The effect of shared vulnerabilities on the survival of complex systems

1. Overview of the Problem

What makes some systems survive and others break down in complex systems of systems?

Biodiversity and Human Capital as Critical Resources: Biodiversity is no doubt a critical resource in today's world. Most pharmaceutical drugs, for example, are derived from one or another natural product of a usually esoteric species. E. O. Wilson (2000), the father of Sociobiology and one of the most important living biologists, wrote that "Community ecology in particular is about to emerge as one of the most significant intellectual frontiers of the 21st century."

Biodiversity as a Source of Energy: The global ecosystem cycles ~2.5 PetaWatts, or the power of 1.5 million nuclear power plants. The Department of Energy, aware of its importance, has recently announced it plans to focus on biomass resources (Biocycle, 2002). *Designing engineered communities that can harvest energy efficiently requires knowing which species can function together*.

Biodiversity as a Model System: Eco-biological systems also constitute a powerful model system, for they are the oldest, most complex, best adapted and best studied of the complex systems we know of. This proposal aims to provide solutions to some fundamental problems in that field, and apply the insights gained from the study of biodiversity to other critical infrastructure. Among these is an important application to understanding perhaps the greatest current risk to what is no doubt our most important critical infrastructure: our human capital.

Biodiversity: A Puzzle of Evolution & Ecology

What is the origin of biodiversity? Why are there so many species? Why haven't the fittest of them all outcompeted the rest to extinction? And a related question: why are close to 30% of gene loci in most every species examined polymorphic? Why have not the fittest of all alleles (gene versions) become fixated (acquired frequency of 100%) in the population? Most relevant to this call, how can we know *which* sets of species can coexist in an environment and which cannot?

Darwin's theory of natural selection sustains that, eventually, only the fittest among competitors will survive. With the exception of rare traits with multiple independent origins, every trait that is fixed in a population must correspond to a bottleneck in evolution whereby *every* individual alive today descends from the *same* common ancestor. This means that, for every genetic difference that distinguishes closest neighbor species in a phylogenetic (evolutionary) tree, there is a corresponding extinction of all individuals not bearing the new and surviving allele (gene).

The standard Darwinian explanation for this set of extinctions is that, in the face of limited resources, populations have a fixed maximum size, and thus any genotype with a fitness advantage, however small, will tend to increase in frequency and thus eventually drive competing genotypes to extinction (i.e. reduce their abundance to below one individual).

In the presence of multiple resources, stable equilibria with more than one species, namely one per resource, can exist. A persistent puzzle in evolution, however, is the empirical fact that multiple species can coexist per resource, contrary to the predictions of Darwinian doctrine. Many species sharing an ecological niche have not driven each other to extinction (Churchfield et al., 1999; Wood et al., 1999). This fact forces us to reevaluate some of the tenets that have been held for the last 150 years.

Interestingly, this coexistence does not happen only between different species. Up to 60% of all genes in a given species are polymorphic, meaning they exist in multiple different forms (alleles) in the population. This happens even for genes that are known not to be neutral. Why does not a dominant gene with maximum fitness outcompete the rest?

2. Proposed R&D

2.1. Technical Approach

Shared Vulnerabilities

We propose to cast the problem of how and when multiple genotypes can coexist stably in a framework of dynamic equilibrium. The probability that multiple species will *de novo* have identical fitness is very low. For a stable equilibrium to exist with multiple coexisting genotypes, what is needed is a restorative force that reduces the ratio of mortality to birth rates when a population's size fluctuates downward, and vice-versa. What is this population-size-dependent force, then?

An often ignored consequence of the Red Queen hypothesis for the origin of sex, a theory which has received considerable empirical support in the last few years, is that *more than half of all deaths* (or losses of fertility) in all sexual species (or their ancestors) are *caused by parasites*. This suggests parasites as a natural candidate for the force determining population sizes at equilibrium. Parasite-driven deaths grow with population density, as required for a stable equilibrium.

A fundamental requirement of stability in the theory is that parasites not spread *equally* across the two coexisting populations, for if they do, the effective population size determining the likelihood of an individual becoming infected will be the

combined total across both populations, and thus a reduction in numbers of one of the two will not lead to a corresponding replenishment, leading to an eventual extinction of the population with lower fitness. This leads to a fundamental prediction of the theory: *species which share all their vulnerabilities will compete until all but one are extinct; species with unique (non-shared) vulnerabilities to frequent parasites can coexist stably*. Note that a vulnerability does *not* mean that *every* member of a species gets killed.

For drowning sailors at sea, survival is not a competition between sailors; it is a battle against the sea. Analogously, species in today's populated Earth live in a sea of parasites. We suggest Darwinian competition between individuals in the same species is not the major factor in the evolution of diversity on Earth today. Instead, *an arms race against parasites is the major force behind biodiversity*. As a consequence, we suggest *more of our genes have evolved to combat parasites than have evolved to adapt to our physical environment*. If this is borne out by experimentation and analysis, it would have major implications for our understanding of genomes and their use for biomedical applications.

We propose that the restorative effect on genotype abundance described above is a principal cause of sequence polymorphisms (persistent variations) in sexual and asexual species alike, as well as of the coexistence of multiple species per resource in an ecological niche. The theory predicts *genes that exhibit polymorphism will be those that confer its bearers differential susceptibility to infectious diseases* (for those show a probability of infection that is dependent on population size), while loci not associated with disease resistance that share susceptibility to parasitic infection by parasites coexisting with the host will tend toward a single neutral cloud in sequence space, through the extinction of all other genotypes. There is empirical support for this prediction: indeed, as noted previously (e.g. Ridley, 1995), some of the most polymorphic genes known are known to be associated with mechanisms to fight disease. The flip side of this coin is that genes responsible for mechanisms to fight disease are also associated with differential susceptibility to particular diseases, as shown previously by the P.I and colleagues (Marcos et al., 1994) and others.

A further application of the theory is in providing a *new functional definition of species* that applies equally well to clonally-reproducing creatures, one based on common susceptibility to parasitic infection.

2.2. Key R&D Goals and Project Milestones

FY 05: Demonstrating the Theory *In Computo:* We will seek to validate the theory by implementing simulations of hosts and parasites evolving in genetic space using genetic algorithms. In particular, we will:

I. Oct. 2004: Verify the assumptions of increasing mortality due to infectious diseases as a function of susceptible population size. Produce plots of outbreak frequency and mortality rate as a function of population size.
II. Nov. 2004: Demonstrate, using ecological models, whether stable co-existence equilibria exist and are limited to species with unshared vulnerabilities to parasites (model details omitted due to space constraints but are available upon request).
III. Dec. 2004: Test whether biodiversity is substantially reduced upon elimination of parasites.
IV. March 2005: Test whether freely evolving sexual and/or asexually reproducing species will cluster in species leaving gaps in the 'fossil record' upon exposure to parasites, and what the effect of removing the parasites is.
V. April 2005: Demonstrate application of the theory to morphogenesis and the creation of multicellular organisms.
VI. May 2005: Illustrate differences between parasitic and mutualistic (positive) interactions.
VII. September 2005: Show experimental support for our hypothesis that the genes that confer susceptibility to infectious disease are the most variable across a population.

FY 06: Modeling Markets as Networks (FY06/Q1&2): Our theory of shared vulnerabilities has many applications outside biology and ecology. We will initially focus on one application to economic markets. Building on the P.I.'s expertise in the study of biological, citation, software, and neurobiological networks, we will model economic markets as networks (graphs) and characterize their network properties (degree distributions, clustering, motifs, …). Individual public corporations will be modeled as nodes, with edges linking nodes with significant interaction. Interaction will be defined as a significant correlation in stock price fluctuations. To build networks out of stock price time series, we will follow an approach we and others have successfully applied to the inference of biological networks from gene expression data (Bhan et al., 2002; Bäcker and Sigman, unpublished results).

Shared Vulnerabilities and Stock Price Fluctuations (FY06/Q2&3)

1. Model market niches as communities with the market network, using the maximum flow network algorithm (Flake et al, 2000).
2. Compute the correlation between market size and probability of significant stock price fluctuation. We predict that stock in larger markets will be subject to more fluctuations than those in smaller niches, due to a greater number of shared vulnerabilities.

FY 07: Testing the Theory *In Vivo:* If our theory's predictions are born out by the models, we will collaborate with experimentalists to test the predictions for bacterial communities in the presence of viral parasites. Can we induce the extinction of a species in an ecosystem by introducing a shared parasite to replace exclusive parasites?

Submission of a SAND report and an external publication upon completion of each of the two stages, aiming for journals of general interest such as Science or Nature.

2.3. Risk and Likelihood of Success

Although the research proposed is of a fundamental nature and challenges long-held assumptions, risk has been significantly reduced by preliminary work done by PI over the past year, showing that the main ideas behind this proposal are correct. The main technical risk is that the parameter conditions under which the above ideas apply are too narrow to be of practical interest. However, prior work suggests this is unlikely. This is a case of all the pieces finally falling in place, integrating ideas from several fields. This provides added assurance that the individual ideas are correct.

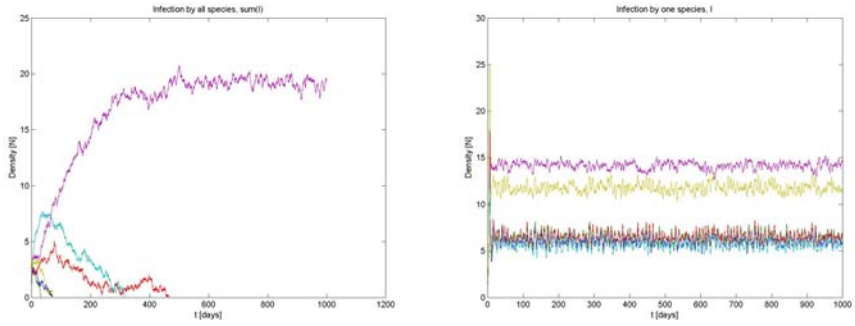
3. Relationship to Other Work

3.1. Previous Work

Faced with the astounding diversity of genotypes both within species and across them, evolutionary biologists developed neutral theory, which argues that most mutations are neutral, and are thus not acted upon by natural selection. Darwin, they argued, was not wrong; his theory is simply excluded from playing a part in the fate of most mutations. And yet the evidence suggests that most accumulated mutations are not in random directions, as would be expected by the neutral theory (Rieseberg et al., 2002).

3.2. Relationship to Other Ongoing Work

Recently, a solution to the coexistence problem was proposed (Huisman & Weissing, *Nature,* 1999*)* that relies on oscillatory dynamics. Beierholm and Adami at Caltech have shown that that solution is not robust to noise and thus unrealistic (unpublished results). Our model is, in contrast, extremely noise robust (Beierholm and Bäcker, unpublished). The left pane shows a single surviving species when vulnerabilities are shared; the right shows multiple species with differential susceptibilities coexisting robustly:



The field of complex real-world networks is growing rapidly. To our knowledge, the interdependencies of stock prices have not been used previously to model economic markets as networks.

3.3. Appropriateness of Approach

Our approach should succeed because it is grounded in real-life coexisting systems, is robust to noise and agrees with experimental observations better than alternative theories.

4. Importance

4.1. Relevance to Laboratory Missions

The theory that we propose to develop emphasizes interdependencies between elements of the infrastructure. Furthermore, it will develop predictions for how to optimize the allocation of the elements in a system of systems so as to optimize surety and minimize vulnerabilities. It will suggest specific recommendations for cyber surety. It will generate practical tools to predict the impact of disruptions on the U.S. economy. It will provide recommendations for the control of bioterrorist attacks and other emergent threats to our most critical infrastructure, our human resources. It will allow the growth and development of a young staff member. It will create competitively differentiating science in an area well differentiated from current laboratory investments. Most of all, we seek to carry out top-rate science with the potential to lead to revolutionary changes in our understanding of the evolution, ecology and survival of complex systems and to provide high visibility to the laboratory in the field.

Predicting the Vulnerability of Economic Infrastructure to Shocks: Our analyses will predict the likelihood of any given corporation being affected by a shock affecting another. Furthermore, statistical predictions on stock market fluctuations will allow the prediction of options pricing, which depend on the magnitude of fluctuations, with important economic applications.

Cybersecurity and Vulnerability of Critical Infrastructure: The fundamental conclusions of the theory are applicable to any complex system of systems as long as a restorative size-dependent force exists that favors rare variants over frequent ones (diseconomies of scale). In particular, we will test the prediction that even highly optimized critical infrastructure which shares vulnerabilities with more prevalent systems would be more vulnerable than custom-made systems even if these are not as optimized. This has obvious relevance to cybersurety in the age of quasi-monopolic highly prevalent operating systems.

Engineering of Microbial Communities: For the harvesting of energy &/or environmental cleanup at CI facilities, for example.

Containing the Spread of Infectious Diseases and Constitution of Teams Robust to Bioterrorist Attacks: Perhaps our most critical infrastructure is human capital. The most devastating loss for Native American people upon the colonization of the Americas by Europeans was not the destruction of temples or any material loss: it was the introduction of novel diseases that ravaged through the locals. Parasites used the socio-genetic networks of their hosts masterfully.

It is well known that susceptibility to disease varies greatly across populations, and yet the source of this variability remains to a large degree unknown. My theory makes the prediction that epidemics will spread preferentially among people with similar genetic composition, and furthermore, that the decimation of a particular allele will lead to a subsequent recovery of its frequency due to an associated decline in the spreading of diseases specific for that allele. At a time when we are on the verge of widespread genetic testing, this prediction is as testable as ever. If borne out, it could be of considerable help in stopping the spread of emerging threats, such as SARS or bioterrorist attacks, and preventing them from becoming endemic, by allowing efforts to be concentrated on individuals likely to acquire and transmit a disease.

From a personnel CI point of view, our findings would allow the constitution of genetically diverse teams, maximizing the probability that a teammember would survive in the case of a bioterrorist attack.

4.2. Programmatic benefit to IA, If Successful

Our proposal directly addresses the call for modeling and simulation of complex systems, and in particular for vulnerability and risk assessment of critical infrastructure. An improved understanding of how vulnerabilities shape survival of complex systems should directly transfer into the ability to create large systems of systems with complex interdependencies. For example, providing systems that share vulnerabilities with differential susceptibilities should protect them from extinction.

Potential customers: Potential customers for follow-up projects with external funding include NIH, for the verification of the prediction of the theory of genes' role in fighting parasites, ONR and DOE's Genomes To Life Project, whose 3rd goal specifically aims for the engineering of microbial communities.