A Theoretical Framework for Abundance Distributions in Complex Systems

Stephan R. P. Halloy
A theoretical framework is proposed to explain how and where complex systems break up into agents or species. Splits lead to diversification and to abundance distributions which are similar to power functions on a rank-abundance representation, and to lognormal functions on a frequency-abundance representation. The combined manifestation of power and lognormal functions is a polo distribution, a situation toward which there seems to be a widespread tendency in complex systems (a polo pattern attractor).

Minimal complex system organisation requires three integrated hierarchical levels, the system, agents and particles. The tendency to polo emerges, or can be explained by, resource particle interaction, in which particles are attracted to each other according to their size and inversely to their distances. Simulation of this simple rule on a preliminary model leads toward polo abundance distributions. The level of abstraction allows the theoretical framework to be applicable to all fields where complex systems are found to have polo distributions. A clearer understanding of the rules and forces leading to diversification can have a range of applications in planning and management for conservation, agriculture, business, health and other areas dealing with complex systems.

1 Introduction

Evolution explains the mechanisms by which organisms change (mutation, recombination, selection, developmental constraints), but how taxa split, and the causes and timing of diversification are not clear. Both biological and economic systems are characterised by trends of increasing diversity. The abundance distributions of the elements (species, businesses, agents) of these systems tend toward characteristic abundance-rank and frequency-abundance patterns. The patterns approach a power function in the first case and a lognormal distribution in the second (polo for short).

Polo distributions are found in many complex systems, particularly with regards to aspects of size (volume, length, biomass) in what has been called the Dyar-Hutchinson rule by May (1978), referring to living systems (Raunkiaer 1934; Hutchinson & MacArthur 1959; Halloy & Mark 1996). Similar distributions are also known in inanimate systems (e.g. nanoparticle sizes, Soderlund et al. 1998).

First I present a theoretical framework to attempt an explanation on how and where splits occur in a complex system, thus leading to both diversification and polo distributions. Then a
simple resource attraction model is proposed to study the implications and effects of the theoretical concepts. The model is based on simple rules, allowing it to simulate reality and be tested.

2 Background

Many different models have been proposed to fit abundance distributions. These range from descriptive (e.g. a mathematical curve that fits empirical data) to attempting theoretical explanations of mechanisms. Preston (1948, 1962, 1980), MacArthur (1957, 1960) and Sugihara’s (1980) lognormal models assume a flat system with no vertical hierarchy and no links (e.g. the canonical lognormal or the broken stick model). Kauffman’s (1993) NK models involve a number of elemental components linked in ways that lead to global system behaviour, without hierarchies. Bak et al.’s (1988) self-organised criticality depends on the variations in size of agents by particle accretion, leading to a distribution of break-points (avalanches or catastrophes) with a power or 1/f distribution in a cellular automaton or topological structure. Accretion is externally determined. Each cell can accumulate particles to a certain threshold after which it ‘collapses’, sending particles to neighbouring cells. If all cells are close to the threshold (critical state), this can lead to a large domino effect. The distribution of ‘avalanches’ tends to a power function. Avalanches represent the agents.

Many more attempts have been made to explain polo distributions on a case by case basis, focussing on a particular field of knowledge (e.g. Barlow 1994; West et al. 1997). The latter models do not take into account the universality of polo distributions. None of the above models has proposed a conceptual framework to serve as a theoretical basis applicable to seek the mechanisms of diversification from an undifferentiated start.

3 Defining the processes and elements of a complex system - toward a theoretical framework

3.1 Processes

In an initial amorphous mass of resources (a ‘simple’ system), breaks may arise through unevenness of attraction. Such a process has been hypothesised more simply in the broken-stick type models mentioned above. Fragmentation at random as in the broken stick, or in crushing an object, tends to produce a lognormal distribution of fragment sizes. Such models and real systems imply external forces and no explicit relationality between elements. Here I postulate that the attraction between resource particles (interactions or links), in combination with some stochastic variation in their sizes, positions or both, leads to a rupture of the amorphous mass and clustering. This clustering changes the amorphous mass into a system with agents (the clusters) separated by boundaries where resources are rarer. The force of attraction is proportional to the mass of the resources, leading to a positive feedback. As agents grow, they attract ever more resources. But the attraction is also inversely proportional to distance. All agents in complex systems can be seen to respond in some ways to this process of attracting resources in proportion to their magnitude (in whatever units this may be measured) and inversely proportional to some measure of distance or difficulty to obtain that
3.2 Elements

**Complex system structure.** Consider two basic types of systems: simple and complex. Simple systems can be thought of as a group of undifferentiated particles. There are two hierarchies: the particles and the system. Complex systems must have a structure requiring a minimum of three hierarchical levels: particles, agents and the system (figure 1). Clumping into agents is the result of the variable attractions between particles described above. Flexible boundaries delimit agents. I postulate that the size of such agents tends to a polo distribution in all complex systems where agents exchange resources in some way (relationality, competition or as used here, attraction) and agents are capable of indeterminate growth. The polo is thus a signature of complex systems.

There are two levels of complex systems according to whether significant amounts of resources flow through the system (e.g. dynamic biological and economic systems) or not (e.g. static physical systems such as the solar system). The static complex systems are at the lower end of the scale of complexity: the pattern is frozen. If resources flow through the system, it continues to evolve into a complex dynamical system. The boundary between simple and complex systems must be like a phase transition. Thus the distinction should be relatively clear in actual systems, constituting a test to the definition.

Systems where the growth of agents is constrained through some information (e.g. size of animals in a population, size of sand grains on the beach) tend to have a normal distribution of the elements and no hierarchy, whether the underlying agents are complex or simple. This in turn depends on the character measured. The human population is normally distributed for size, as size is genetically constrained, but may in some societies be lognormally distributed for wealth when this is not culturally constrained.

![Figure 1. Minimum necessary structure of a complex system.](image)

---

System 🟢 Agents 🌈 Particles 🔹 Interactions →
Particles are the minimum units of resources. From an agent’s viewpoint they are discrete packages of resources of variable size or ‘mass’. Particles are analogous to individuals in a biological population, to quanta of light or space in a plant community, to particles of dust in the cosmos, or to economic elements. In an economic system the component elements are often called agents, but this term should not to be confused with agents as defined here except when it refers to companies.

Resources. Any thing or process for which agents may compete. Resources are designated here as an abstract Mass (M), which can equally be seen to mean actual physical mass, biomass, space, time, energy, financial resources, etc.

Agents constitute the intermediate hierarchical level, which qualitatively distinguishes a complex system from a simple system. Agents arise when an initial undifferentiated mass of particles breaks up or coalesces (i.e. boundaries are formed) into a number of parts. Each agent contains or controls a number of particles. Agents are analogous to species or companies. The system boundary contains all the particles and all the agents.

Boundaries are formed where interactions are proportionately more important between the particles inside the agent than they are between them and particles outside. The same applies for boundaries between systems at a higher level. Boundaries fluctuate and have a certain degree of permeability. This is a fundamental aspect of complex systems which is often overlooked and leads to problems of sampling and definition (e.g. where are boundaries set from our perspective), and of successive nested hierarchical levels of complexity which vary in space and time (e.g. Frontier 1985, or the SWARM simulation system, Langton et al 1998).

3.3 Pattern attractor

Currently defined attractors are based on systems being represented by a single point moving through phase space (Wuensche and Lesser 1992). It is necessary to define here a new type of attractor. Pattern attractors are defined as attractors where the system can only be represented by the relative position of more than one point in phase space. The phase space here is defined by two axes: frequency or rank, and abundance or magnitude. The series of points and their positions creates a pattern which defines the system’s pattern phase state. The lognormal, or any other frequency distribution could be considered as possible pattern attractors. I postulate that the polo distribution is the main pattern attractor for the abundance distributions of complex systems.

By providing this definition, we circumvent the debate on the appropriate mathematical distributions to fit to natural systems, a question which is untestable as there are infinite numbers of natural systems, each modified by its own history. The question is not what distributions fit what empirical data in a given time slice, but toward what curves distributions may be tending. Comparative studies showing variations of goodness of fit to particular models will be more informative than a particular goodness of fit. Thus many natural distributions are rather bad approximations to lognormal models. However, if natural systems consistently approximate to lognormal models when left to their internal mechanisms, while distancing themselves from the lognormal when pressured by external forces, then we can
suspect the presence of a lognormal pattern attractor. Considering a mathematical curve as an attractor becomes a testable hypothesis.

3.4 The Polo Pattern Attractor

The widespread occurrence of polo distributions, and the fact that in many cases where distributions shifted away from a polo pattern return toward that pattern, strongly suggest that the power function, the lognormal distribution, or their combination in the polo distribution act as pattern attractors for complex systems. The power functions and lognormal frequency distributions described in the literature cited above are mathematically distinct. However, both distributions are present in the same natural situations given certain restrictions on scale (Sugihara 1980, Frontier 1985, Tokeshi 1993). A data series with a lognormal distribution will also exhibit a power function rank distribution for the right part of its range, e.g. when its left side is veiled. Lognormal distributions found in nature are generally canonical and are often veiled (May 1978; Preston 1980; Magurran 1988; Brown and Maurer 1989; Brown and Nicoletto 1991). Conversely, empirical data series fitted to a power function (Zipf 1949; Mandelbrot 1983; Magurran 1988) also exhibit a pronounced drop or convexity at the lower right side, a distribution which resembles an exponential function but often indicates lognormality.

The distinction between exponential and lognormal can be seen in a frequency-abundance representation, where the exponential produces a straight horizontal line and the lognormal the typical bell-shaped curve.

The ubiquitous occurrence of polo distributions has often been ignored or downplayed. For example, Solé and Alonso (1998) state ‘species abundance follows a power-law distribution and not a log-normal one, as it is usually assumed.’ This distinction between the two patterns is due to several causes:
- a large array of nomenclature for different expressions describing what is at least in part the same phenomenon, for example, power functions, fractals, 1/f noise, allometric species relations, Pareto principle, lognormal, etc.
- a multiplicity of different representations of abundance distributions (Tokeshi 1993)
- the amount of ‘noise’ in many natural systems, leading to most data only approximating power or lognormal distributions.
- sampling problems: where samples are too small the right part of the lognormal can be identical to a power function.

In their particular fields, polo patterns (in either of the forms stated above) have been variously used to determine the optimum business and marketing strategy for a company (e.g. Pareto principle) or to determine climates on the basis of vegetation structure and vice-versa (e.g. Raunkiaer 1934). Polo patterns have been proposed as diagnostic indicators of ecosystem health (Gray 1979 cited in Frontier 1985; Kevan et al. 1997), predictors of vegetation changes and management tools for environmental risk assessment (Schmoyer et al. 1998), determine sustainability of agricultural systems (Halloy 1994) and to calculate inputs needed to maintain a system away from its ‘harmonic’ polo pattern (Halloy 1997). Such studies implicitly accept that the polo pattern acts as a system pattern attractor. Predictive powers could be refined if we could standardise representations and define the fundamental unifying principles leading
toward the polo distribution. Then we will also be able to search whether there are some other pattern attractors distinguishable as power, lognormal or other. The challenge is not why complex systems have a polo distribution of abundance, which they often do not, statistically speaking, but to define and quantify the forces that push complex systems toward a polo distribution and how they function.

4 A Resource Attraction Model to Explore the polo attractor.

A resource attraction model (RAM) was developed for the exploration of the theoretical framework described and to test whether these assumptions would lead to distributions of agent sizes approximating lognormal or power distributions or both. The model would also allow the exploration of sensitivity to initial conditions and variables and to compare its dynamics with empirical data.

4.1 Description

The model is based on a one-dimensional linear topology, but in principle can be expanded to more dimensions. Resource ‘masses’ are situated along this line with their position defined by a single coordinate. Each resource particle will then have a distance (d) from other neighbouring resources. The coordinate line can be made circular to avoid edge effects or a buffer can be set up. Space and distance here are fundamental abstractions applicable to any system. For example, in a biological community, links between individuals of all species are to some extent a function of spatial distances, but often more importantly of temporal, energy, matter and information ‘distances’.

Following the theoretical postulates, the model assumes that masses attract each other and that the force of this attraction is a function of their mass and an inverse function of their distance. The number of links (L) between particles or agents is a function of the individual force of attraction of each particle (determined by its mass), and its distance to other particles. In theory, L is infinite as all particles can affect each other even at great distances. As only those links which are above a certain strength are important to the agent dynamics, a subset of links to nearest neighbours represents effective links, or Le.

Given an initial distribution of the two variables, particle size (M=mass) and particle position (coordinates), the subsequent positions are calculated by the attraction. The attraction or pull is calculated by the gravity analogy of \( M_1 \times M_2 / d^2 \), where \( d \) is the distance calculated between the particles. At each iteration, masses move according to the magnitude and direction of this pull. The movement of particles to new cells results in clustering to varying degrees, depending on variables, thus forming agents (figure 2).
**Figure 2. Simplified linear resource attraction model. Mean particle size= 1, range of variation= 0.1 and two links (i.e. one on each side).**

The simulation in figure 2 leads to a frozen distribution in just 4 steps, with agents oscillating in some cases between two positions. At time 1 one agent (on coordinate 4) has formed by accretion. Particles in coordinates 6 and 7 have swapped over, both remaining at a value of 1. Circularity is provided by the particles on the extreme left interacting with the particles on the extreme right. The rapid freezing is similar to some physical systems such as the solar system where no substantial quantities of new matter and energy enter the system, compared to the amount already there.

A more realistic situation for dynamic biological systems can be achieved by adding a flow of resources. This can be done by adding and subtracting resource particles at each step distributed uniformly or at random and with a given degree of size variation. This rain of particles simulates a continuous flow of energy or matter that leads to an increase in organisation or maintenance of a characteristic pattern in dynamic systems. With this addition the system remains dynamic. Agents evolve and become extinct, yet the abundance pattern still revolves close to a polo.

### 4.2 Basic rules, inputs and outputs

**New Mass calculation:** Actual mass x relative mass + mass attracted from neighbours + rain - losses

**Attraction calculation:** \(M_1 x M_2 / d^2\)

**Relative mass:** \(M_1 / \text{Sum of neighbouring masses}\)

**Particle rain size:** random or fixed around value decided

**Proportion lost:** \(M_1 \times \text{loss coefficient}\)

The input variables that the operator can modify are:

- total mass of resources
- total universe size (i.e. coordinate space)
- spatial distribution of particles
- number of effective links
- magnitude of the distance exponent
- resource rain, form of distribution and calculation
- particle mass, mean and range (i.e. variability of resource flow)
- resource loss (proportion)
- number of iterations
- minimum viable agent size

Additional constants and variables can be added and modified to explore and fine-tune the model to particular types of systems. Such variations provide ‘vibrations’ that can tip agents out of one cell, catastrophes that can disrupt or destroy them, or slight changes that may help them capture new particles.

The output parameters include:

- mean and deviation of the frequency distribution
- abundance-rank pattern and goodness of fit to power or other functions
- frequency-abundance pattern and goodness of fit to lognormal or other distributions
- total mass
- total agents
- agent dynamic fluctuations over time

5 Results of Simulations

The dynamics of the simple rules described above allow the exploration of the rule-space which leads to diversification. Results vary according to initial conditions, but remain similar within a wide range of situations. For example, the model can simulate the emergence of diversity from an initial amorphous mass of particles with uniform spatial distribution and some (e.g. 20%) variation in particle mass. Particles start to cluster into agents. Larger agents capture neighbouring particles or whole agents (take-over). Smaller agents grow slowly or become extinct. Occasionally, new agents arise (speciation) (figure 3).

The abundance distribution of agents tends to a power function with increasing slope toward the right in a log-log rank-abundance relation or a lognormal (figures 3-5). Also as in natural systems modifying the variables can lead to a veiled or truncated lognormal, which is then almost identical to a power function (see references under 3.4).

Despite the apparently chaotic dynamics of individual agents’ growth and decline, the polo distribution attractor is robust for the RAM. Abundance distributions trend toward this attractor for a range of different variables, in many cases after very few steps (5-20). The similarity of two events arising from different initial sets of conditions in the rank-abundance graph (figure 4), but the clear distinction of the same two in the frequency-abundance graph (figure 5, □ = dotted line of figure 4, ▼ = bold line), suggests that the lognormal has greater diagnostic capability than the power function to discriminate the variables in action. For example, in systems which are large enough in relation to the particles, the dip in slope at the right end of the rank-abundance distribution may become an independent lognormal curve, showing that the system has split into two systems (e.g. as shown in the case of macro- and micro-economic plants in New Zealand agriculture; Halloy 1994, 1998). Such splits are clearly apparent in frequency-abundance representation but not as obvious in abundance-rank representation.
Figure 3. Time sequence of agent development. Linear model starting from an initial uniform distribution of particles of randomly varying sizes, with particle rain at each step. Mean particle size 1, range 0.2, total mass 200-264, 10 links. Dotted: t5, striped: t10, full: t20.

Figure 4. Rank-abundance distribution of agents. Agent size is the mass of resources captured by that agent. Bold line: with particle rain, mean particle size 1, range 0.2, total mass 200-259, 20 steps, 10 links giving 45 agents of mass > 2.8; dotted line: without particle rain, mean particle size 1, range 2, total mass 210, 10 steps (frozen), 10 links giving 33 agents of mass > 2.8.
Figure 5. Frequency distribution of agent sizes. Frequency shows the number of agents with a given resource mass or size. Size shows the resource mass in each class. 

X: without particle rain, mean particle size 1, range 0.2, total mass 200, 10 steps (frozen), 10 links giving 37 agents of mass > 2.8; 

●: without particle rain, mean particle size 1, range 2, total mass 210, 10 steps (frozen), 10 links giving 33 agents of mass > 2.8; 

▼: with particle rain, mean particle size 1, range 0.2, total mass 200-259, 20 steps, 10 links giving 45 agents of mass > 2.8; 

▲: with particle rain, mean particle size 1, range 0.2, total mass 200-264, 20 steps, 10 links giving 50 agents of mass > 2.8; 

■: mean of two last series.

Any distribution which starts as an orderly spatial array (e.g. an arithmetic progression: 1, 2, 3, 4, 5...), ends up concentrating all resources on one coordinate; there is no diversification. Thus, the way the particles are distributed with respect to their links (i.e. space-distance) is of importance to the development of diversification. A completely uniform initial distribution (i.e. all 1, or all 2) remains frozen in that same condition.

6 Discussion

The theoretical model and the simple simulation described in this paper are clearly distinct. The theoretical framework attempts to define the fundamental components and concepts relating to complex systems. The computer model is an extension of a ‘pencil and paper’ exercise to visualise the consequences of the theory. The model is only one of a range of different approaches to test to the theory.
6.1 Links

Kauffman (1993) suggested that in a Boolean network the relation between number of links and number of elements was critical to the system dynamics. Only intermediate values lead to polo distributions. Too few links would make the system vary at random within a normal distribution (i.e. particles clump in random group sizes according to an initial random distribution). Too many links would freeze the system whichever way it started (historically determined). The RAM tends very often to polo, so how is the ratio of effective links to number of particles \(L_e/N\) close to the intermediate values needed? In Boolean networks, the number of links is a discrete number imposed from the outside. In the RAM the magnitude of \(L_e\) is a continuous function of the force determined internally by the rule of attraction. Boundaries to agents and to the system arise at the distance where the internal attraction reaches a certain ratio to external attraction. The boundary encloses \(N\) particles or resource mass. In this way \(N\) and therefore \(L_e/N\) are set within certain bounds by the internal rule.

6.2 Relationality: Attraction, repulsion, competition, cooperation

The theoretical framework may help clarify some debated aspects of the behaviour of complex systems.

The rule for attraction is analogous to gravity in a physical system, reflecting the observable fact that agents attract or pull in resources to grow whether they belong to physical, biological or economic systems. The agent’s attraction is a function of its mass and distance. The laws of increasing returns and of diminishing returns in economics are some of the manifestations of this phenomenon in the practical world. The law of increasing returns expresses the positive feedback between the size (e.g. market share) of a company and its increasing ability to capture more resources (see Arthur 1994). The law of diminishing returns observes that as an agent accumulates resources from a particular source, the initial accumulation is cheap and easy, but as the resource is depleted and/or competitors arise, capturing the resources becomes increasingly costly. In the terminology of this model the resource is initially clustered, proximate and dense leading to a high attraction, but at a later stage more distant and dispersed, with a higher entropy, and therefore more difficult to attract.

The attraction force can also be modelled as a repulsion force, as one is the inverse of the other (e.g. Douady and Couder 1992). Other phenomena arise from such a force, without having been explicitly included. The notions of competition and cooperation arise as a result of attraction rather than as incompatible theoretical options. Agents compete to attract resources in between them, and cooperate to attract resources on either side of both.

6.3 Hierarchy and diversification

An aspect of behaviour that can be inferred from the framework is that once evolved to a large size an agent can, in turn, become the basis for further differentiation on a new hierarchical level, and levels can cross over. For example, socio-economic systems arise as agents (e.g. nation-states) which break up into a series of new agents (provinces on a spatial scale, companies and ethnic groups on an economic and cultural scale) or join into super-systems such as nation-blocks or alliances.
6.4 Different one-dimensional self-gravitating models

Physicists have developed one-dimensional self-gravitating systems to explore multi-bodied dynamics of attraction (Cuperman et al. 1971; Rybicki 1971; Yawn and Miller 1997). Such models differ from the resource attraction model in that particles accelerate until they either bounce elastically or pass through each other. There is no accretion into larger agents. The dynamics of diversification and abundance variations which occur in the resource attraction model do not occur in these physical systems.

6.5 Convergence with other models

It is no coincidence that many models produce polo distributions starting from many different explanations (e.g. Zipf 1949; John Conway’s Game of Life, Berlekamp et al. 1982; Barlow 1994; Pahl-Wostl 1995), as the trend results from similar fundamentals such as neighbourhood interaction or competition for resources. Bak (1997) suggested that such models evolve to criticality, as do natural systems (e.g. Lockwood and Lockwood 1997). Whatever the underlying principles, there is a need to describe these in a way valid for all systems. For example, the analogies with fundamental quantum theory are intriguing (‘quanta’ of energy or mass, and attraction) and are may not be coincidental. These analogies reflect deeper mathematical laws (e.g. Tegmark 1997; Chown 1998). The theoretical framework proposed here is one attempt to identify such fundamental rules in a way which can be tested across natural systems. The resource attraction model behaves robustly because the rules define the system at each level independently of the user.

6.6 Issues to explore

The resource attraction model allows the exploration of a variety of important questions concerning the behaviour of complex systems. For example:
- explore the limit to diversification or polo: are there limits to diversification or to polo? If there is a limit, what is it? are most systems at the limit? if not, why?
- explore the pattern attractor phase space by modifying rules and variables (e.g. the variables leading to particular slopes or veil lines, conditions for splitting and diversification, or the collapse of diversity)
- under what conditions is the system most stable?

7 Conclusion

A theoretical foundation is proposed to explain the features and mechanisms of power and lognormal distributions so widely found in nature. The framework is based on fundamental elements which are observable in nature (e.g. attraction, distance), which lead to the breakup of resources into new agents (speciation, diversification) and emergent mechanisms to set boundaries. The level of abstraction allows the model to be applicable to any complex system fitting the proposed definitions.

The model proposes that relatively simple rules lead to the emergence of polo distributions in complex systems starting from a range of initial conditions. The prevalence of polo
distributions suggests that they could be used as management tools in conservation, agriculture, economy and other disciplines dealing with complex systems (e.g. to determine ecosystem health, to predict agricultural inputs, to pre-empt weed invasions or to determine the cost of an egalitarian society). The fact that it is not used is to some degree due to the lack of an appropriate theoretical framework, to a confusion between fundamental rules and noise, and to non-standardised language leading to lack of communication.

Acknowledgements

I am grateful to B. Barratt, K. Dodds and J. Lockwood for allowing me to sound out these theories and discuss drafts of the manuscript; also to the editor, Russell Standish and an anonymous referee for helpful comments. I would also like to thank E. Jen for inviting me to visit the Santa Fe Institute where further insights were achieved.

References


Simulation System, Santa Fe Institute, Santa Fe, New Mexico.


