The Ecology of Echo

Peter T. Hraber
Dept. of Biology
University of New Mexico
Albuquerque NM 87131
pth@santafe.edu

Terry Jones
Santa Fe Institute
1399 Hyde Park Road
Santa Fe NM 87501
terry@santafe.edu

Stephanie Forrest*
Dept. of Computer Science
University of New Mexico
Albuquerque NM 87131
forrest@cs.unm.edu

To appear in *Artificial Life*
December, 1996

Echo is a generic ecosystem model in which evolving agents are situated in a resource-limited environment. The Echo model is described, and the behavior of Echo is evaluated on two well studied measures of ecological diversity: relative species abundance and the species-area scaling relation. In simulation experiments, these measures are used to compare the behavior of Echo with that of a neutral model, in which selection on agent genotypes is random. These simulations show that the evolutionary component of Echo makes a significant contribution to its behavior and that Echo shows good qualitative agreement with naturally occurring species abundance distributions and species-area scaling relations.

*Corresponding Author.

1 Introduction

Many interesting systems are difficult to analyze and control using traditional methods. These include natural ecosystems, immune systems, cognitive systems, economies and other social organizations, and arguably modern computers. One source of difficulty arises from nonlinear interactions among system components. Nonlinearities can lead to unanticipated emergent behaviors, a phenomenon that has been well documented and studied in physical, chemical, biological, and social systems (e.g., see [7, 29]), as well as in some forms of computation [13]. Nonlinear systems with interesting emergent behavior are often referred to as *complex systems*. A second form of complexity arises when the primitive components of the system can change their specification, or evolve, over time. Systems with this additional property are sometimes called *complex adaptive systems* [20]. Here, we will use the term *complex adaptive system* (CAS) to refer to a system with the following properties: a collection of primitive components, called "agents"; nonlinear interactions among agents and between agents and their environment; unanticipated global behaviors that result from the interactions; agents that adapt their behavior to other agents and environmental constraints, causing system components and behaviors to evolve over time.

It can be quite difficult to model systems with these properties analytically. Useful and predictive mathematical treatments are difficult, due primarily to the following: nonlinearities (as discussed earlier), discreteness, spatial inhomogeneities, and the changing behavior of the primitive elements of the system. Discreteness arises, e.g., in time (as in the case of generations in population genetics), state spaces, and internal variable values, whereas spatial heterogeneity arises through resource gradients, nonuniform operating conditions (e.g., different mutation rates affecting different parts of the body [26]), or even random drift. Both discreteness and spatial heterogeneity can have a significant effect on system behavior. For example, Durrett and Levin describe how both of these attributes affect the predictive power of conventional models, including ordinary differential equations and reaction-diffusion systems, on a classical problem in ecology [8]. As a consequence, many of the standard approximations for infinite-sized systems and techniques developed for studying asymptotic behavior of continuous nonlinear dynamical systems cannot be directly applied to discrete or spatially heterogeneous systems. Finally, adaptation is central in CAS. Because the primitive components of the system can change over time, agents behave according to different rules at different times. Individual variants are important and can determine the overall system trajectory, which precludes modeling only their aggregate behavior. Although the underlying evolutionary mechanisms themselves can in principle be modeled, as suggested in [12], this is a difficult undertaking and this level of detail is omitted in most ecological models.

An alternative to analytical models of CAS is simulation. Detailed simulations are also problematic, because it is often impossible to get all of the details correct. Consider, for example, the vertebrate immune system which has been estimated to express over 10^7 different receptors simultaneously [27]. Modeling the physical chemistry of just one receptor/ligand binding event, even at an abstract level, requires enormous amounts of computation, and it is therefore infeasible to model the total expressed repertoire of receptors explicitly and precisely. This problem exists for many large complicated systems, but because nonlinear systems can be highly dependent on seemingly small details, even a trivial inaccuracy in the model could lead to wildly erroneous results. One approach, commonly used in artificial life models, is to strip away as much detail as possible, retaining only the essential interactions.

This paper describes one such model, called Echo, whose emphasis is on the essential components and interactions of ecosystems. Many CAS can be viewed as ecologies (e.g., [24, 46]), but our focus in this paper is on the analogy with natural ecologies. Our goal for Echo is to capture in an artificial ecosystem a set of behaviors that are robust with respect to the details of the interactions (e.g., avoiding parameter tweaking to coax a system to produce desired behaviors) and which produce the broad qualitative categories of ecological phenomena in which we are interested.

Echo is a mechanistic model in the sense that it encodes (as a computer program) a theory about which mechanisms are most relevant in ecosystems. In Echo, certain primitive components and interactions are built in, and when the model is "run," i.e., simulated, these mechanisms give rise to various macro-level properties. Our goal is that ecologically significant behaviors will arise spontaneously as a consequence of the primitive mechanisms. This is quite a different kind of explanation than simply predicting what will happen next without representing the underlying mechanisms explicitly. Of course, prediction is important for models like Echo, but it is quite a different kind of prediction than that typically associated with simulations. In this paper, we use the word "model" to describe the design decisions we have made about which components and interactions are included and which are not. This is potentially confusing because the model is described in terms of computational structures, like "agent," "stack," and "rules." Our model is really a high-level simulation of generic ecological behavior.

In Echo, an agent replicates (makes a copy of itself, possibly with mutation) when it has acquired enough resources to copy its genome. The local state of an agent is exactly the amount of these resources it has stored and its location in the world. Agents acquire resources through interactions with other agents (combat or trade) or from the environment. This mechanism for endogenous reproduction is much closer to the way fitness is assessed in natural settings than conventional fitness functions in genetic algorithms.

Echo specifies certain structural features of the environment in which agents evolve. There is a two-dimensional grid of "sites" and each agent is located at a site, although it is possible for agents to move between sites. There are usually many agents at one site, and there is a linear neighborhood within a site.[1] Each site produces renewable resources. Resources are represented by different letters of the alphabet, and genomes are constructed from the same letters. The number of resources in an Echo world is typically small. These are denoted by lower-case letters: a, b, c, and so forth. Resources can exist as part of an

[1] This definition of locality within a site will likely be changed in future Echo implementations.

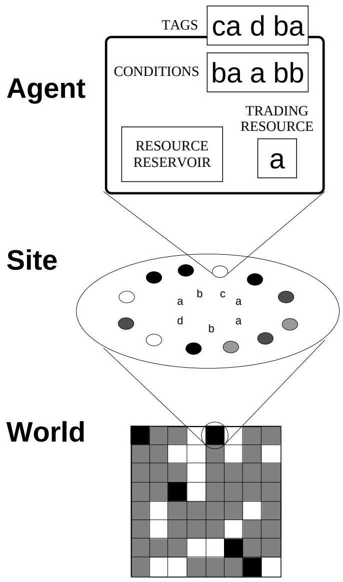

Figure 1: Structure of the Echo world. Agents consist of a genome (tags, conditions, and trading resource) and a resource reservoir. Genomes are made of resources (letters) which code for interactions. Agents store excess resources in their reservoir and reproduce when they have enough excess to copy their genome. A site is a population of agents, with resources added at a specified rate. Agents within a site are organized in a linear array (shown as a ring). A world is generally comprised of multiple sites, although some studies use a world with only one site. Under certain circumstances, agents can migrate to neighboring sites.

agent's genome, as part of an agent's local state (in its reservoir), or free in the environment. Figure 1 gives an overview of the structural aspects of the Echo model.

There are three forms of interactions among agents: trade, combat, and mating. In trade, resources stored in the reservoir are exchanged; in combat, all resources (both genetic and stored) are transferred from loser to winner (as in a predator-prey interaction); in mating, genetic material is exchanged through crossover, thus creating hybrids and providing a primitive form of sexual reproduction. Mating, together with mutation during the replication process, provides the mechanism for new types of agents to evolve. Resource constraints provide the selective pressure for agents to develop alternative strategies for survival and reproduction.

In each Echo run there is a fixed number of resource types, which is determined by the user of the system. These might represent of resources in a real-world system, or they might correspond to a more abstract notion of something that is required to ensure survival. For example, an experiment could be designed to require that agents possess a certain resource, which some agents can only obtain through trade. In this situation, the resource need not be thought of as corresponding to a physical entity, but as representing some type of agent-agent interaction that is necessary for agent survival.

Echo is related to several earlier CAS models. Genetic algorithms focus on the evolutionary component of CAS [17]. They are reasonably well understood and mature, but ignore several important features, including resource allocation, heterogeneity, and endogenous fitness. Classifier systems [21] apply genetic algorithms to a cognitive modeling framework. Similarly, Echo extends genetic algorithms to an ecological setting, adding the concepts of geography (location), competition for resources, and interactions among individuals (coevolution). Echo is intended to capture important generic properties of ecological systems, and not necessarily to model any particular ecology in detail. Echo's contribution to ecological modeling lies in the fact that evolution is built in as a fundamental component of the system. Few existing models consider both ecological interactions and evolutionary dynamics, or the potential for open-ended evolution of ecological function in communities comprised of many interacting species [31, 36, 5, 6, 9], a recent exception being [15]. Without a model that incorporates evolution, it is difficult to address topics such as flows of information and resources in many-species assemblages, patterns of speciation and extinction in ecological communities, and effects of spatial heterogeneity on population dynamics.

Echo also resembles more recent CAS computational systems. These include Swarm [28], Sugarscape [11], Tierra [46], and the Evolutionary Reinforcement Learning (ERL) model [1]. Echo describes a family of models, but it is not a generic modeling platform like Swarm. Swarm supports a wide range of agent types and interaction rules, but Echo makes specific commitments about the form of system components (agents, resources, interactions). Echo does resemble Sugarscape in several respects, but it differs in specific details (e.g., the complexity of each individual agent) and its overall goals. Echo is intended as a general CAS model, but Tierra and ERL are much more specific systems. Like Tierra, Echo is resource-driven, but the context is quite different. In Tierra, the limiting resources are CPU time and memory space, agents consist of short sequences of computer instructions, and the behavior of interest is how quickly these sequences replicate themselves. Finally, ERL provides two levels of adaptation, compared with one in Echo.

The original idea for Echo, including motivation, design decisions, and overall structure was introduced in [17, 19, 20]. In that work, a family of progressively more elaborate Echo models is outlined. Our goal in this paper is to describe more fully one specific Echo model and ask how well it compares with natural ecological systems. Section 2 describes the Echo model we have implemented. An important aspect of our study is the idea of a neutral model. We propose a neutral model that allows us to assess the effect of evolutionary pressures on Echo's behavior, and we compare its behavior with that of our original Echo implementation. Sections 3 and 4 report experiments (with and without the neutral model) on the species abundance relations in Echo, a characteristic feature of ecological systems. This raises some fundamental questions, such as what are the operative mechanisms producing canonical distributions and what scientific purpose is served by such models, which we also discuss.

2 An Overview of Echo

Echo consists of a collection of *agents* distributed across a two-dimensional array of *sites*, resources (represented as letters), and various kinds of interactions, both between pairs of agents and between agents and the environment. The following subsections describe Echo in more detail. Much of this is devoted to describing agents and the interactions that can occur.

2.1 Echo Agents

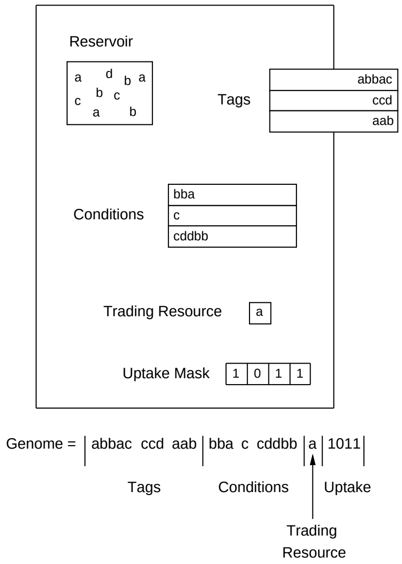

Figure 2: The structure of an Echo agent. Each agent consists of a genome and a resource reservoir. The genome has $r + 7$ genes, where r is the number of resources in the world. Six of these, the *tags* and *conditions* are composed of variable-length strings of resources (i.e., of the lower-case letters that represent resources). Tags are visible to other agents. Conditions and other properties are not.

Agents have a genome that is roughly analogous to a single chromosome in a haploid species, as shown in Figure 2. Tags are genes that produce some externally visible feature of the phenotype, e.g., color. Conditions are genes that cannot be directly observed by other agents. They encode internal preferences, behavioral rules, and so forth. Thus, one agent will interact with another on the basis of its own internal conditions (rules for interaction) and the other's external tags (appearance). This allows the possibility of sophisticated interactions between agents, including mimicry, bluffing, other forms of deception, and some intransitivities. For example, in mimicry, an agent can appear dangerous but actually be unwilling to fight. An example of intransitive combat relationships would be the following: an agent A might attack an agent B, and B attack C, but it does not follow that A will attack C. This has obvious parallels in real-world systems, e.g., in food webs. See [17, 18, 20] for discussions of the significance of these kinds of interactions.

The six (external) tag and (internal) condition genes possessed by every agent are the *offense tag*, *defense tag*, *mating tag*, *combat condition*, *trade condition* and *mating condition*. These genes are used to determine what sort of interaction will take place between a pair of agents, and what the outcome will be. The use of these genes is described below. The current implementation conforms largely to the early description in [17], but not to those in [19, 20].

The r genes correspond to the agent's *uptake mask*, which determines its ability to collect each resource type directly from the environment. If an agent does not have a "1" allele for the uptake gene corresponding to a certain resource, it will not be able to collect that resource directly from the environment. Consequently, if the agent requires this resource (because the site at which it is located charges a tax that includes it, or because the agent needs it to replicate), it will either have to fight or trade for it. The designer of an Echo world can create trading webs among agents by requiring them to trade in various ways to ensure survival, i.e., by seeding the system with agents whose genotypes enforce this constraint (see Figure 5 for an example). These webs can be quickly altered through mutation, crossover and migration.[2]

The final gene is the *trading resource* which is the resource type that the agent will provide to another agent if trading takes place. An agent may trade with another agent even if it does not possess any of its trading resource to exchange. In this case, the other agent in the interaction receives nothing. Agents also have a *reservoir*, or stomach, in which to store resources. Resources from the reservoir are used to pay taxes, to produce offspring, and for trade.

2.2 Agent Replication

An agent is replicated when it acquires sufficient resources. The agent's genome is copied using the resources it has stored in its reservoir to construct the daughter agent. There are two parameters that control the replication process:

- The *self-reproduction threshold* is used to determine when enough resources have been collected. As an example, if the genome contains three as and the self-reproduction threshold is two, then the reservoir must contain at least six as before self-reproduction can take place. A similar requirement must be satisfied for each resource type.
- The *self-reproduction fraction* determines the division of excess resources between the original agent and the daughter agent after replication. The fraction indicates how

[2] We expect that webs might be more stable if each agent were restricted to collecting only one resource type directly from the environment. Future versions of Echo are likely to reflect this change.

the proportion of extra resources (i.e., those not needed to make the clone) should be divided between the two. A setting of one will result in offspring with no internal reserve of resources. At the other extreme, a setting of zero allocates all extra resources to the offspring, leaving the original vulnerable.

Both of the above parameters are defined as constants for the entire Echo world. Thus, there is no variation among agents in life-history strategy. This means there is no possibility in the current model for selection to operate on tradeoffs between size and number of offspring, or to relative investments in growth versus reproduction, or to the timing of reproduction versus dormancy. In our experiments, we chose the parameters in such a way that Echo agents would reproduce by giving the daughter agent half of the parent's available resources, as soon as there are enough resources to do so. In this sense, the reproductive strategy we studied is more akin to that of bacteria than of higher-order (multicellular) organisms.

During replication, there may be spontaneous mutations; their frequency is controlled by the *mutation probability*. Mutations in different parts of the genome are treated differently:

- If a mutation occurs in the trading-resource gene, the agent's trading resource is set to a randomly selected resource (including the current resource).
- If a mutation occurs in the uptake mask (see Section 2.1), its bits are flipped independently, each according to this probability.
- If a mutation occurs in one of the tags or conditions, it occurs in the following way. If there are r resource types and the chromosome is of length n, a number is selected uniformly at random from the interval $[1..n+1]$. If this number lies in the interval $[1..n-1]$ the corresponding allele is set to a randomly chosen value (resource type). This case is analogous to point mutation in conventional genetic algorithms. If the number is $n+1$, a symbol is added to the end of the chromosome and assigned a random value (an insertion). Finally, if the number is n, a number is selected uniformly at random from the interval $[1..r+1]$. If this second number is $r+1$, the last symbol is deleted from the chromosome. Otherwise, the last symbol in the gene is set to the corresponding resource (a point mutation) Consequently, gene length can change through additions or deletions at the last position in the gene. These possibilities are illustrated in Figure 4.

We note that the mutation operator makes it possible to alter the ecological function of an agent's genome by a single mutation event. Thus, the mutation rate might seem to have too great an influence on the relationship between ecological and evolutionary dynamics. In natural living systems, a mutation in a protein-coding region of the genome is likely to be evolutionarily neutral or produce little effect on the individual's phenotype. However, a mutation in a gene regulatory region can have a great effect on how much of the regulated gene is expressed, and can greatly effect the phenotype. We consider the mutation operator in Echo to be more representative of the latter type of mutation, which seems to be the most appropriate abstraction for the level of detail at which the agent genome is modeled.

2.3 Agent Interactions

There are three direct forms of agent-agent interaction: combat, trading and mating. All of these interactions take place between agents that are located at the same site and all involve the transfer of resources between agents. Tests for interactions are always conducted in the same order: first for combat, then for trade, and finally for mating. The order of interactions constrains community composition in a realistic way. Testing for a competitive or predator-prey interaction prior to testing for mutualistic interactions approximates the behavior of natural organisms.

Combat

Combat is an idealization of any antagonistic interaction between real-world entities. If two agents in a real-world system are behaving in a competitive fashion, this would be modeled in Echo by designing the agents that engage in combat. When combat occurs, one agent is killed (and its resources are placed in the reservoir of the survivor), [3] unless it flees first.

When two agents meet, it is first determined whether or not either agent will attack the other. An agent A will attack an agent B if A's combat condition is a prefix of B's offense tag. If attacked, an agent is given a chance to flee (which it does with a probability equivalent to the probability of it losing in the combat encounter). The calculation of the probability of victory in combat is based on matching A's offense tag with B's defense tag and vice versa. The resource characters that comprise these genes are used as an index into the *combat matrix*, with special provisions for zero-length genes and for genes of unequal length. This is somewhat complicated and is not described fully here (see [20] for the detailed description).

As a result of this computation, each agent receives some number of points. If α and β are the points awarded to A and B, then A will win the combat with a probability of $\alpha/(\alpha+\beta)$. The resources that comprise the loser (both its genome and the contents of its reservoir) are given to the winner, and the loser is removed from the population.

Trade

If two agents are chosen to interact and they do not engage in combat, they are given the opportunity to trade and mate. Unlike combat, trading and mating must be by mutual agreement. Agents A and B will trade if A's trading condition is a prefix of B's offense tag and vice versa. Notice that the offense tag is used here as well as in determining whether combat will occur.

When trade takes place, each agent contributes its excess trading resource. Excess is defined to be the amount of resource that an agent possesses above that required to replicate

[3] In more recent formulations of Echo [19, 20], the interaction is not always so extreme and results in resources being transferred (possibly in both directions, and possibly in a very uneven fashion) between the agents.

its genome, plus some reserves (world-level parameters control how much reserve an agent retains). Thus, an agent provides some fraction of the resource that it does not need for the next self-reproduction. This may be zero, in which case an agent does not provide anything in the trade. This behavior is analogous to a form of deception or bluffing. An agent cannot know in advance if another agent will supply a positive quantity of a resource, or what that resource might be. This might seem to be an odd form of trade, but agents can learn to recognize each other based on their trading tags. Agents whose tags involve them in disadvantageous trades will reproduce less quickly and have smaller probabilities of being able to meet taxation demands.

Mating

Agents that interact and do not engage in combat may exchange genetic information through recombination. As in many genetic algorithms, the new agents replace their parents in the population. Recombination occurs between two agents A and B if A finds B acceptable and vice versa. A will find B acceptable if (1) A's mating condition is a non-zero prefix of B's mating tag or (2) both A's mating condition and B's mating tag are zero length. The restriction to nonzero prefixes is designed to prevent agents with zero-length mating conditions from proliferating rapidly. Such an agent finds all other agents desirable (including copies of itself). To prevent this, an agent with a zero-length mating condition will only find an agent with a zero-length mating tag acceptable. This is a slight departure from the description of mating given in[20]. Figure 3 shows a simplified view of the two-way matching process used to determine whether mating will occur.

When mating does occur, a form of two-point crossover is employed. This is complicated by the fact that genomes are variable length; one can choose a crossover point in one agent but find that the same crossover point does not exist in the other. Briefly, crossover proceeds by (1) selecting two genes to contain crossover points, (2) choosing crossover points in each gene in each agent, and (3) crossing over in the manner of two-point crossover. The operation conserves resources (i.e., resources are not created or destroyed) but the ratio of genetic material from each parent in each of the children is not guaranteed to be 50:50. Mating, together with mutation during the replication process, provides the mechanism for new types of agents to evolve, as shown in Figure 4.

2.4 The Echo Cycle

There is a fixed sequence of events that takes place in a single Echo cycle, which is described below. This description mentions several system parameters, which are defined more carefully in Appendix A.

1. Interactions between agents occur at each site. These include trade, mating, and combat. The proportion of agents that interact on each time step is controlled by a world-level parameter, the interaction fraction. For each interaction, one agent is

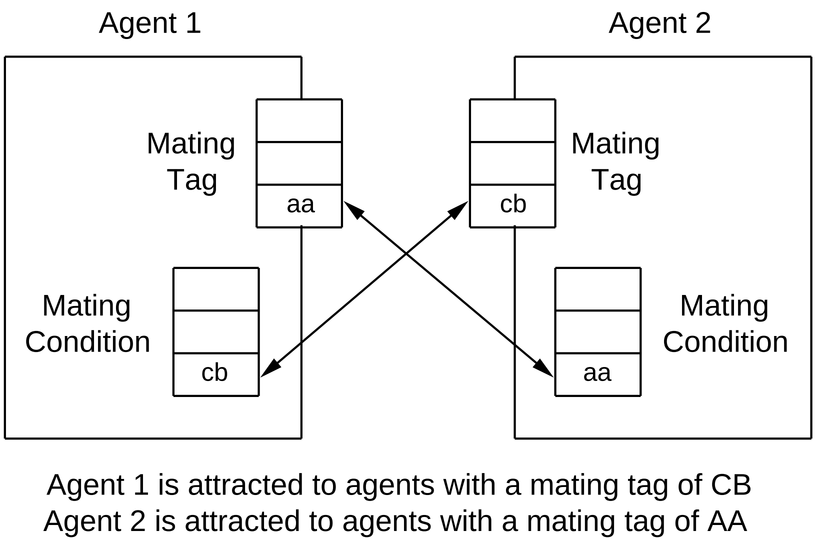

Figure 3: A simplified view of the two-way tag and condition matching that is used by agents to determine whether mating will occur.

selected randomly. A second agent is then selected in the vicinity of the first. The first agent is moved adjacent to the second in the one-dimensional array of agents at the site. If the first agent attacks the second, the second can run away by moving a small distance in the array. In both cases, distances are likely to be small.

2. Agents collect resources from the site if any are available. The site produces resources according to its site parameters, and these are distributed as equally as possible among the agents at the site that are genetically able to collect them.

3. Each agent at each site is taxed (probabilistically). Each site exacts a resource tax from each agent with a given (worldwide) probability. If an agent does not possess the resources to pay the tax, it is deleted (killed) and its resources are returned to the environment. Tax in Echo can be thought of as economic taxation, or as the cost required to live at the site. Biologically, this can be thought of as metabolic cost.

4. Agents are killed at random with some small probability. This can be interpreted either as bad luck or as a mechanism that prevents agents from living indefinitely. Agents that are not killed some other way (through combat or taxation), will eventually be randomly killed. When an agent at a site dies, for whatever reason, its resources are returned to the environment, thereby becoming available to other agents at that site.

1. Mutation during self-replication.

Point mutation	AAA	$\longrightarrow$	ABA
Deletion	ACC	$\longrightarrow$	AC
Insertion	BB	$\longrightarrow$	BBE

2. Genetic change via crossover.

ABC	ADBB	BA	$\longrightarrow$	BBC	ADBB	AA
BBC	DACC	AA	$\longrightarrow$	ABC	DACC	BA

Figure 4: Operations introducing genetic change in Echo agents.

5. The sites produce resources. Different sites may produce different amounts of each resource. For example, one site might produce ten a's and ten b's on each time step, whereas another might produce five b's and twenty c's. The idea is that agents will replicate frequently if they are located at sites whose resources match their genomes, if the site is not too crowded. When an agent at a site dies, its resources are returned to the environment and become immediately available to other agents at that site.

6. Agents that do not acquire any resources during an Echo cycle (either through picking them up or through combat or trade) migrate to a neighboring site. The neighboring site is selected at random from among those permitted by the geography of the world.

7. Agents that can replicate do so (asexual reproduction). An agent may replicate when it acquires sufficient resources. In replication, an agent makes a copy of its genome using the resources it has stored in its reservoir. A parameter controls how many resources are required to be stored beyond those needed to make an exact copy. The replication process is noisy: random mutations may result in genetic differences between parent and child.

This cycle is iterated many times during the course of a run.

Each run of Echo involves a *world* that contains a fixed number of *sites* (see Figure 1). Each site can contain an arbitrary number of *agents*, including zero. These components are designed by the user, typically as an abstraction of some aspect of a real-world CAS. In the Echo world used in this paper, there are four resources and one site.

The design of an Echo world requires the choice of a geography of sites. In the current implementation, this geography must form a rectangular array, although the general class of Echo models makes no such restriction. Various properties of sites must be specified by the user, described in Appendix A. Table 1 shows parameter values for the experiments described in later sections.

An Echo world also requires specification of the *combat matrix*. The matrix is used in the calculation of combat outcome probabilities.

An Echo site initially contains an arbitrary number of agents. These are arranged in a one-dimensional array. The probability that a pair of agents will be chosen to interact falls off exponentially with increasing distance between agents in the array. The user determines which agents initially reside at each site, and in what order they should appear in the array. There are three parameters for each site, shown in Appendix A.

3 Ecological Considerations

Does the model described in the previous section behave similarly to natural ecosystems? There are many relevant behaviors that we might consider, some of which are listed below:

- Stable patterns of interactions among genetic variants, or trophic networks, such as the ant-fly-caterpillar triangle described in [20, 22], or the trophic cascade described in [47].

- Different levels of resources, i.e., what happens when the amount of resources pumped into the system at each time step is increased or decreased?

- Relative species abundance (trajectories, duration, distributions).

- Community assembly patterns (see, e.g., [23]).

- Cataclysmic events such as meteors, droughts, etc.

- Isolation effects—what happens when a number of sites are allowed to evolve independently (with no migration) and then the sites are connected (i.e., migration is enabled).

- Flows of resources—are there resource sources and sinks? How quickly do resources flow through the system? With what are rates of flows correlated?

- Transitions from single-cellular to multi-cellular organization (this last question requires extensions to the Echo model described in Section 2).

We have studied several of these questions informally by observing running Echo models. In these studies, we start Echo either with a population of randomly constructed agents or (more typically) with a fixed collection of agents (e.g., that implement a trophic triangle).

Beyond these informal explorations, one earlier study used a modified form of Echo, called Gecko, to study the population dynamics of a tri-trophic food chain [47]. The goal of this work was to identify which biological features lead to stability of artificial food chains under various conditions. For the simulations, the adaptive operators (mutation and crossover) were disabled. This allowed the experiments to isolate the effects of "ecological" dynamics from the effects of evolution. The experiments showed that the three species could persist only when heterotrophs (herbivores and carnivores) could select for high-quality (nutritious) prey (this behavior in natural predator-prey relations has been named "prudent predation"[48]), and when adjustments were made to the number of interactions per species per time step. That is, all species could persist when herbivores were allowed many opportunities to attack plants per iteration, and carnivores were allowed fewer chances to attack herbivores.

In the remainder of this paper, we describe a second case study, one in which the adaptive operators are enabled and in which we consider many species communities. Rather than explicitly limiting our populations to three species and testing for ecological realism in the limited assemblage, we take a different approach to verifying the Echo model. We start with three genetic variants (species), and allow their genotypes to mutate and recombine. In our study, corroboration will come not from the persistence of a three-species community, but from an increase in diversity with increasing habitat availability. Specifically, we study species abundance patterns, with a provisional definition of species as a unique agent genotype.

Species diversity in ecological communities is arguably the single most interesting ecological variable [25, 32, 33, 3]. One reason for this interest is the overwhelming number of species found on Earth; approximately 30 million species live today, and a great many more that have lived and suffered extinction [4, 44]. A second reason is that species diversity is believed to play a role in the stability of ecosystems to perturbations or invasions (see [39] for a review). Ultimately, the science of ecology seeks means for predicting species diversity, by identifying those mechanisms that regulate it, for example [3]:

> One of the greatest remaining challenges in biology is to explain the diversity of living things. What determines the number and kinds of animals, plants and microbes that live together in one place? What causes variation in diversity from one place to another? What accounts for changes in the abundance and identity of species over time? How do individual species contribute to the diversity and stability of the natural world? (p. 877)

A related property of ecological communities is the relative abundances of species. Consider a sample of vegetation from a semi-arid desert or tropical forest. What is the distribution of individuals into species? Many factors will contribute to this distribution. These include ecological factors (e.g., the presence or absence of various predators or mutualists, and relative population sizes and distances), environmental factors (e.g., temperature, availability of precipitation, and soil texture and quality), and experimental factors (e.g., size of sample, scale of observer, and location of the sample). Ecologists have performed such sampling countless times from many biota. One purpose of such sampling is to describe the way N individuals present are partitioned into S species, and in particular, to determine whether the distribution of species fits a mathematically distinct pattern. If it does, then it is relevant to ask what biological processes could produce this pattern. Such analyses help to identify processes that regulate species diversity.

The experiments described in the following two sections show how species abundance patterns can be used to test whether or not Echo exhibits ecologically plausible behavior. We use two methods to test the hypothesis that diversity patterns in Echo resemble those of natural systems: comparison of Echo with a neutral model and comparison with quantitative and qualitative expectations. Because evolution of agent genotypes is the mechanism of principle interest in Echo, we define a neutral model that removes direct sources of selective pressure and compare its behavior with that of the original Echo model. To make a falsifiable comparison with biological systems, we consider two ecological patterns, each of which has similar quantitative properties in many different ecosystems—the Preston curve and the species-area scaling relation. This allows us to compare Echo diversity patterns with empirical patterns of species abundance. In the case of the Preston curve, careful quantitative comparisons are problematic for several reasons, so we report detailed results only for the species-area scaling relation.

3.1 The Preston Distribution

When sampling experiments are performed on natural ecosystems, a common result is that most species sampled have few representatives. That is, most species are rare, but a few are common [41, 32, 3, 4]. Preston's canonical log-normal distribution is the most widely accepted formalization of the relative commonness and rarity of species, but see [38] for a discussion of alternative schemes. When species counts are plotted on linear axes, the distribution is unimodal and right-skewed with an extremely long tail. Because it is impossible to observe less than a single individual representing a species, these distributions are truncated on the left at what Preston called the "veil line." When plotted on a log-transformed x-axis, the distribution above the veil line closely resembles a normal distribution. Preston observed this and postulated that species counts follow log-normal distributions.

To construct these curves, Preston counted abundances of sampled species and grouped them into a series of "octaves" [40]. An octave is simply a logarithmic (base 2) grouping of species abundances. His octaves were labeled "< 1", "1–2", "2–4", and so forth. Preston plotted octaves on the x-axis and the number of species per octave, a frequency of frequencies, on the y-axis. If a species abundance fell within octave boundaries, it counted as one species in that octave. If an abundance fell on the boundary between octaves, (as will any value which is a power of 2), one-half was counted in the two neighboring octaves.

Preston constructed these curves for a number of biota, and found their general shape was well approximated by a Gaussian (normal) distribution of the form

$$y = y_0 e^{-(aR)^2}$$

where y is the number of species falling into the R^{th} octave left or right of the modal octave, y_0 is the value of the mode of the distribution and a is a constant, related to the logarithmic standard deviation (σ) such that $a = (2\sigma)^{-\frac{1}{2}}$ [40, 30]. In particular, Preston found that the value of a tended to be in the vicinity of 0.2. This observation gave rise to the "canonical" log-normal distribution [41, 42]. In the canonical distribution, the general log-normal distribution is reduced to a family of log-normal distributions with similar variance. This relationship makes it possible to predict relative species abundances given only the number of individuals or the number of species [41, 50].

3.2 The Species-Area Scaling Relation

As more individuals are collected, more species will be found. Preston observed that doubling the number of individuals sampled will "unveil" another octave in the species curve by moving all species counts one bin to the right [40, 41]. One way to collect more individuals is to increase the area of the sample. This can be achieved either by taking successively larger sample sizes, or by sampling from similar habitats of increasing area. Conveniently, islands in an archipelago provide ecologists with natural sampling units of varied size, all with similar climate and terrain.

When species richness is tallied for islands of increasing area, the following scaling relation holds: $S = cA^z$, where S is the total number of species, A is the area of the habitat (e.g., an island or mountain top), and c and z are regression constants. For empirical ecological communities with a canonical log-normal distribution of abundances, $z \approx 1/4$. This relation and the value for z can be predicted from the Preston distribution [41, 32, 38] and is well supported empirically [41, 30, 10].

The species-area scaling relation is an important ecological pattern. The theory of island biogeography is predicated on species-area relations [30]. In conservation biology, this relation has been used to predict the effects of different reserve sizes on species diversity. Although the species-area relation can be derived from Preston's canonical log-normal distribution, a satisfactory explanation of the processes that regulate this relation has not been advanced [10]. Exceptions to the species-area relation can be found (see, e.g., [43]), which help identify mechanisms that constrain the species-area curve. For instance, a value for z larger than $1/4$ is typically found when sampling at the continental scale. This is taken to reflect the effect of greater habitat heterogeneity represented by vast regions of complex terrain [30].

4 Echo Experiments

The Preston curve and species-area scaling relation are both robust ecological phenomena. This section describes experiments that ask whether Echo exhibits these phenomena, both for the original and the neutral versions of Echo. An earlier study examined more carefully whether Echo populations resemble Preston distributions [14]. Thus, the species-area curve

Parameter	Value
Number of Resources	4
Trading Fraction	0.5
Interaction Fraction	0.02
Self Replication Fraction	0.5
Self Replication Threshold	2
Taxation Probability	0.1
Mutation Rate	0.02
Crossover Probability	0.07
Random Death	0.0001
Initial Vector	10 10 10 10
Production Vector	10 10 10 10
Maximum Vector	100 100 100 100
Maintenance Vector	1 1 1 1

Table 1: Parameters used for Echo experiments described in text. Parameters above the line are world-level parameters; those below are site-specific. See Appendix A for details.

is our primary focus for formal quantitative comparison, although we include some data on Preston distributions.

Our experiments were designed to test two things: (1) the effect of evolution (by allowing the production of new genotypes) on Echo's behavior, and (2) how closely Echo exhibits canonical Preston distributions and species-area scaling relations. We consider these experiments to be one step in testing whether Echo exhibits ecologically plausible dynamics. To test whether observed diversity patterns in Echo result from random processes or from evolutionary pressures (question 1), we compare the behavior of the original Echo model with that of a neutral model that replaces Echo's interaction rules with random interactions between agents.

4.1 Simulation Parameters

The parameters used in our Echo experiments are shown in Table 1. We vary the area by defining worlds with multiple identical sites. Agents which fail to acquire resources during a time step are allowed to migrate to one of the site's four neighbors to the north, south, east or west. Periodic boundary conditions are imposed on worlds comprised of 1, 4, 16, or 64 sites. Thus, agents live in toroidal worlds of fixed size; this is roughly analogous to living on an island of given area. Independent simulations were run for both 10^3 and 10^4 iterations.

We have run experiments with a variety of initial conditions. The two most common setups are as follows: (1) start with a single founding agent that is one mutation away from trading with itself, fighting with itself, and mating with itself, and (2) start with a

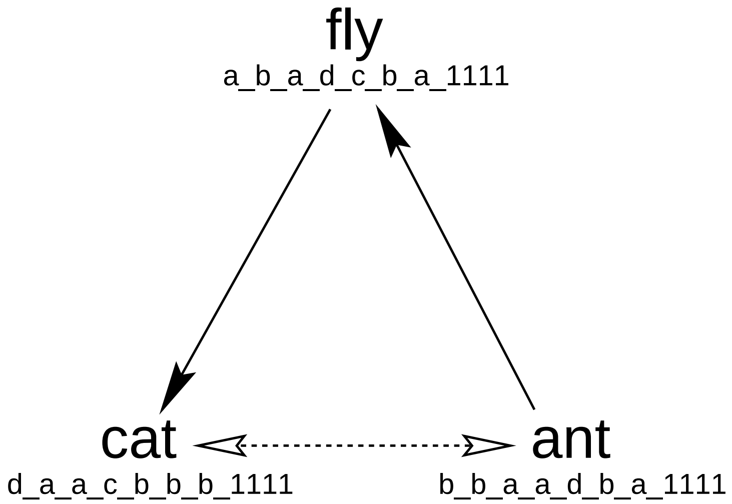

Figure 5: Echo agents initially present in experiments described in this section. Interactions among ants, caterpillars ("cat") and parasitic flies are coded by agent genotypes. Open arrows indicate mutualistic trade between ants and caterpillars. Closed arrows indicate direction of an attack in combat (flies attack caterpillars; ants attack flies). The genotype of each agent is indicated with loci in the following order: offense tag, defense tag, mating tag, combat condition, trading condition, mating condition, trading resource, and uptake mask.

small set of hand-designed agent types that encode some interaction pattern of interest. The experiments reported here all use the second setup, consisting of three agent types that form a trophic triangle. However, we get similar results when the first style of initial condition is used (data not shown).

Figure 5 illustrates the initial conditions used for our experiments. This ant, caterpillar, fly interaction network was described by Holland [19] and is based on observations by Hölldobler and Wilson [22]. The caterpillars secrete a nectar which is collected by the ants. In exchange, ants nurture the caterpillars. Parasitic flies attempt to consume caterpillar larvae, while the ants defend their symbionts. Initially, each site in a world has ten agents of each genotype present, with ten letters of each resource type in reserve. No mating occurs initially, although mutation could produce such an interaction during simulation.

4.2 The Neutral Model

Neutral models are used in biology to evaluate which mechanisms are necessary to produce a phenomenon of interest [35]. Nonrandom mating, competition or predation, and direct or indirect mutualisms are all believed to regulate species diversity in ecological communities. The question for both Echo and the real world is whether such mechanisms are

	Combat	Trade	Mating
	$P_{combat} \cdot P_{escape}$	$(1 - P_{combat}) \cdot P_{trade}$	$(1 - P_{combat}) \cdot P_{mating}$
Neutral model	$1/4 \cdot 1/2$	$(1 - 1/4) \cdot 1/8$	$(1 - 1/4) \cdot 1/8$
Original model	$1/r^n \cdot P_{escape}$	$(1 - P_{combat})(1/r^n)^2$	$(1 - P_{combat})(1/r^n)^2$
n=0	$1/1 \cdot 1/2$	$(1 - 1)(1/1)^2$	$(1 - 1)(1/1)^2$
n=1	$1/4 \cdot 1/2$	$(1 - 1/4)(1/4)^2$	$(1 - 1/4)(1/4)^2$
n=2	$1/16 \cdot 1/2$	$(1 - 1/16)(1/16)^2$	$(1 - 1/16)(1/16)^2$
n=10	$1/4^{10} \cdot 1/2$	$(1 - 1/4^{10})(1/4^{10})^2$	$(1 - 1/4^{10})(1/16)^2$

Table 2: Comparison of expected likelihood of interactions between agents with random alleles. Neutral model probabilities were chosen to match those of the original model. Probability of combat is the chance of matching either agent's offense tag against the others combat condition times the chance of fleeing. Probabilities of mating and trade equal the chance of no combat times the chance of matching both agents' mating tags with mating conditions (for mating) or offense tags against trading conditions (for trade). The alphabet size for genotypes is the number of resources used ($r = 4$). For the original model, probabilities are shown for conditions of length (n) 0, 1, 2, and 10.

necessary to produce patterns such as the species-area scaling relation. The neutral model we implemented substitutes the tag-based interactions described in Section 2.3 with random interactions. When two agents are chosen to interact, the decision of whether they will fight, trade, or mate is based on random tosses of a biased coin, rather than the string-matching criteria described in Section 2.3. This adjustment is analogous to an ecology in which species interact with one another randomly. Aside from this modification, all other bookkeeping (i.e., resource assimilation, self-replication, migration, taxation, and random death) is performed identically in both models.

Parameters for the neutral model were chosen to produce the same interaction probability between two randomly selected agents as in the original Echo model. The chance an interaction will occur between two randomly chosen agents is the sum of the probabilities that combat, trade, or mating will occur (Table 2). For the neutral model, probabilities of combat, trade, and mating were set at 1/4, 1/8, and 1/8, respectively. If an agent is attacked by another agent in combat, the attacked agent is given a chance to flee from the attacker. The probability of escape equals the probability of losing in combat. In the neutral model, an agent will escape attack with a probability of 1/2. In the neutral model, the chance of being attacked and defeated is the product of the probability of combat and the chance of escape ($1/4 \cdot 1/2 = 1/8$). The probability that any interaction will occur between two selected agents in the neutral model is 10/32 (0.31).

In the original model, the probability any interaction will occur is based on the chance of randomly matching a condition as a prefix of a tag. Mating and trade occur only in the absence of combat and if each agent's condition is a prefix of the other's tag. Denoting n

as the mean length of the conditions in agents' genomes, we calculate the probability of any interaction occurring between two agents in Echo to be $P_{int} = 1/2$ (0.5) for $n = 0$, 7/32 (0.22) for $n = 1$, and $\approx 1/32$ (0.03) for $n = 2$. In our experiment, the mean length of conditions in the original model was 0.98, varying between 0 and 6, with some conditions being much longer. Comparing P_{int} for $n = 1$ with the total probability an interaction will occur in the neutral model (0.31), we conclude that the average interaction probabilities agree for the two versions of Echo.

4.3 Experimental Design and Data Analysis

The simplest way to study relative abundance in Echo is to sort the genomes by their abundance (consider each different genome to be a different species), and to plot these by rank on the x-axis and by number of individuals on the y-axis. In Figure 6, we report population data from various Echo runs (see caption) and group it into octaves, as was done by Preston. Although other definitions of species in Echo have been considered (see [14]), here we use the term species to refer to a unique genotype.

In the species-area experiments, we tested whether or not $S = cA^z$ in both the original and neutral versions of Echo. We used a $4 \times 2 \times 2$ factorial design, with 20 replicates for each combination of world area (1, 4, 16, or 64 sites), number of iterations (10^3 or 10^4 iterations), and model type (original or neutral model). For each replicate population, we recorded S (number of species), N (number of individuals), and N_{max}/N (an index of dominance, the fraction of the population represented by the most abundant species [38]). We also calculated the mean number of species and individuals per site (S/site and N/site). To test for significant treatment effects, we used two-way analysis of variance (ANOVA) on three response variables: mean S and N per site, and dominance. Separate analyses were conducted for runs lasting 10^3 and 10^4 iterations.

We also compared regression coefficients of species-area functions for the original and neutral models. A variable was eliminated from our regression equation if it had less than 90% probability of being significantly different from 0. The full regression equation was of the form $\widehat{S} = (\beta_1 + \beta_2)A^{(\beta_3+\beta_4)} + \epsilon$ for the original model and $\widehat{S} = \beta_1 A^{\beta_3} + \epsilon$ for the neutral model. To interpret these regression equations, note that eliminating β_4 from the best regression equation implies that model type has no significant effect on z, and that β_3 can be used to predict z for both models. Similarly, eliminating β_2 implies that model type has no significant effect on c, and can be predicted by β_1 alone. The ϵ term describes residual error.

4.4 Results

Preston curves from typical populations are shown in Figure 6. Populations from both the original and neutral model simulations are plotted, for short runs in small worlds, and for longer runs in larger worlds. Distributions expected under Preston's hypothesis are plot-

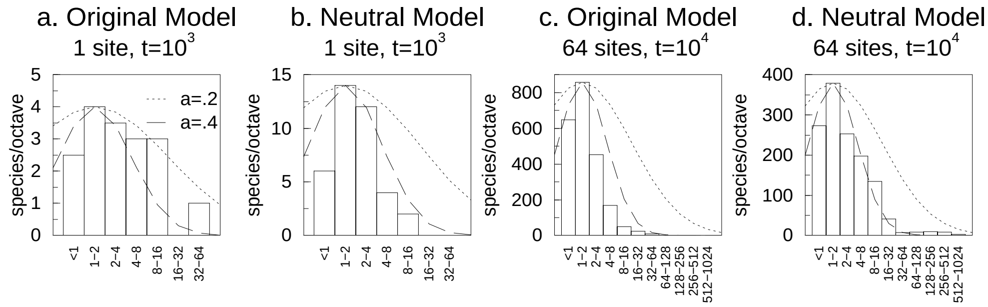

Figure 6: Preston curves from four Echo populations. See Section 3.1 for a description of Preston curves. Populations are from original (a, c) and neutral (b, d) models, either for a one-site world at 10^3 iterations (a, b), or for a 64-site world at 10^4 iterations (c, d). Each population was randomly chosen from 20 replicates. Expected values under Preston's canonical hypothesis are plotted as dotted and broken lines, for $a = 0.2$ and $a = 0.4$, respectively.

ted for comparison. These data show that qualitatively the pattern of genome abundances in Echo populations resembles the general patterns found in biological systems, although there are some differences. We have observed this general pattern in a wide variety of Echo runs under a wide variety of different parameter settings and initial conditions. Overall, abundances in the original model are more clumped than in the neutral model (note differences in scale on the y-axis). That is, the original model has a greater bias towards rare species (singletons) than common species. We will see a similar pattern in the species-area experiments.

For a number of reasons we do not believe that a detailed quantitative comparison of Echo species diversity with that of natural ecosystems would be meaningful, so the remainder of our experiments focus on the species-area scaling relation. A more meaningful test can be obtained by studying the relation of species richness (S) to size of habitat (or sample), known as the species-area relation. The species-area relation avoids many of the procedural artifacts associated with the Preston Curve (see Section 5).

In testing for effects of model type and world area on response variables in shorter runs (Figure 7a, b, c), the neutral model yields more species per site than the original model in simulations lasting 10^3 iterations ($P = 0.0001, d.f. = 1, F = 294.66$). Increasing area decreases number of species per site in the neutral model ($P = 0.0001, d.f. = 3, F = 78.68$) and decreases dominance by a particular species in the original model ($P = 0.0001, d.f. = 3, F = 58.67$). Number of individuals per site is not affected by model type or area of the world.

Variances in all response variables are greater in longer runs than in shorter runs (Fig-

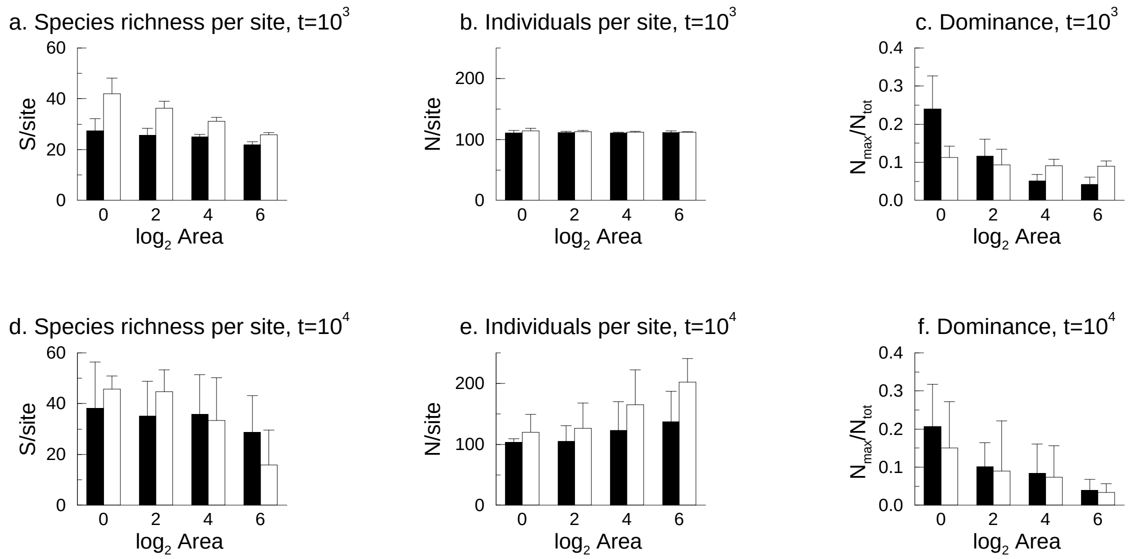

Figure 7: Response variables from simulations of 10^3 iterations (a, b, c) and 10^4 iterations (d, e, f). Species richness per site (a, d), population size per site (b, e) and dominance (c, f) are plotted as a function of area. Solid (black) bars are from the original model and unfilled (white) bars are from the neutral model. Each bar represents a mean from twenty replicates. Error bars indicate one standard deviation.

ure 7d, e, f). Dominance and number of species per site do not differ with model type, but the neutral model has more individuals per site than the original model ($P = 0.0001, d.f. = 1, F = 31.24$). Increasing area increases the number of individuals per site ($P = 0.0001, d.f. = 3, F = 17.27$), and decreases dominance ($P = 0.0001, d.f. = 3, F = 17.31$) and number of species per site ($P = 0.0001, d.f. = 3, F = 15.28$).

A scaling relation describes the species-area function in both models at 10^3 iterations, and in the original model at 10^4 iterations (Figure 8). Mild curvilinearity might be present in the species-area function for the neutral model at 10^4 iterations. With this minor exception, we note that both models demonstrate the hypothesized scaling relation. Multiple regression analyses indicate all terms in the full regressions equation are statistically significant (Table 3). That is, we reject with more than 99% confidence the null hypotheses that $\beta_2 = 0$ and $\beta_4 = 0$. Thus, values of c and z for the original and neutral models differ significantly. Actual values for z at 10^3 iterations are 0.95 and 0.87 for the original and neutral models, respectively (Figure 8a). Simulations of 10^4 iterations yield z values of 0.93 and 0.70 for original and neutral models (Figure 8b).

T	n	β_1	β_2	β_3	β_4	r^2
		(c)		(z)		
10^3	160	5.39	-0.62	0.87	0.08	0.995
10^4	160	5.78	-0.62	0.70	0.23	0.877

Table 3: Coefficients from multiple regression analysis. Treatments of different durations (T) were analyzed separately. Parameter estimates are shown for the best regression equation, after backwards elimination of parameters with $P > 0.10$. See section 4.3 for the multiple regression equations. All parameters are significant at $P \leq 0.001$.

5 Discussion

Our results show that the original and neutral versions of Echo behave differently with respect to the number of species and population size per site, as well as the scaling exponent in the species-area scaling relation. This suggests that the dynamics of the two models are different. Both models produce clumped distributions of abundance among species although neither produces a canonical log-normal distribution. Both the original and neutral versions of Echo yield positive species-area relationships, but with scaling exponents far greater than the expected value of 1/4. This result might be attributed to a variety of factors, discussed later.

Several considerations lead us to question whether the log-normal distribution is an important consequence of identifiable ecological processes or whether it is simply the expected outcome of a random process with certain properties. However, a qualitative comparison provides an important first check of how Echo compares with empirical distributions of abundance. Our data show a good qualitative resemblance, under a wide range of Echo operating conditions, as shown in Figure 6 and [14].

It is well-known that log-normal distributions often arise when random processes are the outcomes of multiplicatively compounded subprocesses [2, 32, 33, 16, 34, 50]. Randomly compounded subprocesses can account for log-normal distributions in general, but they fail to account for Preston's "canonical" log-normal, in which distributions always have similar variance. However, a number of the steps used to construct Preston curves might limit species abundance distributions to the canonical family, a point made by several other researchers [32, 38, 33]. Location of the veil line near the mode of the distribution, sampling and the histogram binning procedure, and problems with fitting a specific nonlinear function to discretized data all complicate the task of testing formally for a canonical log-normal distribution. For example, the procedure of splitting a count between octaves forces an inflection of the curve about a mode of one. In the case of singletons (species with only one representative), half of the counts are assigned to the octave labeled "< 1", and half of the counts are assigned to the "1–2" octave. Species with 2, 3, or 4 representatives will also

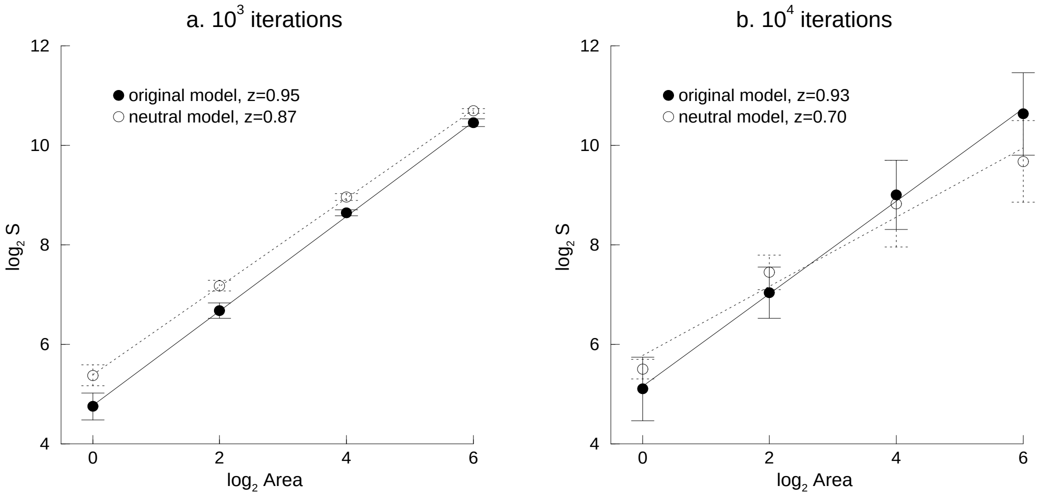

Figure 8: Species-area curves for 10^3 (a) and 10^4 iterations (b). Species counts for the original model (•) and neutral model (○) are plotted against area of the world. Note that both axes are $\log_2$-transformed. Lines are regression fits using parameters in Table 3. Slope of each regression line (z) is indicated in the legend. Each point is a mean of 20 replicates. Error bars indicate one standard deviation.

fall into the "1–2" octave, so this octave will generally have a greater value than the "< 1" octave. Thus, the observation of an inflection near the veil line is caused by the binning procedure used in constructing the histogram. If an alternative procedure were used, we might not observe an inflection, and perhaps not infer a unimodal distribution. All of these complications make it difficult to explain what mechanisms produce the Preston distribution.

There are many factors in Echo that might contribute to the discrepancy between the expected and observed values for z. These include specific parameter choices (although the model is intended to be robust to a wide range of parameter settings), the way species are defined, and intrinsic rates of species formation and extinction.

Our choice of simulation parameters was largely arbitrary. A different choice of parameter settings might produce different results than we observed. However, our experience with Echo has shown that the general patterns reported here appear to be relatively insensitive to parameter settings. We have obtained qualitatively similar results in experiments with lower mutation rates, increased interaction fractions, and different initial populations.

A second factor affecting our results is the definition of species in Echo. We used the provisional definition of unique genotypes as species. Another possible definition is based on clustering genomes by a distance metric (number of mutations to produce one genotype from another) [14]. This clustering technique produces curves that more closely approximate the canonical Preston distribution. In our current experiments, such a definition might reduce the number of singletons in a population, and thus lower the slope of the species-area curve. We decided not to use the clustering technique here because of the problem of defining species by an arbitrary genetic distance, but it could significantly change the outcome of our experiment by reducing the number of singletons.

Species-area curves are typically constructed from the total number of individuals present in a standardized sample, rather than a complete census. We tested whether sampling affected our results by constructing species-area curves using sub-sampling. Rather than including each individual in a sample with a probability of one, we selected individuals with probabilities of 0.1 and 0.25. That is, we tried sampling both 1/10 and 1/4 of the population and recalculating the species-area curves. This procedure did not noticeably change the slope of the curves (data not shown). Sampling and complete censusing of Echo populations produce similar species-area curves.

Yet another factor in Echo that affects our results is the balance between rates of speciation and extinction. New species are formed from mutations of existing species, and by recombining genotypes through mating. Formation of new species is thus controlled by the mutation and recombination rates. Species go extinct on the death of the last surviving individual, whether through competition or by chance. Thus, all the extinction events in Echo are intrinsic (cf. Van Valen's law [45]). However, the mechanisms of extinction in natural systems can include exogenous events never before experienced by the species.

Raup proposes that mass extinctions could be caused by events that occur on a time scale greater than species' evolutionary time scales [44]. This idea is promising, because it parsimoniously explains how widespread species might suffer extinction. If the exogenous events were to occur on an evolutionary time scale, natural selection might favor species that can tolerate such adversity. Species with small populations are more likely to suffer extinction by such mechanisms as fluctuations in population size or competition than widespread species. Again, this would have the effect of flattening the species-area curve. For a firmly established (widespread) species to suffer extinction requires extreme circumstances, or rare events. Raup calls such unnatural events "first strikes" and argues that these rare events can play a major role in driving common species to extinction. Unlike earth, the Echo model we used here is a closed system; there are no meteorite impacts or global changes in climate or sea level to drive species nearer to extinction through previously unencountered adversity.

From all of the above considerations, we conclude that the mechanisms in the current implementation of Echo do limit species' persistence and abundance to a certain extent. However, they do not accurately reflect the balance observed in nature. Competition and random death as means for extinction apparently fail to produce the quantity of extinctions necessary to agree with those seen in nature. However, Echo does exhibit a robust positive species-area scaling relation as well as a log-normal species-abundance curve similar to a Preston curve. The degree of quantitative agreement between these curves and those found in natural ecosystems is perhaps less important than the fact that we see similar kinds of curves.

More generally, these results represent a first step towards validating Echo as a model of natural ecological systems. Echo is clearly quite a different kind of model than those typically used in ecological studies, and it is relevant to ask what we can hope to learn from a highly abstract model that by design does not directly correspond to any real system. There are at least three possible answers to such a question:

- Echo as a "patch of dirt,"
- Echo as an ecological abstraction, or "flight simulator"[20], and
- Echo as a theory of complex adaptive systems.

By "Echo as a patch of dirt," we mean to suggest the possibility of constructing Echo worlds that correspond directly to some real ecosystem. This would allow careful quantitative validation of Echo's behavior and should lead to concrete predictions. This kind of modeling is well beyond the current state of the Echo research program. Although it would be an important achievement for Echo to simulate one real ecosystem accurately, this is not likely to be its most important contribution. Other modeling approaches (e.g., most simulation studies and many analytical models) have traditionally focused on exactly this problem of making quantitative predictions about the behavior of specific systems with given parameter settings.

A second use of Echo is as an ecological abstraction to be used for building intuitions about how ecosystems work, what is critical to their stability, and so forth. Under this view, Echo itself is a rich enough ecology to be worth studying in its own right, along the lines that we have outlined in the previous sections. We can study patterns of behavior, e.g., how resources flow through different kinds of ecologies, how cooperation among agents can arise through evolution, and arms races [19]. We can also use such a model to identify parameters or collections of parameters that are critical, i.e., to perform sensitivity analysis. As with any simulation tool, it is much easier to run hypothetical what-if experiments than to conduct experiments on a real system. If a model like Echo were successful, it would enable users to build deep intuitions about how different aspects of an ecological system affect one another, important dependencies, and an appreciation of how evolution interacts with the internal dynamics of an ecosystem. The neutral model that we introduced in this paper is an example of how these intuitions can be developed, explored, and used to understand what kinds of forces produce the diversity we observe in nature. This is perhaps the most important contribution that models like Echo can make.

A third, and more ambitious, view of Echo is as a theory of complex adaptive systems [20]. Echo can be viewed as the implementation of a theory that captures what we we believe to be the essential components and interactions in a wide variety of CAS. As such, it articulates our beliefs about which mechanisms are relevant and which are not. The theory can be confirmed or disconfirmed by running the model under a wide variety of operating conditions and observing the extent to which the expected macro-level behaviors arise spontaneously. Our experiments on species diversity are an example of such a confirmation process, and our introduction of the neutral model provides another way of testing the extent to which the current Echo mechanisms are all required to produce the behaviors of interest. It should be noted that the version of Echo reported here is an early attempt at a broad CAS theory, and we expect that many of the details will be modified over time (e.g., to address the problem of extinctions, metabolic transformations, and so forth) and when careful comparisons to other CAS (e.g., economic models) are made. Nevertheless, an important contribution of computational systems like Echo is to distill features that are common across many CAS in a testable model.

Although we are optimistic about the insights that artificial life models like Echo can contribute to understanding real ecosystems and other CAS, it is important to note that such models have several apparent drawbacks. These include a correspondence problem, what concrete questions they can address, scaling issues, and nonlinear interactions (already discussed). We discuss the some of these issues briefly in the following paragraphs. Despite these drawbacks, we believe that there is much to be learned from discrete, agent-based models such as Echo.

Because CAS models tend to strip away many details, it is often impossible to say what any component of one of these models corresponds to in the real world. In the immune system, for example, many theoretical immunologists use string matching to model receptor/ligand binding [37]. Patterns of bits (or other symbols) are used to represent both molecular shape and electrostatic charge. Consequently, it is difficult to say what one bit in the model corresponds to in the immune system. Because different alphabets and different matching rules can have very different properties, the challenge is to select an alphabet and matching rule that have general properties similar to the real system without worrying too much what each bit really stands for [49]. By contrast, most theories of modeling are based on the premise that a direct correspondence can be established between the modeled system and the primitive components of its model.

As a consequence of this correspondence problem, it is not always clear what scientific questions are being addressed by CAS models. In more conventional simulation-based modeling, models are used to make quantitative predictions based on certain inputs, for example, to determine optimal parameter values. Agent-based models of CAS are rarely able to make this kind of quantitative prediction, and as a result the focus is on identifying broad categories of behavior and critical parameters (but not necessarily the exact critical parameter values), as we discussed earlier.

A third problem faced by agent-based models is one of scale. Because they are simulations, agent-based models often operate on vastly different time scales of evolution and with much smaller population sizes than those of the systems they model. Also, we tend to be intolerant of high failure rates such as those often observed in nature. For example, consider the selection algorithms typically used in genetic algorithms. Selection pressure is maintained at an artificially high rate and often scaled to maintain increased pressure near the end of a run. Evolution thus occurs orders of magnitude more quickly than in natural systems, and as a result, we may lose some of the richness of the natural evolutionary process.

6 Conclusion

Systems comprised of many interacting components can be categorized as adaptive or non-adaptive. Whereas nonadaptive systems are studied by analysis or numerical simulation, adaptive systems are less amenable to these methods. Innovative approaches to simulation (e.g., genetic algorithms) enable simulation of the dynamics of adaptive systems. In particular, Echo extends the genetic algorithm concept to model dynamics of adaptive agents in a spatially-distributed and resource constrained setting. In this paper, we have discussed both the difficulties and successes of analyzing and modeling CAS. We described a software implementation of a class of models called Echo, introduced by Holland [19, 20]. In addition, we compared the evolutionary behavior of Echo to real evolutionary systems.

To verify that Echo mechanisms are sufficient to model evolutionary dynamics, we performed simulations to compare species diversity and abundance patterns with patterns commonly seen in natural ecosystems. Specifically, we considered the relative abundances of species and tested for a scaling relation between available habitat and species richness. Because it is unclear whether random processes or adaptive behavior produce the patterns in empirical systems, we used a neutral model to test whether all of Echo's mechanisms are necessary. The neutral model replaces genome-mediated interactions with random inter-agent interactions.

Qualitative agreement with empirical systems was observed for both models, although neither model agrees exactly with quantitative predictions. Original and neutral versions of Echo exhibit significantly different patterns of species richness and population structure.

7 Acknowledgments

This work was supported by the Santa Fe Institute (Project 2050 and ONR grant N00014-95-1-1000). Hraber was partially supported by NSF grant BSR-9107339 awarded to B.T. Milne. Support for Forrest was provided by NSF grant IRI-9157644 and ONR grant N00014-95-1-0364. John Holland designed the original Echo model on which our work is based, and he suggested many of the ideas in this paper. James Brown, Marc Feldman, Simon Fraser, Timothy Keitt, Andrew Kosoresow, and Bruce Milne have all provided many useful comments on Echo. Brown suggested the examination of Preston's work on species abundance. We also thank Josh Epstein, Rob Axtell and all those involved in Project 2050 at the Santa Fe Institute, the Brookings Institute and the World Resources Institute for their comments.

8 Appendix A: Echo Parameters

The parameters are described below:

- The number of resource types that will exist in the world.

- The *interaction fraction* determines the number of agent-agent interactions that will occur at a site in each Echo cycle. This is determined by multiplying the interaction fraction by the population at a site. Use of a constant fraction to determine interaction rates regardless of population size does not accomodate density-dependent effects, which is an unfortunate oversimplification.

- The *trading fraction* determines what proportion of an agent's excess trading resource it will exchange in a trading interaction with another agent. Trade is described in section 2.3.

- The *taxation probability* is the probability that an agent is taxed in an Echo cycle (see Section 8 for details).

- The *neighborhood* determines the possible directions of migration between sites. The three possibilities, given a rectangular array of sites, are none (i.e., no migration), the four compass points (used in the results described in this paper) or the compass points plus the diagonals.

There are three parameters for each Echo site:

- The *mutation probability* determines the probability that the genome of an agent is mutated during self-reproduction. This is a per-gene probability. Mutation is handled slightly differently for different genes (see Section 2.2).

- The *crossover probability* is used similarly to a genetic algorithm. It determines the probability that two agents that mate will be subject to crossover. If crossover is not performed, the agents are unchanged.

- At the end of an Echo cycle, there is a (typically small) probability of each agent being killed. This is called the *random death probability.*

In addition to these parameters, every site must specify four resource-level vectors. Each of these consists of a vector of non-negative integers whose length is the same as the number of resource types in the world. These vectors are used as follows:

- The *production* vector contains the number of resources of each resource type that will be produced by the site in each Echo cycle.

- The *initial* vector contains the number of resources of each resource type that will be present at the site when it is first created.

- The *maximum* vector contains the maximum level of each resource type that can accumulate at the site.

- The *maintenance* vector contains the number of each resource type that will be charged to an agent when it is taxed.

References

[1] D. H. Ackley and M. L. Littman. Interactions between learning and evolution. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 487–507, Reading, MA, 1992. Addison-Wesley.

[2] J. Aitchison and J. A. C. Brown. *The Log–Normal Distribution.* Cambridge University Press, Cambridge UK, 1957.

[3] J. H. Brown. Two decades of homage to Santa Rosalia: Toward a general theory of diversity. *American Zoologist*, 21:877–888, 1981.

[4] J. H. Brown. Complex ecological systems. In George A. Cowan, David Pines, and David Meltzer, editors, *Complexity: Metaphors, Models and Reality*, volume XIX of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 419–433. Addison–Wesley, Reading, MA, 1994.

[5] H. Caswell. *Matrix Population Models: Construction, Analysis, and Interpretation.* Sinauer Associates, Sunderland, MA, 1989.

[6] D. L. DeAngelis and L. J. Gross. *Individual-Based Models and Approaches in Ecology: Populations, Communities and Ecosystems.* Chapman and Hall, New York, NY, 1992.

[7] R. L. Devaey. *Chaotic Dynamical Systems.* Addison-Wesley, Redwood City, CA, second edition, 1989.

[8] R. Durrett and S. A. Levin. The importance of being discrete (and spatial). *Theoretical Population Biology*, 46:363–394, 1994.

[9] R. Durrett and S. A. Levin. Stochastic spatial models: A user's guide to ecological applications. *Philosophical Transactions of the Royal Society of London B*, 343:329–350, 1994.

[10] R. Durrett and S. A. Levin. Spatial models for species area curves. *Journal of Theoretical Biology*, in press.

[11] J. M. Epstein and R. L. Axtell. Growing artificial societies: Social science from the bottom up. Technical report, Brookings Institution, 1994.

[12] J. Doyne Farmer. Connectionism: A Rosetta Stone for adaptive dynamics. *Physica D*, 42(1–3):153–187, 1990.

[13] S. Forrest. *Emergent Computation.* MIT Press, Cambridge, MA, 1991.

[14] S. Forrest and T. Jones. Modeling complex adaptive systems with Echo. In R. J. Stonier and X. H. Yu, editors, *Complex Systems: Mechanisms of Adaptation*, pages 3–21. IOS Press, Amsterdam, 1994.

[15] G. Hartviggsen and W. T. Starmer. Plant-herbivore coevolution in a spatially and genetically explicit model. *Artificial Life*, 2(2):239–259, 1995.

[16] T. Hogg. The dynamics of complex computational systems. In W. Zurek, editor, *Complexity, Entropy and the Physics of Information*, volume VIII of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 207–222. Addison-Wesley, Reading, MA, 1990.

[17] J. H. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, 2nd edition, 1992.

[18] J. H. Holland. The effect of labels (tags) on social interactions. Technical Report 93–10–064, Santa Fe Institute, Santa Fe, NM, 1993.

[19] J. H. Holland. Echoing emergence: Objectives, rough definitions, and speculations for Echo-class models. In George A. Cowan, David Pines, and David Meltzer, editors, *Complexity: Metaphors, Models and Reality*, volume XIX of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 309–342. Addison–Wesley, Reading, MA, 1994.

[20] J. H. Holland. *Hidden Order: How Adaptation Builds Complexity*. Addison–Wesley, Reading, MA, 1995.

[21] J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. Thagard. *Induction: Processes of Inference, Learning, and Discovery*. MIT Press, 1986.

[22] B. Hölldobler and E. O. Wilson. *The Ants*. Belknap Press of Harvard University Press, Cambridge, MA, 1990.

[23] P. T. Hraber and B. T. Milne. Community assembly in a model ecosystem. *Ecological Modelling*, in press.

[24] B. A. Huberman. The performance of cooperative processes. In S. Forrest, editor, *Emergent Computation*, pages 38–47, Cambridge, MA, 1991. MIT Press.

[25] G. E. Hutchinson. Homage to Santa Rosalia, or Why are there so many kinds of animals? *American Naturalist*, 93:145–159, 1959.

[26] T. B. Kepler and A. S. Perelson. Somatic hypermutation in b cells: An optimal control treatment. *Theoretical Biology*, 164:37–64, 1993.

[27] N. R. Klinman, J. L. Press, N. H. Sigal, and P. J. Gerhart. Need title. In A. J. Cunningham, editor, *The Generation of Antibody Diversity*, page 127. Academic Press, New York, NY, 1976.

[28] C. G. Langton. personal communication.

[29] C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors. *Artificial Life II.* Addison-Wesley, Redwood City, CA, 1992.

[30] R. H. MacArthur and E. O. Wilson. *The Theory of Island Biogeography.* Princeton University Press, Princeton, NJ, 1987.

[31] R. M. May. *Stability and Complexity in Model Ecosystems.* Princeton University Press, Princeton, NJ, 1973.

[32] R. M. May. Patterns of species abundance and diversity. In M. L. Cody and J. M. Diamond, editors, *Ecology and Evolution of Communities*, pages 81–120. Harvard University Press, Cambridge, MA, 1975.

[33] R. M. May. The search for patterns in the balance of nature: Advances and retreats. *Ecology*, 67(5):1115–1126, 1986.

[34] E. W. Montroll and M. F. Schlesinger. On $1/f$ noise and other distributions with long tails. *Proceedings of the National Academy of Sciences (USA)*, 79:3380–3383, 1982.

[35] M. H. Nitecki and A. Hoffman. *Neutral Models in Biology.* Oxford University Press, Oxford UK, 1987.

[36] S.W. Pacala. Neighborhood models of plant population dynamics. 2. multispecies models of annuals. *Theoretical Population Biology*, 29:262–292, 1986.

[37] A. S. Perelson and G. F. Oster. Theoretical studies of clonal selection: minimal antibody size and reliability of self- non-self discrimination. *J. Theoretical Biology*, 81:645–670, 1979.

[38] E. C. Pielou. *Mathematical Ecology.* Wiley, New York, NY, 1977. (second edition).

[39] S. L. Pimm. The complexity and stability of ecosystems. *Nature*, 307:321–326, 1984.

[40] F. W. Preston. The commonness, and rarity, of species. *Ecology*, 29(3):254–283, July 1948.

[41] F. W. Preston. The canonical distribution of commonness and rarity: Part I. *Ecology*, 43(2):185–215, Spring 1962.

[42] F. W. Preston. The canonical distribution of commonness and rarity: Part II. *Ecology*, 43(2):410–432, Spring 1962.

[43] F. W. Preston. Noncanonical distributions of commonness and rarity. *Ecology*, 61(1):88–97, 1980.

[44] D. Raup. *Extinction: Bad Genes or Bad Luck?* W. W. Norton, New York, NY, 1991.

[45] D. M. Raup. Biological extinction in earth history. *Science*, 211:1528–1533, 1986.

[46] T. S. Ray. An approach to the synthesis of life. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408. Addison-Wesley, Redwood City, CA, 1992.

[47] O. J. Schmitz and G. Booth. Modeling food web complexity: the consequence of individual-based spatially explicit behavioral ecology on trophic interactions. *Evolutionary Ecology*, in press.

[48] L. B. Slobodkin. *Growth and Regulation of Animal Populations.* Rinehart and Winston, New York, NY, 1961.

[49] D. Smith. Towards a model of associative recall in immunological memory. Master's Project, University of New Mexico, 1994.

[50] G. Sugihara. Minimal community structure: An explanation of species abundance patterns. *The American Naturalist*, 116(6):770–787, 1980.