trade-offs in a highly analytical manner.

Glum cyberheads will be happy to know that, for a nominal charge, software ("Overcoming Depression 3.0") is available to accompany the book. The program provides both lessons derived from the book and interactive tutorials that help you identify and modify self-defeating thinking patterns.

NOTES

[1]See my "Dynamic Optimism: Epistemological Psychology for Extropians", esp. the first section, in *Extropy* #8, (vol.3 no.2, Winter 1991-92).
[2]Stephen Covey, *The Seven Habits of Highly Effective People*; Alan Lakein, *How to Get Control of Your Time and Your Life.* Both of these books have been reviewed in *Exponent*, the Extropy Institute members' newsletter.
[3]See my essay "Technological Self-transformation" in *Extropy* #10 (vol.4, no.2), Winter/Spring 1993.
[4]For example, *The Ultimate Resource.*

EXTRO 1

The First Extropy Institute Conference on Transhumanist Thought

Proceedings and audio tapes

PROCEEDINGS (approx. 160 pages)

Contents

AUDIO TAPES OF EXTRO 1 SESSIONS

Michael Perry: Mathematical Immortalism $10

Max More: Pancritical Rationalism: An Extropic Metacontext for Memetic Progress $11

Tom Morrow: Why Respect the Law? The Polycentric Justification of Jurisdiction $11

Simon! D. Levy: SIMNET: A Neural Network Simulator for Modeling Complex Dynamical Systems+ **Romana Machado:** Five Things You Can Do To Fight Entropy Now $11

Hans Moravec: The Age of Robots $12

Chris Heward: The Endocrinology of Aging: Can We Prevent Senescence? $11

Ralph Merkle: Cryonics, Cryptography, and Maximum Likelihood Estimation $12

Extropian Virtual Community: Past, Present, and Future $10

Videotapes of some of the sessions may become available in the near future.

Prices for audio tapes include postage. Exl members & conference attendees may subtract $1 per tape.

Proceedings: $18 + $2 postage

Make checks payable to "Extropy Institute" and mail to 13428 Maxella Avenue, #273, Marina Del Rey, CA 90292.

More Dynamic Optimists — NOT!

"Lee De Forest has said in many newspapers and over his signature that it would be possible to transmit the human voice across the Atlantic before many years. Based on these absurd and deliberately misleading statements, the misguided public... has been persuaded to buy stock in his company." U.S. District Attorney, at the mail fraud trial of Lee De Forest, "Father of Radio", 1913.

"That Professor Goddard and his 'chair' in Clark College and the countenancing of the Smithsonian Institute does not know the relation of action and reaction and of the need to have something better than a vacuum against which to react – to say that would be absurd. Of course he only seems to lack the knowledge ladled out daily in high schools." *The New York Times*, January 13 1920.

"The ordinary "horseless carriage" is at present a luxury for the wealthy; and although its price will probably fall in the future, it will never, of course, come into as common use as the bicycle." *The Literary Digest*, October 14 1889.

I must confess that in my imagination, in spite even of spurring, refuses to see any sort of submarine doing anything but suffocating its crew and floundering at sea." H.G. Wells, in *Anticipations*, 1901.

"In my opinion such a thing is impossible for many years... [P]eople... have been talking about a 3000 miles high-angle rocket shot from one continent to another, carrying an atomic bomb and so directed as to be a precise weapon... I think we can leave that out of our thinking." Dr. Vannevar Bush, testimony to Senate Committee, 1945.