Modeling Complex Adaptive Systems with Echo

Stephanie Forrest
Department of Computer Science
University of New Mexico
Albuquerque NM 87131, USA.
forrest@cs.unm.edu

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

Abstract.
Complex adaptive systems (CAS) consist of many interacting and adapting components. Echo is a computational CAS model in which evolving agents are situated in a resource-limited environment. Different views of the notion of species within Echo are compared to biological experiments on relative species abundance, specifically to Preston's "canonical" lognormal distribution.

To appear in: *Complex Systems: Mechanism of Adaptation*, R.J. Stonier and X.H. Yu (Eds), Amsterdam: IOS Press, September 1994, pp. 3–21.

1. Introduction

Many interesting systems are difficult to describe or control using traditional methods. These include natural ecological systems, immune systems, economies and other social systems. 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 as well as in some forms of computation [1]. 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*. 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."
- Interactions among agents and between agents and their environment.
- Unanticipated global properties often result from the interactions.
- Agents adapt their behavior to other agents and environmental constraints.
- As a consequence, system behavior evolves over time.

Building models of CAS is difficult for several reasons. First, useful and predictive mathematical analyses rarely exist. This is due to both nonlinearities and the changing behavior of the primitive elements of the system. Second, detailed simulations are problematic because it is virtually impossible to get all of the details correct. Consider, for example, the vertebrate immune system which in some cases has been estimated to express over 10^7 different receptors at a time. 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 expressed repertoire of receptors precisely. This problem exists for all 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 to this dilemma is to strip away as much detail as possible, retaining only the essential interactions. The goal is then to develop models whose behavior is robust with respect to the details of the interactions (e.g., avoiding parameter tweaking to coax a system to produced desired behaviors), and which produces the broad categories of behaviors in which we are interested. An implication of this approach is that such models will rarely, if ever, be able to make precise quantitative predictions. Adaptation is central in CAS, and this is a third reason that modeling CAS is difficult. The underlying rules of the system are changing over time which means that different agents behave according to different rules at different times.

Because of these difficulties, a class of models, variously called "artificial worlds," "particle-based," and "agent-based," have been a popular approach to studying CAS. This style of modeling is quite different from the differential-equation style of models used most frequently to model nonlinear dynamical systems. In agent-based models, each "actor" and each interaction among actors (i.e., not just each type of interaction) is represented (simulated) explicitly. Individuals are capable of quite different kinds of behaviors (the agents in the system are heterogeneous). Agent-based models are discrete in most dimensions, typically time, state, and update rules. Thus, the standard approximations for infinite-sized systems and the techniques developed for studying asymptotic behavior of continuous nonlinear dynamical systems often do not directly apply. As a result, these systems tend to be more difficult to analyze.

Agent-based CAS models have several apparent drawbacks. These include the mapping problem, the problem of asking the right question, scaling issues, and nonlinear interactions (already discussed). 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. Continuing with the immune system example, many theoretical immunologists use string matching to model receptor/ligand binding [2]. 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. Since different alphabets and different matching rules can have very different properties, the challenge is to select an alphabet and matching rule that has general properties similar to the real system without worrying too much what each bit really stands for [3]. Most theories of modeling are based on the premise that a correspondence can be established between the modeled system and the primitive components of its model. As a consequence of this mapping 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 predicated 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). A third problem faced by agent-based models is one of scale. Because they are simulations, agent-based models typically 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.

We have studied several different CAS models over the past fifteen years. Genetic algorithms [4] focus on the evolutionary component of CAS. They are reasonably well understood and mature, but ignore several important features, including resource allocation, heterogeneity, and endogenous fitness. Classifier systems [5, 6] 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. What can we hope to learn with a model that by design does not correspond to any real system? 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 [7]. We can also use such a model to identify parameters or collections or 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 and correct, 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 ongoing dynamics of an ecology. This is perhaps the most important contribution that models like Echo can make. The original idea of Echo, including motivation, design decisions, and overall structure were introduced in [4, 7]. Our goal in this paper is to describe more fully one specific Echo model (Echo really refers to a class of models) and to show how one might study the extent to which Echo does or does not capture important properties of ecological systems. Towards this end, we report preliminary results on the relative abundance of species, an important feature of any ecological system. This feature raises some fundamental questions, such as how to define precisely the concept of "species" in Echo, which we also discuss.

2. Echo

Echo was designed to capture the essential features of ecological systems in an agent-based model. All of the entities and interactions in Echo are highly abstract, and it is not yet known whether Echo can be used to model real-world phenomena effectively. Many CAS can be viewed as ecologies (e.g., [8]), but our focus in this paper is on the analogy with natural ecologies. Echo resembles some other CAS models. These include Swarm [9], Sugarscape [10], and the Evolutionary Reinforcement Learning (ERL) model [11]. Unlike Swarm, Echo makes specific commitments about agent types and interactions; it differs from Sugarscape, both in specific details, and in its focus on ecological principles;

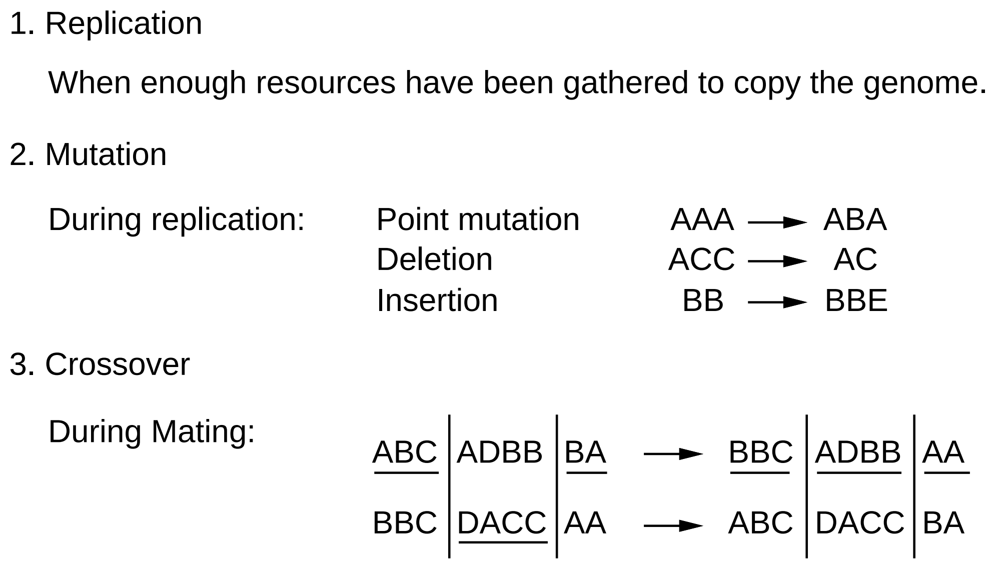

Figure 1: The ways in which an Echo agent can undergo genetic modification.

ERL provides two levels of learning (there is only one in Echo) but is not intended as a general ecological model.

Echo extends classical genetic algorithms in several important ways: (1) fitness is endogenous, (2) individuals (called agents) have both a genome and a local state that persists through time, and (3) genomes are highly structured. 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. Agents acquire resources through interactions with other agents (combat or trade) or from the environment. This mechanism for "endogenous" reproduction comes much closer to the way fitness is assessed in natural settings than conventional "fitness functions" in genetic algorithms.

Along with these extensions to the evolutionary component, Echo specifies certain structural features of the environment in which agents evolve. Specifically, 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 notion of neighborhood within a site. Each site may produce renewable resources. These resources are represented by different letters of the alphabet, and genomes are constructed from the same letters. Resources can exist in three places: as part of an agent's genome, as part of an agent's local state, or free in the environment. There are three forms of interactions among agents: trade, combat, and mating. In trade, resources stored internally (the local state) are exchanged; in combat, all resources (both genetic and stored) are transferred from loser to winner; in mating, genetic material is exchanged through crossover, thus creating hybrids. Mating, together with mutation during the replication process, provides the mechanism for new types of agents to evolve, as shown in Figure 1. Resource constraints provide the pressure for agents to diversify and occupy new niches.

In each Echo run there is a fixed number of resource types which is determined by the user of the system. These may be representative of resources in a real-world system, or may correspond to a more abstract notion of something that is required to ensure survival. For example, the environment can be designed to require that agents possess a certain resource, which some agents may only obtain through trade. In this situation, the resource need not be thought of as corresponding to a physical entity, but as something that requires a certain type of agent-agent interaction for agent survival. The number of resources in an Echo world is typically small. These are denoted by lower-case letters: a, b, c and so on. In the Echo world used in this paper, there are four resources and one site.

The following sections describe Echo in more detail. Much of this is devoted to describing agents and the interactions that can occur, both between pairs of agents and between an agent and its environment.

2.1. *Echo Structure*

Our implementation of Echo divides Echo into a structural hierarchy. Each run of Echo involves a *world* that contains a fixed number of *sites*. Each site may contain an arbitrary number of *agents*, including zero. Each world specifies certain system-wide parameters, including the number of sites, the number of resource types, the taxation rate, parameters controlling replication, and the probability of random death. See [12] for details of these parameters. Each site specifies its own mutation, crossover, and random death probabilities, as well as some parameters controlling the details of how resources are managed at the site (e.g., the maximum amount of a resource that can accumulate at the site).

Each of these components is designed by the user of the system, typically as an abstraction of some aspect of a real-world CAS. In each case, the use of Echo requires decisions about the structure of these objects and the ways they will behave when the result is set in motion. This paper refers briefly to the elements of worlds and sites. A full description of these elements can be found in any of [12, 4, 7]. Section 2.3 describes the structure and properties of agents.

2.2. *The Echo Cycle*

The sequence of events in an Echo cycle consists of the following:

1. Interactions between agents are performed at each site. These include trade, mating, and combat. The number of interactions is controlled by a "world" parameter.

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 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 as bad luck or as a mechanism that prevents agents from living forever. If they are not killed some other way (through combat or taxation), they will eventually be randomly deleted.

5. The sites produce resources. Different sites may produce different amounts of each resource. For example, one site may produce ten a's and ten b's on each time step, whereas another may produce five b's and twenty c's. The thought 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 have not received resources this cycle migrate. If an agent does not acquire any resources during an Echo cycle (either through picking them up or through combat or trade), it will migrate to a neighboring site. The neighboring site is selected at random from among those permitted by the geography of the world. This is not the same as the local movement within a site that occurs as the result of the agent-agent interactions that are described in section 2.4.

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."

2.3. Agents

Figure 2 illustrates an example Echo agent. Agents have a genome which is roughly analogous to a single chromosome in a haploid species. The chromosome has $r+7$ genes, where r is the number of resources in the world. Each of these genes can be altered by the mutation operator. 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). The mutation operator can alter the allele value at any locus, and can also cause a tag or condition to grow or shrink in length.

Tags are genes that produce some easily observable feature of the phenotype. Conditions are genes that do not produce observable phenotypic effects, and their result cannot be detected by other agents. Thus an agent will interact with another on the basis of its own conditions and the other's tags. This allows, for example, the possibility of agents that appear dangerous but are in fact usually unwilling to fight. It also allows for the evolution of intransitive combat relationships. For example, an agent A might always attack an agent B, and B always 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). The importance of this kind of relationship among agents in CAS has often been stressed [4, 13].

The six tag and 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. It should be noted that the current implementation conforms to a very large extent with the description given in [4], but not with that in [7].

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

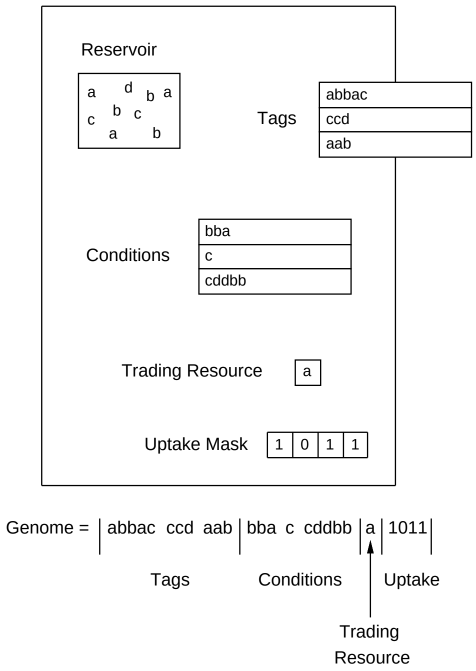

Figure 2: The structure of an Echo agent. Tags are visible to the outside world. Conditions and other properties are not.

'1' allele for the uptake gene corresponding to a certain resource, it will not be able to collect that resource if it encounters some amount of it at a site. Consequently, if the agent requires this resource (for example 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. Of course there is nothing in Echo to guarantee that such webs will not soon be greatly altered through mutation, or that they will survive at all. The final gene is the *trading resource* which is the resource type that the agent will provide to another agent if trading takes place. Each agent also has a *reservoir* in which it keeps some amount of each resource type. Resources from the reservoir are used to pay taxes, to produce offspring and for trade. The reservoir corresponds exactly to the local state of the agent.

Agents at a site 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 this array. The user must decide which agents initially reside at each site, and in what order they should appear in the array.

2.4. Agent-Agent Interactions

There are three main forms of agent-agent interaction: combat, trading and reproduction. All of these interactions take place between agents that are located at the same site and all involve the transfer of resources between agents

2.4.1. *Combat*

Combat is an idealization of any interaction that might occur between real-world entities that is antagonistic. It does not necessarily imply that the agents are actually fighting, though of course this is not precluded. If two agents in a real-world system are behaving in a competitive fashion, this would be modeled in Echo by designing the agents in such a way that they would engage in combat. When combat occurs, one agent is always killed, and its resources are given to the survivor. In a more recent version of Echo [7], the interaction need not be so extreme and results in a transfer of resources (possibly in both directions, and possibly in a very uneven fashion) between the agents.

When two agents encounter each other, the system first checks to see if either would attack the other. An agent A will attack an agent B if its 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 somewhat complicated and is not described fully here. It 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 a *combat matrix*, with special provisions for zero length genes and for genes of unequal length.

As a result of this computation, each agent receives some number of points. If A_p and B_p are the points awarded to A and B, then A will win the combat with a probability of $A_p/(A_p + B_p)$. 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.

2.4.2. *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 which is required to replicate its genome, plus some reserves (system 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 may be. This may seem an odd form of trade, but agents can "learn" to recognize each other based on their trading tags. Agents whose tags tend to involve them in disadvantageous trades will tend to reproduce less quickly and tend to have smaller probabilities of being able to meet taxation demands.

2.4.3. *Sexual Reproduction*

Agents that interact and do not engage in combat may produce offspring through recombination. As in many genetic algorithms, the offspring replace the parents in the population. Sexual reproduction occurs between two agents A and B if A finds B

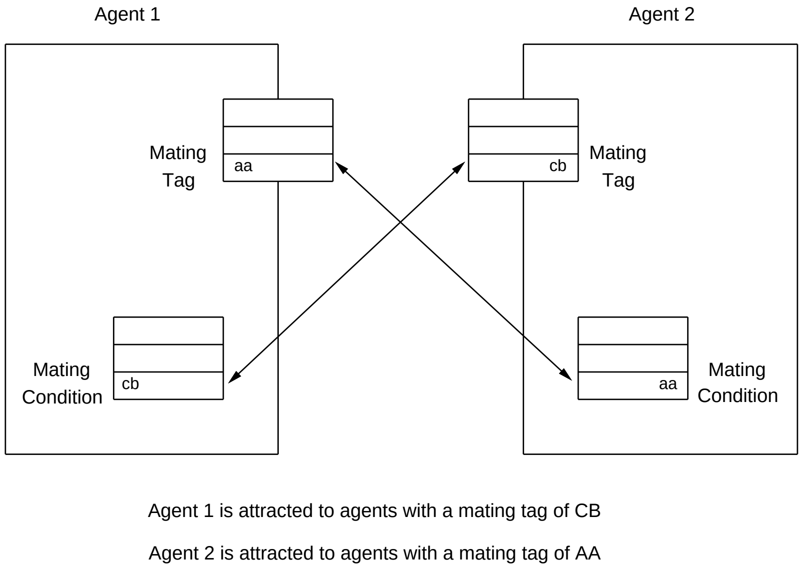

Figure 3: A simplified view of the two-way tag and condition matching that is used by agents to determine whether mating will occur.

acceptable and vice versa. A will find B acceptable if either 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 non-zero prefixes is designed to stop agents with zero-length mating conditions from rapid proliferation. 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 [4]. Figure 3 shows a simplified view of the two-way matching process used to determine whether mating will occur.

When sexual reproduction does occur, a form of two-point crossover is employed. This is complicated by the fact that agent genomes are variable length. Thus one can choose a crossover point in one agent and find that the same crossover point does not exist in the other agent. Without going into detail, two genes are selected to contain crossover points. Then the actual crossover points are chosen in each gene in each agent, and the crossover is performed. 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 will typically not be 50:50.

2.5. Agent Movement

There are two forms of agent movement in Echo: within a single site and between sites. Intra-site movement is the result of an agent-agent interaction. In each of these, one agent is first selected. A second agent is then selected in the vicinity of the first. The first agent is moved next to the second in the one-dimensional array of agents at the site. If the first agent would attack the second, the second may run away by moving a small distance away in the array. In both cases, distances are likely to be small, with the probability of a large distance being used falling off exponentially.

Inter-site movement occurs if an agent does not acquire any resources during an Echo cycle (either through picking them up, combat or trade). In this case it will migrate to a neighboring site, selected at random from among those permitted by the geography of the world.

3. Experimental Results: Species Abundance and Echo

In this section we present preliminary results comparing Echo populations with previous work on relative species abundance. Our overall goal is to confirm or disconfirm the hypothesis that Echo exhibits many of the same broad classes of behaviors as natural ecological systems. Because Echo emphasizes evolution, a natural starting point in the confirmation process is to ask whether or not evolution in Echo produces distributions of agents that are similar to or different from those observed in natural systems. Although we are still in the early stages of this investigation, our results to date are encouraging.

As we discussed earlier, it can be quite difficult to say what the individual components of a CAS model like Echo actually correspond to in the modeled system. To address the question of species abundance, for example, we need to define exactly what we mean by a species. The concept of species is not directly built into Echo, and there are a number of ways in which species could be defined. The simplest of these is to simply interpret individual Echo agents as species. A second interpretation, perhaps more appealing, is to group genetically related agents together in species. In the following we consider both of these interpretations.

3.1. Introduction to Species Abundance

Suppose we took the catch from a laden fishing boat returning to harbor and sorted the fish according to species. What would the distribution of fish into species look like? The answer, of course, will depend on many factors—weather, bait (if any), the depth at which the fish were caught, the water temperature at that depth, the size of the catch, and myriad others. Experiments of this nature have been performed many times by biologists, with samples of many sizes drawn from taxa including birds, snakes, fish, snails, lepidoptera, phytoplankton, arthropods, mammals and many others. A general perspective on such experiments is to consider the ways in which the n individuals that are sampled can be partitioned to represent a (typically unknown) number of m species. From a biological perspective, the interesting questions are: Does the distribution into species follow a pattern that can be characterized mathematically? And if so, are there biological theories that can account for this pattern? In many cases it is possible to fit mathematical models of distribution to observed patterns and to give plausible biological explanations for why these patterns should arise. See, for example, [14, 15, 16, 17, 18, 19, 20, 21].

A commonly observed phenomenon, is that the vast majority of species in a sample are made up of relatively few individuals. The conditions under which distributions of this kind are seen include early successional communities, environments perturbed by toxins or pollutants, and in appropriately sized samples [18, 22]. Relatively stable "climax" communities consisting of many species typically do not exhibit this qualitative pattern.

In examples where this general pattern is seen, Preston's canonical lognormal distribution has often proved the most accurate model, e.g. [23]. Preston [16] took the counts for the various species in observed data and grouped them into a series of "octaves." This was simply a (base 2) logarithmic grouping of the species counts. His octaves were labeled "< 1", "1–2", "2–4" and so on. Octaves were plotted on the x-axis and the counts of the species in each octave, a frequency of frequencies, was plotted on the y-axis. If a species count fell within octave boundaries, it counted 1 for that octave. If a count fell on the boundary between octaves, (as any count that is a power of 2 will), one-half was counted for the neighboring octaves.

Preston plotted these "species curves" for a number of experiments, and found that 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, to be determined from the data [16].

Because it is not possible to observe less than a single individual from a species in a sample, these distributions were truncated on the left at what Preston called the "veil line." As the distribution of octave counts is reasonably approximated by a normal distribution, the original species counts were postulated to come from a lognormal distribution. In particular, Preston found that the value of a that was calculated for the experiments he examined tended to be in the vicinity of 0.2. This gave rise to the "canonical" lognormal distribution of [19, 20]. In the canonical distribution the general lognormal distribution is reduced to a family of lognormal distributions dependent on a single independent variable. This relationship makes it possible to form good predictions of species relative abundance given only the number of individuals or the number of species [19, 21].

There are a number of conditions under which Preston's canonical distribution might be expected to arise, mentioned above. Alternative explanations for the occurrence of this distribution have also been advanced [24]. These range from arguments that such distributions are an artifact of the Central Limit Theorem, to simple statistical arguments. When these do account for the lognormal distribution, they fail to account for the fact that a wide range of experimental data is is not only lognormal, but also is close to Preston's canonical family of lognormal distributions. Sugihara [21] discusses these attempts and presents a biologically plausible alternative that generates the canonical distributions.

3.2. Species Abundance in Echo

In this section we consider different groupings of Echo agents, any one of which could be potentially considered a species in Echo. This section does not provide details of the various Echo worlds that have been observed to produce the effects described. To a large extent, this is not necessary as these can be seen in a wide range of Echo worlds in populations that are of reasonable size (roughly several hundred agents) that have been allowed to evolve for a reasonable number of iterations (at least two hundred cycles). In all of the figures of this section, the populations are taken from Echo worlds that were stopped after 1000 generations. The parameter settings that have been held constant throughout the experiments reported in this section are summarized in Table 1. Details on the precise meaning of these parameters are provided in [12].

These effects were also observed in earlier versions of the program in which several properties of the model were slightly different. In fact, Echo agents at one point managed to find and exploit a hole (bug) in the function that calculated the points agents receive in combat. When exploited, this typically results in an agent becoming relatively powerful and that agent and its offspring will tend to quickly dominate the

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
Number of Sites	1
Mutation Probability	0.02
Crossover Probability	0.7
Random Death Probability	0.0001

Table 1: The world and site parameters that were held constant throughout this section. Those above the line are the worldwide parameters. These parameters are described in [12].

world. Nevertheless these agents and those that found ways to survive, produced graphs of ranked genome abundance that were similar to those of the corrected program. All of this suggests that species abundance patterns in Echo are very robust.

The simplest way to study relative abundance in Echo is to sort the genomes by their abundance, and to plot these by rank on the x-axis and by number of individuals on the y. This was the method used by MacArthur [17, 18] and the data shown in Figure 4 are similar to his graphs. This figure was produced by simply examining the number of copies of individual genomes in the population after 1000 generations of an Echo run.

Taking the population data from the same Echo run and organizing it into octaves using the method described by Preston [16] results in Figure 5. This figure bears a strong resemblance to those of Preston, especially those in which the veil line is close to the mode of the distribution. It is clear that the character of genome abundances in Echo populations tends to follow the general patterns found in some biological systems. The question is how close is the correspondence.

There are two important aspects of this correspondence: (1) how Echo agents are grouped into "species, " and (2) how the result is sampled. In Echo, there is no *a priori* grouping, one has to be defined. We have already seen the simplest case, in which each genome is considered a group, and that this gives rise to graphs that resemble those of biological systems. We have considered several possible strategies for grouping, including clustering based on genetic distance (e.g., see Figure 6), clustering based on functional properties (agents that act alike are grouped together), and clusterings based on evolutionary history (agents that evolved together are grouped together). Here, we examine groupings based on genetic distances, and we use a simplistic method of deciding where to impose "species" boundaries between clusters.

The second dimension is of great importance in both biological systems and in Echo. Sample size (and location) can completely determine whether distributions such as those shown will appear. This has been mentioned in virtually every work cited in this section. It may be the case that a very large sample does not exhibit certain properties, but if that sample is divided into a set of smaller samples at random, then each of the smaller samples will show the highly skewed distribution. The locality from which the sample is drawn will also have a great affect since most species show considerable variation in relative density over their entire range of habitats. Thus even

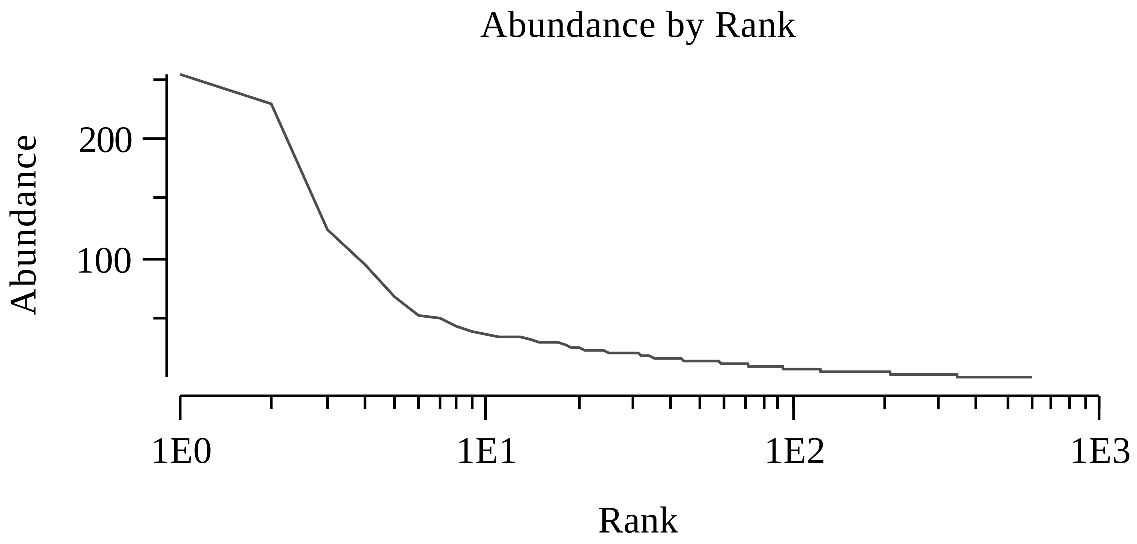

Figure 4: An example of the abundance of Echo genomes in a population after 1000 cycles. Abundances are ranked from commonest (left) to rarest (right), with the actual abundance given on the y-axis. The final population contained 603 different genomes.

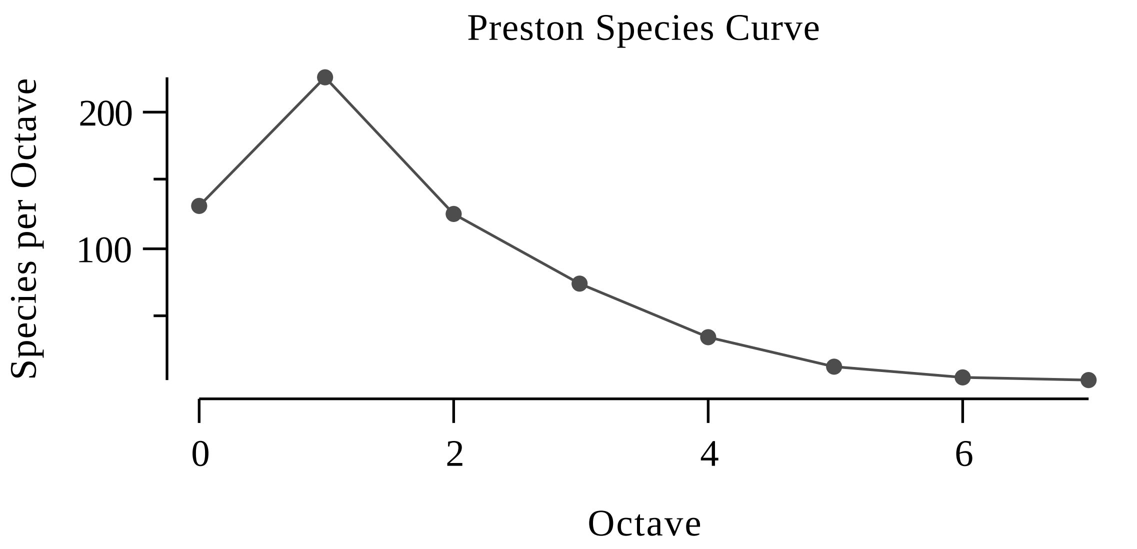

Figure 5: The population data from Figure 4 organized into octaves according to the method of Preston [16].

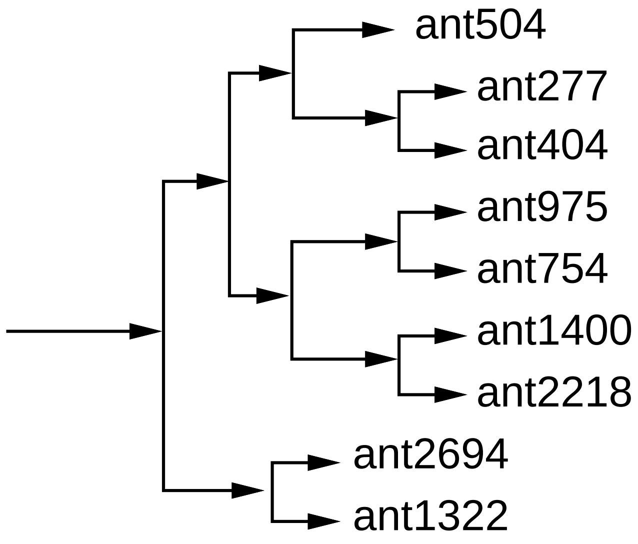

Figure 6: A fragment of a cluster analysis of Echo agents based on genetic distance.

if all species contained exactly the same number of individuals, this variation could produce a skewed distribution if the sample size were small relative to the total number of individuals.

In Figures 4 and 5 there is no grouping and no sampling. As a result, all curves for Echo derived in the method of Preston will have a mode of one, since every single individual is present in the data and there is great variation at the level of individual genomes. Such sampling is rare (but not unheard of) in biological systems.

Using the minimal number of mutations required to transform one agent into another as a distance metric, we used a hierarchical clustering algorithm to cluster the genomes of populations. At each iteration, the clustering algorithm locates the two clusters at minimum distance and combines them. By imposing a maximum on this distance, the algorithm can be restricted from proceeding all the way to a single giant cluster. We then consider each of the clusters that has been formed to represent a species. When the limit is reached, any agents that have not been included in a cluster will be considered singletons—the sole representatives of a species. Table 2 shows the number of species that are produced from three bounded clusterings of three example Echo runs.

Figure 7 plots an example of the data in Table 2. The curve was obtained from the experiment in which the site produced 300 units of each resource in every Echo cycle. Here the clustering algorithm was prevented from combining clusters with an average distance of greater than 20. This can be compared to Figure 8 which shows exactly the same experiment (i.e. started with the same random seed) but with the clustering limit set to 10. There are several differences between the graphs that are not difficult to account for. The first has a larger number of octaves expressed, which is a direct result of grouping agents into fewer categories (140 species as opposed 462, as shown by Table 2). On average, categories will tend to be larger and thus more octaves will be represented. The heights of the modes of the two figures also differ considerably. This is to be expected since the higher clustering distance limit will gather more singletons into clusters before halting. This results in far fewer species falling into the lower octaves. The first figure, with the higher clustering limit, more closely resembles the figures

Cluster limit	Resource level	Total species	Non-singleton agents	Non-singleton species	Singleton species
10	100	187	128	43	144
	200	294	214	72	222
	300	462	235	95	367
15	100	77	225	30	47
	200	149	331	44	105
	300	303	410	111	192
20	100	19	265	12	7
	200	50	412	26	24
	300	140	514	52	88

Table 2: The number of species resulting from different bounding conditions on genetic clustering of Echo agents. The experiments all consider the same world with differing resource levels provided by the site. The sizes of the final populations in the three experiments were: 1191, 2388 and 3509.

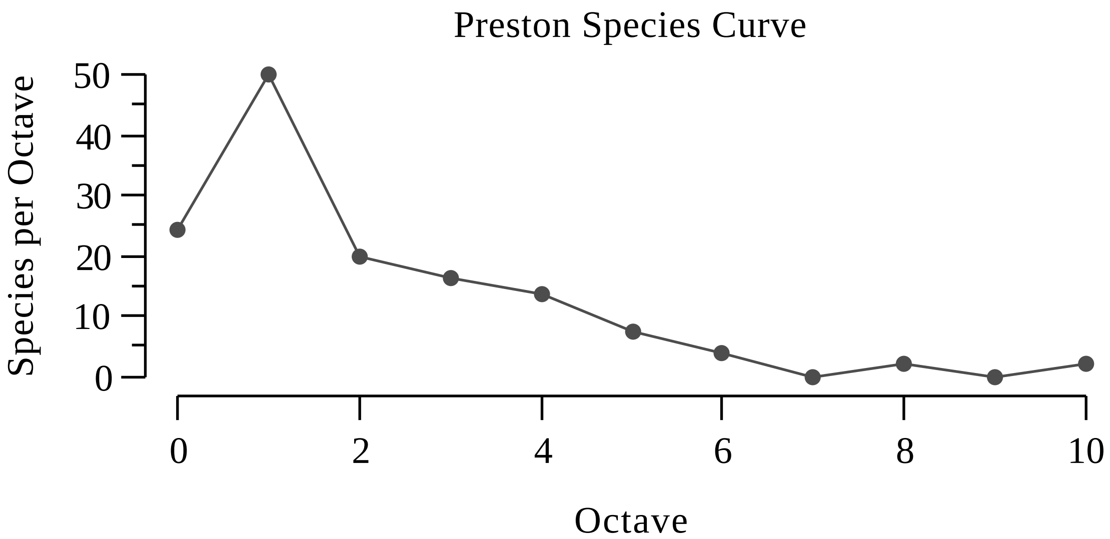

Figure 7: The species curve resulting from genetic clustering of 3509 Echo agents. Clustering was restricted to distance 20 or less.

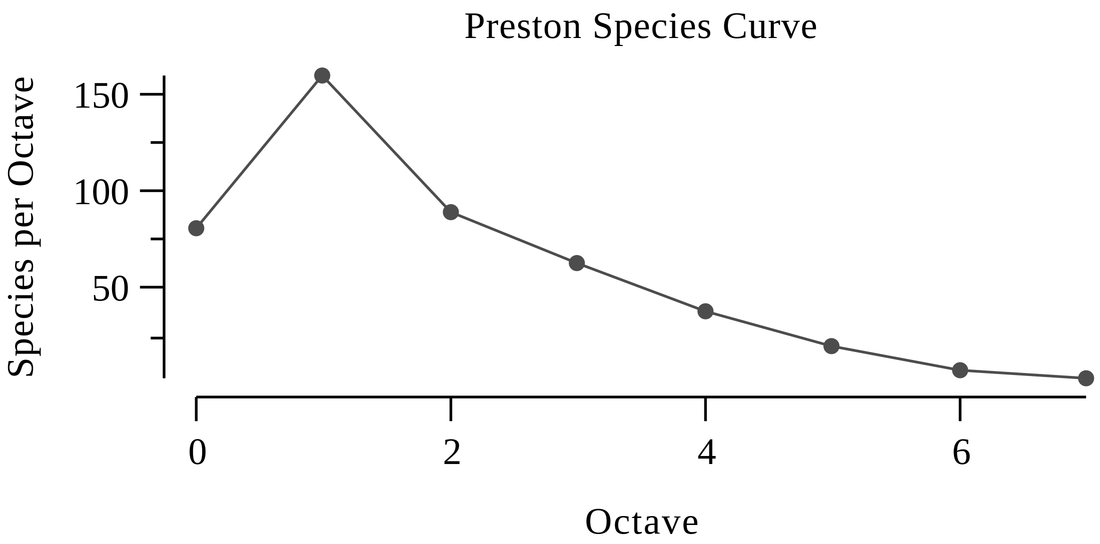

Figure 8: The species curve resulting from genetic clustering of 3509 Echo agents. Clustering was restricted to distance 10 or less.

found in [16]. The clustering method, in all the cases examined, reduces the height of the mode of the species curve significantly.

We tried a simple sampling method (results not shown), which does not appear to produce any change in distributions. In it, each agent in the population is sampled with some fixed probability. However, we expect that sampling based on Echo's geography will produce marked changes.

4. Conclusion

Our preliminary work on species abundance is encouraging, and there are several directions in which we plan to extend it. These include deciding how to limit the clustering algorithm based on population size; examining other methods for grouping, in particular clustering based on agent behavior and evolutionary history; investigating sampling methods; and finally, fitting Echo data obtained from different choices of grouping and sampling to that of the various models of relative species abundance. These directions are not independent. The extent to which Echo data will fit existing work on species abundance will, as described above, depend on how species are delineated in the model and on how populations are examined. Given the tendency for this qualitative behavior to be present in several different versions of Echo, it seems likely that there will be no single correct answer. Rather, we expect to identify some perspectives on Echo that are most appropriate for modeling biological ecologies.

Examining species abundance is our first formal step in the validation of Echo. Informally, a number of interesting phenomena have also been reported, such as the evolution of "arms" races. This suggests that Echo is quite a rich system. Our approach to validating Echo as an ecological model is to perform a series of small experiments, each of which is designed to explore one aspect of Echo's behavior. If the system performs realistically on this set of experiments, we will have much more confidence in Echo's relevance to real-world systems. We believe that such a validation will increase the confidence with which the model can be applied.

It will be a long time before models like Echo can be used to provide quantitative answers to many questions regarding CAS. A more realistic goal is that these systems might be used to explore the range of possible outcomes of particular decisions and to suggest where to look in real systems for relevant features. The hope is that by using such models, people can develop deep intuitions about sensitivities and other properties of their particular worlds. High-level knowledge of this kind could be very valuable. In many CAS, a small increment in intuition would translate into large gains. For example, even a very small increment in our intuitions about the likely behavior of some aspect of the economy or environment could be used to great effect. We view Echo as an early step in the building of CAS models. The process of validating such models is a daunting task. We hope that by examining carefully the model's behavior we will learn lessons that are also valuable to the development of future models with similar aims.

5. Acknowledgments

This work was supported by the Santa Fe Institute (Project 2050) and by the National Science Foundation (grant IRI-9157644). John Holland conceived of and designed the original Echo model on which our work is based, and he suggested many of the ideas in this paper. James Brown and Timothy Keitt, both of the Biology department at the University of New Mexico, have both 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.

References

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

[2] 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.

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

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

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

[6] S. Forrest. *Parallelism and Programming in Classifier Systems.* Pitman Press / Morgan-Kaufmann, London, 1991.

[7] 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.

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

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

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

[11] 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.

[12] T. Jones and S. Forrest. An introduction to SFI Echo. Technical Report 93–12–074, Santa Fe Institute, Santa Fe, NM, 1993. Available via anonymous ftp from ftp.santafe.edu:pub/Users/terry/echo/how-to.ps.Z.

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

[14] A. S. Corbet. The distribution of butterflies in the Malay peninsula. *Proceedings of the Royal Entomological Society of London (A)*, 16:101–116, 1942.

[15] R. A. Fisher, A. Steven Corbet, and C. B. Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12:42–58, 1943.

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

[17] R. H. MacArthur. On the relative abundance of bird species. *Proceedings of the National Academy of Science*, 43:293–295, 1957.

[18] R. H. MacArthur. On the relative abundance of species. *The American Naturalist*, 94:25–36, 1960.

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

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

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

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

[23] C. B. Williams. *Patterns in the Balance of Nature.* Academic Press, London, 1964.

[24] 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.