OBSERVING ARTIFICIAL LIFE EVOLUTION: A FORMAL ALGEBRAIC FRAMEWORK

JANARDAN MISHRA

(M.Tech., Indian Statistical Institute Kolkata)

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

SCHOOL OF COMPUTING

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgments

It was a long and memorable journey with deep lessons to learn during the course of the work. I would like to express my gratitude and thanks to Dr. Martin Henz for his patience to read my drafts and listen to me for long hours and all the support he extended on administrative matters as my supervisor. His inputs during the weekly discussions were of essential help in carrying the work forward. Working in a research area such as Artificial Life has indeed been a mental voyage into a deeply mysterious phenomenon of life with sometimes disturbing philosophical realizations regarding our own existence and meaning.

I can hardly forget intellectually delighting talks with Daniel Högberg and the help, which he extended to me time to time. Indeed some of the clarifications on the philosophical insights came to me only during our discussions. Discussion with J. Vedvyas and Bee Peng on their honor's year projects under Dr. Martin were so important to learn the field in a comprehensive manner. I feel fortunate and thank Dr. Martin for giving me the chance to work with them.

For the proof reading of my drafts I could have found none better than Hanna Kurniawati as well as Hugh Anderson.

Last but not least, my deepest gratitude goes for all the moral support and emotional comfort given by my fiancee Anjeline Daniel, who even though staying so far off never let me feel tired while working on the thesis.

Contents

Summary

Establishing the presence of evolutionary behavior as a defining characteristics of 'life' is a major step in Artificial life (ALife) studies, though seldom specified explicitly and formally. We present in this thesis a general abstract algebraic formal framework for this aim. The framework is sufficiently generic to be applicable to a wide variety of ALife studies, and does not depend upon the low-level dynamics and the structure of the underlying model universe.

The framework is based upon the notion of high-level observations made on the ALife model (chemistry) at hand. An observation process is defined as a computable transformation from the underlying dynamic structure of the chemistry to a tuple consisting of abstract components needed to establish the evolutionary processes in the chemistry. Starting with defining entities and their evolutionary relationships observed during the simulations of the model, the framework prescribes a series of definitions, followed by the axioms that must be met in order to establish the level of evolutionary behavior in the model.

The framework is defined with the assumption that presence of life-like phenomena in any ALife model requires that evolutionary processes are effective in that model universe during its simulations. These evolutionary processes are defined along the lines of neo-Darwinistic view of the evolution of biological life on earth. The framework defines in algebraic and statistical terms major components of the evolution - the presence of reproduction in the entities, the variation in the characteristics of the entities because of mutations, the heritability of the characteristics across generations in order to maintain the variation, and the natural selection which results owing to the differential rates of reproduction among the entities in a population.

The framework is illustrated on four different kinds of ALife models including Cellular Automata based Langton Loops and Evoloops, Lambda calculus based Algorithmic Chemistry, and two new experimental artificial chemistries - the Reduced Instruction Set Artificial Chemistry and the Artificial Graph Chemistry. Generic design principles for the ALife research are drawn based upon the framework design and case study analysis.

List of Figures

Part I

The Background

Chapter 1

Introduction

1.1 The Phenomenon of Life and the Problem of Definition

The phenomenon of "life" on earth is one of the most intriguing one, which have evaded clear definition for long time. This lack of precise definition to characterize living systems can partly be attributed to the vast variety and complexity of forms in which life is found on multiple levels ranging from microbiological scale to higher level animals and plants. These living systems in myriad of morphological forms are known to be exhibiting a vast array of properties and characteristics. Nonetheless due to persistent scientific endeavors over the course of past three centuries our understanding of life has increased tremendously and at present there are very specialized branches in biology dealing with specific forms and characters of life - on micro organismic level (microbiology) to complex ecological level (ecology), from life under sea (marine biology) to search of life on extra terrestrial levels [Biology05].

Even after such detailed studies, a clear definition of life is yet to be formulated which can encompass all distinguishable properties which life-forms observed to possess. One of the acceptably comprehensive definition of life is proposed by Ernst Mayr, the leading evolutionary biologist of the 20th century, in [Mayr83] as a cluster of properties which can distinguish "the process of living" from "inanimate matter" such that not all of these properties are found to be present in any known non living forms. The following is a slightly changed version of original list (emphasis on certain terms is mine):

1. All levels of living systems have an enormously complex and *adaptive*

organization.

2. Living organisms are composed of a chemically unique set of macromolecules.
3. The predominant phenomena in living systems are *qualitative* not quantitative.
4. All levels of living systems consists of *highly variable population* of unique individuals.
5. All organisms possess historically evolved genetic structures which enable them to engage in 'teleonomic' (purposeful) processes and activities.
6. Classes of living organisms are defined by historical connections of common descent.
7. Organisms are the product of *natural selection.*
8. Biological processes are essentially unpredictable.

This list comprehensively compiles major aspects of life found on both individual level as well as population level. As discussed in [SS99] there are indeed two major schools of biology which focus on two different aspects while defining life. The first school is system-theoretic, which places emphasis upon metabolism and self organizing properties of (individual) living systems. For example, Kauffman defines life in terms of "autonomous agents with self organizational and open ended adaptation capabilities " [Kauffman89] and Maturana and Valera have defined life as an "autopiotic system, which continually reproduces itself through a network of interactions" [MV80]. Another school of thought is based upon neo-Darwinistic conception, which associates life with its population level evolutionary adaptation described in terms of multiplication, heredity, variability and natural selection [SS97], [SS99], [Kuppers90]. In a related theme on Artificial life (discussed in the next chapter), Langton [Boden96] identified list of properties defining life, which include the dynamic processes organized in specific ways, self organization, self reproduction, emergence, evolution, and epigenetic gap between genotype and phenotype.

In this thesis we will focus upon the later evolutionary approach for defining life and system theoretic (individual level) metabolic and self organizational properties will

not be considered in the current scope of the work. In the following discussions, the biological life on earth will be referred to as 'real life' in order to distinguish it from the artificial life (discussed in the nest chapter). The rest of the chapter briefs upon the known status of evolution for the real life.

1.2 Evolution in Real Life

Evolution basically refers to the changes within a population of living things across generations and (observed) over long periods of time. In its basic structure evolutionary theory assumes that all changes basically originate in populations of species and then carried on to the next generations, if those changes are inheritable and selected by the natural selection. Thus cross mingling between species does not lead to any new forms of life due to the homologous nature of sexual reproduction and non inheritability of phenotypic changes, though may sometimes result into some kind of cooperative systems.

The molecule called DNA (deoxyribose nucleic acid) provides the physical mechanism of heredity in all living beings, finer structure of which is inferred by x-ray diffraction. All DNA molecules differ only in one specific portion of their structure known as nucleotide base and there are only four possible such bases commonly represented as **A, C, G**, and **T**. Another interesting property of the DNA molecules is their complementary chemical affinity whereby **A** makes stable and strong chemical bonding with **T** and **C** with **G**, not otherwise. This complementarity is very fundamental for the DNA to work as basic coding structure controlling the reproduction process in the organisms. Long double sequences of DNA molecules (called duplex structures) make up all genetic material for any organism.

The living systems interact with the environment not directly through their DNA based genetic material but through elaborated phenotype structures. The exact translation from genotype to phenotype is not yet well understood but is controlled through the proteins which are encoded in the genes. DNA to protein coding is defined in terms of triplets of DNA molecules consisting of three nucleic acids coding for one of the twenty different amino acids comprising proteins.

Living forms are divided into two distinct categories based upon their cellular structure: the Prokaryotic organisms and the Eukaryotic organisms. Eukaryotic cells differ

from prokaryotic cells in the sense that eukaryotic cells contain many membrane bound organelles (small structures inside the cell which carry out specialized functions) while prokaryotic cells on the other hand possess no such membrane bound structures. For example, the nucleus is surrounded by a nuclear membrane and contains most of the hereditary material (DNA) of the eukaryotic cells. Most of the bacterial, viral, and primitive unicellular organisms are prokaryotic, while almost all multicellular organisms are eukaryotic. The sequences of the nucleotide bases are grouped into genes, chromosomes[1] and then genomes. Special interest organisms among eukaryotic are those, which reproduce sexually. In sexually reproducing organisms, the genes are inherited in characteristic Mendelian ratios which express the proportion of different genotypes in the offspring of parents with particular combination of genotypes. These Mendelian ratios are very important because they maintain variation in the population and thus allow natural selection to work. In absence of such mechanism for maintaining the variation, evolution by natural selection will not be possible as we will discuss further in later chapters.

The major evidence for evolution comes from the observation of certain kind of similarities (called homologies) between living things on various hierarchal levels (molecular, organismic, groups, species, higher taxa etc), which will not be expected if species originated independently. Again existence of adaptation has no non evolutionary explanation though exact evolutionary explanation for various adaptations might differ. Observation of evolution on the small scale in controlled studies combined with the extrapolative principle of 'uniformitarianism' , which states that natural laws must have operated (and thus evolution worked) the same way even in the past in absence of any observations, suggest that all life might have evolved from same common ancestor. Universal homologies - such as universal genetic code found in all living things also suggest that all species descended from a single common ancestor.

In the following two sections we will give a brief description of the currently held theoretical conception of evolution as the synthetic theory of evolution, which will also lay conceptual foundation for the work presented in this thesis in subsequent parts.

[1] All sexually reproducing organisms possess two copies of chromosomes (one inherited from the father and one from the mother) in each cell, a combination of two genes present at a particular locus in the genome is called a *genotype*. Prokaryotic organisms do not contain chromosomes rather have single copy of genetic material.

The discussion is based upon the standard texts on evolutionary theory - [Ridley96, Stephen00, MSmith98, Futuyma98, SS97, SS99, Volkenstein94, MB97, Ridley97].

1.3 The Synthetic Theory of Evolution

The first theory of evolution was formulated at the beginning of the nineteenth century by J. B. Lamarck [Capra97, pp. 217] who also coined the term 'biology'. Lamarck observed that animals changed their characters [2] during their life times due to the environmental pressures and he believed that they could pass on these acquired changes in their characteristics (characters) to their offspring. Though his idea of this kind of hereditary transfer does not hold in case of real life, the very idea of evolution or changes being generated and propagated across generations was fundamental in shaping later theories in understanding life as compared to static theories propounded by Creationists, which hold that life forms never change and are in this state since when they were "created" (by some external entity.)

Coupled with his initial observations on professional breeding and domestication of animals, Charles Darwin, on the other hand made keen observations on the fauna of the Galapagos Islands and other ecological environments and speculated that geographical isolation could have an profound effect on the formation of species. He synthesized his ideas on the role of environmental effects in evolution of life in his well acclaimed theory of natural selection in [Darwin59]. Darwin based his theory on two key ideas - chance variation and natural selection. The core insight of Darwinian thought is that all living organisms are related by common ancestry and all forms of life emerge due to continuous process of chance variations and natural selection.

The current structure of Darwinian theory (called 'neo-Darwinism') differs from its original structure on the actual 'dynamics' of evolution though the central role played by variation and natural selection remains unchallenged. The major shift in Darwinism came through its synthesis with Mendelian laws of genetical heredity [Ridley97]. The neo-Darwinism thus can be seen as a theory of population dynamics (natural selection working on populations) with Mendelian genetics. According to the neo-Darwinistic theory, major evolutionary variation results from random genetic mutations followed by

[2] observable properties of organisms are called characters. We will use 'characteristic' and 'character' synonymously.

natural selection.

In the mathematical theory of evolution [Evolution], the heritable structure, the genome, is abstracted as linear sequence of mutable genes, each present in one of the many possible allelic forms. The Evolution thus has been defined more specifically as a change in the frequency of alleles in a population from one generation to the next as a result of genetic drift, gene flow and natural selection discussed next.

1.3.1 Natural Selection

According to the Darwinistic theory of evolution [SS97], there are four essential preconditions for the occurrence of evolution by natural selection:

1. Reproduction of individuals in the population. (Section 6.2)
2. Variation in characters within population, which affect likelihood of survival of individuals based upon interaction with the environment. (Section 6.1, 6.2)
3. Heredity in reproduction, that is, like begets like or in other terms mechanism to maintain variation.(Section 6.3)
4. Variation in (fitness of) individuals as per the individual characters which affect their rates of reproduction. (Section 6.4)

These factors arguably result in natural selection which changes the characteristics of the populations over time. Natural selection thus resulting because of above factors can be either *directional* or *stabilizing* or *disruptive* [Volkenstein94] (see Figure 1.1). Directive selection is a driving force and leads to a change in the genetic constitution or phenotypic distribution of characters in a population. Disruptive (or diversifying) selection favors preservation of extreme forms and the elimination of average forms, thereby contributing to increase in diversity of population as well as promoting speciation. The stabilizing selection on the other hand, performs a protective function by discarding extreme forms. Stabilizing selection is more widespread in case of real life.

1.3.2 Artificial Selection in-Vitro

There have been successful experiments in vitro to demonstrate that selection pressures (artificial selection) indeed work even in short periods of time under controlled laboratory

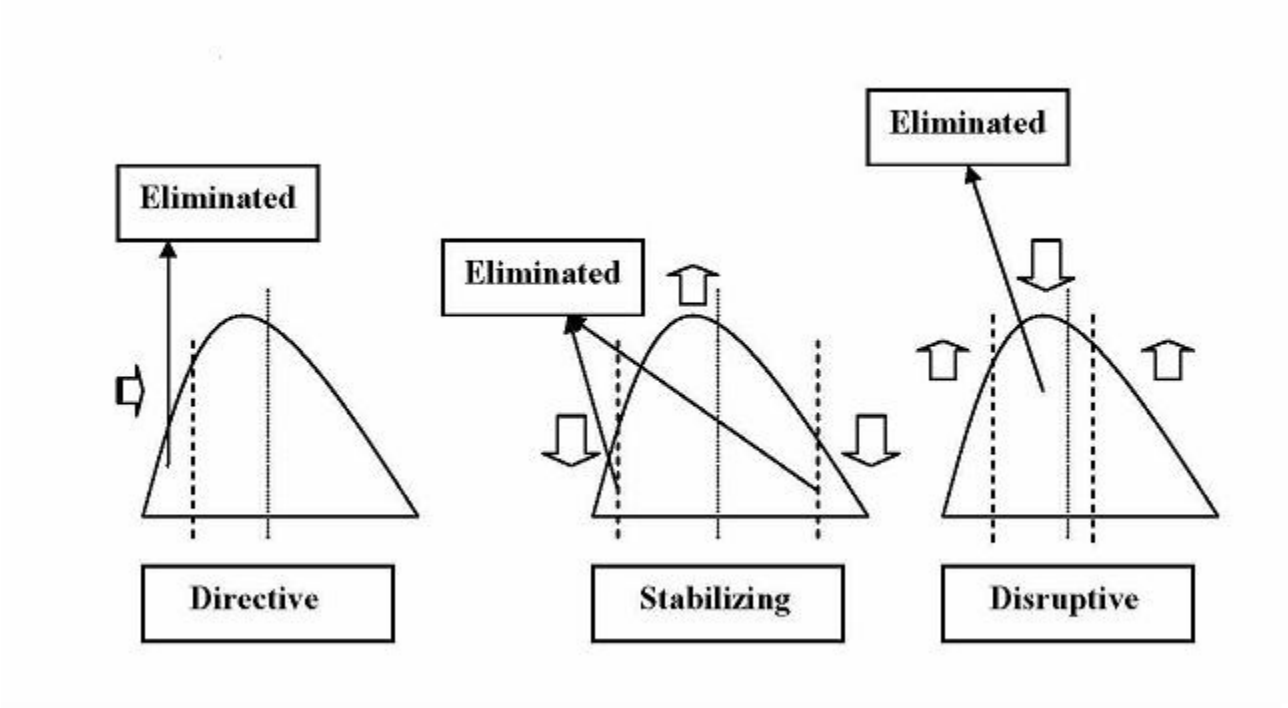

Figure 1.1: Three kinds of selection pressures: directional, stabilizing, and disruptive.

environments. These are [MSmith98] - first by Orgel using enzyme $Q\beta$ replicase and RNA and second on *Drosophila* population . These experiments shed significant light on the working of (natural) selection and form of genetical variation.

In case of Orgel's experiments, shorter and fast replicating RNA strands were naturally selected over the course of just 100 or so generations when starting point was only a single randomly chosen RNA strand added with the adequate supply of RNA monomers and using $Q\beta$ replicase as reproducing machinery. The repeatable experimental results could be used to draw some important conclusions like final population of RNA molecules depended upon the initial conditions (e.g., ionic concentration) and mutation errors during copying of RNA strands by $Q\beta$ replicase were the sole factors for introducing variability (RNA strand sizes, sequencing) and that in some experiments, RNA strands even developed resistance for antibiotic elements introduced in the test tubes (adaptation). Finally the RNA which evolved in these experiments was extremely unlikely to have evolved by pure random chance since probable cases were ultra exponentially large.

The laboratory experiments on *Dorophilia* population exhibit the variability and response for artificial selection pressures. Apart from two or three exceptional cases, in

almost all of phenotypic characteristics, genetical variations found to be present in the naturally occurring populations.

1.3.3 Stability and Variability of Hereditary Transmission

The way genetic material gets transmitted from the parent to the child is the single most important factor in determining genetic stability and the genetic variability in the population. For evolution to occur, genetic transmission must be simultaneously stable and variable. For example, the transmission of genetic material from parent to child must, with high likelihood, pass parental characteristics to the child. This is because otherwise without such stability in inheritance, natural selection could select a parent with good characters but children of that parent would not be likely to inherit those good characters. On the other hand genetic transmission should also allow, at the same time, new traits to appear and get inherited by children so that the variation in traits is present in the population for the natural selection to work upon.

Sources of Hereditary Variability

Genetic variability in a population is the result of the four principle forces:

Mutation

Variation resulting from random changes in DNA, such as mutation is the main source of variability in evolution. There are many types of mutations, including [MSmith98]: *point mutations*, which change the genetic sequence by one symbol. These mutations often produce neutral variations but can also lead to useful new sequences. *Frameshift mutations* are the result of addition or deletion on genetic sequence and in general have harmful effects. Lastly there are large sequence *rearrangement mutations*, which are almost always fatal. Mutations are indeed the prime source of variability in asexually reproducing organisms like bacteria.

Horizontal and Vertical Genetic Transmission in Bacteria

The exchange of genetic material among bacteria through mechanisms such as phages, plasmid conjugation[3], and transposons[4] is also a source of bacterial genetic variability. In case of horizontal conjugation a bacterium actually injects a copy of part of its genetic material into another bacterium, where homologous recombination occurs. In case of transposon, it is able to insert an entire gene into the genetic material of recipient bacterium. Transposons are indeed supposed to be carriers of genes for antibiotic resistance among bacteria. Vertical genetic transmission refers to usual parent to child transmission.

Homologous Recombination

Exchange of genetic material in sexually reproducing species happens through recombination. Genomes from both the parents are recombined to produce an entirely new genome for the child. This recombination process is homologous in its nature and thus is major source of stability as well as variability in natural evolution. Homologous genetic exchange occurs during "crossover" in sexual reproduction as well as bacterial plasmic conjugation. In homologous exchange, genetic material is exchanged such that it largely preserves the function of genes as well the genomic length. The homologous exchange works only for almost identical genomic sequences which can properly align with each other and that is the reason inter species reproduction is not seen in nature.

Symbiogenesis

This is yet another recently accepted source of variability in real life forms proposed by Lynn Margulis [Margulis81]. While the primary source of the origin of varieties in above cases is the genetic material per se, in symbiosis that is not the case. In symbiosis, different elements with even different genetic material seem to come to some kind of permanent seamless cooperative coexistence and thus giving rise to novel variety of life forms. The best evidence for symbiogenesis ('process of origin of new life forms through symbiotic relationships') comes from the presence of mitochondria in nucleated cells

[3]Conjugation refers to the physical transfer of genetic material between prokaryotes.

[4]Transposons is a sequence of genetic material that can move around in the genome, also called "jumping gene".

and chloroplast in plant cells. These mitochondria which are also the source of cellular energy supply contain their own genetic material and even reproduce independently at different times than the rest of cell. Indeed mitochondrial genetic material is actually used to measure mutation rates because it is directly passed from the mother to the child.

Whereas the conventional theory of evolution sees the unfolding of life as a process in which species only diverge from one another, in symbiogenesis, new life forms originate as a result of formation of composite entities through the symbiosis of formerly independent organisms. Lynn Margulis has also argued that major source of speciation is not the mutational variation and subsequent fixation through natural selection but rather the process of acquisition and integration of new genomes by symbiotic merger [Margulis02].

Hereditary Stability

As pointed out before stability in hereditary transmission is as necessary as variability in order for selected variance to propagate across the generations. The principle mechanisms of stability in real life include:

Redundancy There is a high level of redundancy in the genetic material in almost all life forms. Sometimes this redundancy is so high that only 5% of all genetic material is used. The positive effect of this redundancy is that various harmful mutations are absorbed in the redundant portions of genetic material.

Repair Complex repair mechanisms operate on the genetic sequences for repairing their damage and correcting the copying errors so that almost identical genetic material is transmitted from the parents to the children.

Homologous Recombination Homologous recombination during sexual reproduction as explained before tends to prevent the fixing of negative mutations in the population (thereby reducing variability) [SS99].

1.3.4 The Structure of Synthetic Theory

The formal structure of neo-Darwinistic theory of evolution also known as population genetics, is primarily concerned with changes in genotype and gene frequencies across

generations under the effect of random drift and natural selection where heredity preserves the basic genetic structures from the parent generation to the child generation.

The general model of population genetics starts with some arbitrary chosen point in a generation n for a particular species (or "gene-pool") and estimates the structure for the next generation $n + 1$. Thus it considers the relative frequencies of genotypes in population in generation n and specifies how these genotypes recombine as per some mating criteria. The mating criteria for a particular species determines which individuals successfully mate based upon their fitness for sexual selection. At this point it applies Mendelian ratios to estimate frequencies of genotypes among the offspring population. Next these genotypes are subjected to selection weights which determine which of these offsprings correspond to surviving phenotype individuals. Here natural selection operates in two ways - by creating differences in survival of genotypes and secondly by differences in fertility (indeed no survival can also be equated with zero fertility). This general model of population genetics is further simplified sometimes using well known Hardy Weinberg theorem which predicts the allele frequencies in sexually reproducing populations across generations in absence of selection. In other words, under Mendelian laws of heredity, in absence of selection, population converses towards Hardy - Weinberg equilibrium [Ridley96].

Next focus of the synthetic theory is to define fitness of a genotype, which consists of the relative probability of survival from birth to the childhood. The fitness can also be inferred from the changes in the gene frequencies between generations. Sometimes fitness of a genotype also depends upon its frequency and the migration between the populations, which acts to unify frequencies such that the balance between the selection and the migration can maintain genetic differences between sub-populations. Harmful mutation which are selected against but persistently arise can be settled with low frequencies in the populations, which is also known as selection-mutation balance.

Random Drift

The formal model of population genetics assumes very large populations so that the changes due to the random sampling for reproduction can be eliminated. But in case of small finite populations effects of such random drift cannot be ignored. Random drift

arises because only a randomly selected fraction of parental population is allowed to reproduce and therefore successive generations form a random sample from the parental gene pool. Under genetic drift, frequencies of characters even with the same fitness change at random. Nonetheless, in the absence of mutations, over longer periods, random drift causes certain characters to get fixed in the population. Random drifts can also supplement natural selection by enabling populations to explore valley bottoms of the adaptive topographies.

Neutral Theory of Molecular Evolution

Another interesting formal development in the history of synthetic theory is the Kimura's theory of neutral selection and neutral drift for molecular evolution (observed as protein evolution), which suggests that the mutations which have been substituted in the evolution were selectively neutral with respect to the genes they replaced [Kimura83]. The observations of rapid but constant rate of molecular evolution in the functionally less constrained parts of the genome and the high polymorphism are cited as supportive evidence for the neutral theory. The main property of the neutral selection is that it can drive high rates of evolution and can maintain high levels of variation without imposing any genetic load. Kimura pointed out the contrast between the observed uneven rates of macro-morphological evolution and the constant rates of molecular evolution, and thus argued that the natural selection could not have maintained constancy in molecular level evolution without neutral drift.

Quantitative Genetics

Whereas population genetics deals with the changes in the genotype frequencies for small number of genes (loci), quantitative genetics on the other hand deals with those characters which are controlled by the large number of genes and for which exact genotype(s) producing related phenotypes are not known. The estimates are made for the changes in the frequencies of phenotypes and the corresponding genotypes between generations. Influences on the phenotype variance of these characters are divided into environmental and genetic factors. The genetic factors are further classified based upon their heritability.

In quantitative genetics all quantities are expressed with respect to the population

means. The value of a character is thus expressed as deviation from the population mean and heredity is defined as the proportion of phenotype variance which adds the corresponding variance in the child population. Intuitively, if a parent differs from its population mean by certain amount then the degree of heritability is reflected in the amount by which the child also deviated from the mean of child population, thus in the absence of heritability the child should not deviate much from the mean. Similarly response to the artificial selection is measured as a difference between the offspring population mean and the parental population mean, which is actually determined by value of heritability of the character and is also controlled by the genotype-phenotype relationship (if non-linear).

The Units of Selection

As discussed before, adaptation is one of the basic explanatory component for any theory of life on earth. Evolutionary theory explains adaptations using the process of natural selection. Thus if a character is an adaptation, then natural selection works against its mutant alleles (i.e., reduces their reproductive success). Adaptations are found on almost every level of organizational hierarchy from nucleotide to gene, through organelles, cell, organ, organism, group, population, species, and higher taxa. The question remains for the benefit of which level in this hierarchy does natural selection work for adaptations to emerge and evolve? In some cases it may not matter that much because what might benefic an individual organism may also benefit the population and its species at higher level, and all of the parts of the organism, on the lower level. However conflict arises because in some cases what benefits on one level might not benefit on other levels or indeed may be at the cost of other levels. This question in more generic terms is restated as to why adaptations evolve at first place? If the natural selection works only to produce adaptations for one specific level and that works incidently for other levels then the question is what is that level? Restated, “What is the unit of selection?”

Segregation distortion is an example of an adaption of a gene against its allelic alternatives. Similarly, in some organisms having separate germ and somatic cell lines, selection weakly works on the cell level. These kind of adaptations are rare but adaptations on organism level are abundant. Forces of groups selection which might work on

specific circumstances are also weak in nature.

From an evolutionary view point the accepted answer to the units of selection is that the adaptations are produced for those levels in the hierarchy of life which possess heritability or the unit of selection is that entity whose frequency is adjusted directly by natural selection over long periods. In such a case, the gene and the organisms seem to be the most potential candidates for the unit of selection out of which gene frequencies are definitely adjusted by the selection over an evolutionary time scale. In terms of Richard Dawkings [Dawkins99], this unit of selection is called *replicator*, which has permanence over long evolutionary time scales for the selection to adjust its frequency. The replicator can be physically mapped to small unit in the genome known as cistrons [Ridley97]. *Kin selection* where genetically closely related entities in a population behave in altruistic manner (e.g., laborer honey bees) by sacrificing their reproductive success in favor of the reproductive success of the other kin can also be explained as an adaptation which increases only (altruistic) gene frequencies.

1.3.5 Macro - evolution and Coevolution

Life on earth has long history of more than 3.5 billion years. During such a long span life evolved in many directions and forms such as origin of new species, new organs, new higher taxa. Macro-evolution is the study of such large scale historical events and changes shaped over the course of long periods. These large scale historical changes spread across a wide range of species' populations can be distinguished from the micro-evolution, which refers to changes in gene frequencies over generations within a population. Speciation in some sense the defines the dividing line between the micro- and the macro- level evolutions.

Macro-evolutionary studies are mainly conducted using fossil records, which also provide the data for estimating the rates of evolution. The major question in macro-evolutionary studies concerns whether these observed large scale macro-evolutionary events are extrapolation of micro-evolutionary events over longer time scale or are there some other higher level mechanisms in operation. Changes in abiotic environment, such as temperature etc are definitely the factors, which might cause sudden or large scale changes in the populations, which will not be just accumulated results of micro-level

events.

The observations from the measurements of evolutionary rates using fossil records indicate that the faster evolution seemingly takes place at shorter time intervals, which led to an alternative hypothesis called 'punctuated equilibria' maintaining that the rates of the evolutionary changes happen faster during speciation and then come to a halt [GN77]. Opposite to this, is the idea of 'phyletic gradualism', which maintains that evolution has constant tempo and small mutational changes accumulate over time at a constant rate. Unfortunately due to the insufficient fossil data, it is not possible to determine which one of these more correctly explains the pattern of evolutionary rate changes.

Most of the macro-evolutionary transitions are supposed to be the results of the macro-evolutionary changes in developmental pathways in the organisms.

Coevolution

Coevolution happens when two or more species reciprocally influence one another's evolution. Coevolution is usually used to explain the coadaptations found in nature on several levels. The example of ants feeding on the honey produced by the caterpillar of butterfly and in return providing protection from the parasite wasps and flies is a classic example of such coadaptations between the species [Ridley97]. The morphological structure and the behavior patterns of both the ants and the caterpillar suggests that they have evolved in relation to each other. Thus in contrast to ordinary evolution where species evolve in relation to their abiotic or physical environment, in coevolution they define each other's environment which itself evolves. Thus species exert selection pressures on each other and evolve themselves in response to that.

Parasites and hosts provide another example of a different kind of coevolution, where species evolve characters not to cooperate with each other, as suggested by above example of the ants and the caterpillars, but in order to compete and defend from each other. Thus parasites evolve more improved properties to penetrate the hosts while hosts devise more effective protective characters. If the range of genetic variants in the parasites and the hosts is limited then coevolution might be cyclic. Observations suggest that the phylogenetic branching (i.e., speciation pattern) of parasites and hosts have mirror

structures and are simultaneous. Another consequence of the coevolution is the arm's race between the predator and the prey species, which might sometimes also result into escalation.

Any set of competing species that coevolve against each other under the constant supply of resources, can generate four kinds of patterns of extinction: *the contractionary* (species lags behind its competitors and goes extinct), *the expansionary* (species evolves ahead of its competitors and expands), *the stationary* (competing species evolve to a optimal state and stay there), and *the Red Queen equilibrium*, where species continually evolve but balance each other exactly. The Red Queen hypothesis is proposed by Van Vellen to explain the generic pattern found in fossil records that rate of extinction of species is independent of their age. Red queen hypothesis suggests that the species evolve continuously to maintain a level of adaptation against other competing species for constant resources (giving rise to the 'zero sum game'). In this red queen mode of coevolution, natural selection continually works on each of the species to keep up with improvements made by other competing species. On average, each of the species have balanced level of adaptation against their competing species.

Having described in brief, the major aspects of the neo-Darwinistic conception of the evolution in real life, we will end this chapter with a note on the ongoing controversy related to the very scientific status of this evolutionary idea.

1.4 Scientific Status of Evolutionary Theory

The scientific status the evolutionary conception to explain life was questioned by the well known scientific philosopher, late Karl Popper on the grounds of his refutability principle [Popper63, Gould81]. Refutability is the most accepted criterion to justify the scientific status of theories and for the demarcation of the scientific theories from the pseudo scientific theories or meta physical theories. However refutability is in general defined for the physical theories which are mostly "current" - in the sense that they model currently occurring phenomena or describe concepts founded on the currently observable facts and therefore can be refuted by demonstrating counter experiments or observations. Unfortunately evolutionary theory cannot be said to be completely refutable in this sense because of its very subject of study—evolution of life—mostly

being a historical phenomena. But this does not invalidate its scientific status because as Mark Ridley noted in [Ridley97], evolutionary theory is based upon the principle of "*uniformitarianism*", that is, the principle that what is observed in small scale experiments for evolution, indeed carries backward to whole history of evolution of life on earth in more or less the same way. This is important because evolution on earth has been occurring for past several billions of years and we have started studying it only now and due to the extremely slow pace of evolution, experiments of similar nature cannot be performed on the same scale as is easily possible with other kind of physical phenomena. And that is where, exactly the role of Artificial Life studies come into picture - as discussed in the next chapter.

Chapter 2

Artificial Life

Having discussed the nature of biological life on earth from evolutionary perspective, we will present a brief survey of the field of Artificial Life (ALife) in this chapter. Discussion in the chapter is based upon various aspects of ALife studies including the overall context in which ALife studies aim to make contributions, the broad research goals, and a generic design for ALife models. Some of the underlying philosophical issues related to the problem of defining life, the problem of epistemological cut and the problem of emergence are discussed after that. We compare ALife research with real life and discuss the synthesis of these two fields. This is followed by a brief discussion on a subfield of artificial chemistries.

2.1 Introduction to Artificial Life

As discussed in detail in the previous chapter, after its conception by Charles Darwin [Darwin59] in the late 19-th century, evolutionary theory has been regarded as the most plausible explanation for the emergence of the enormous variety of complex living forms [Ridley96], [Futuyma98], [Ridley97], [Stephen00]. Though we understand and can explain several aspects of life around us, we still do not have a comprehensive and complete understanding of the principles underlying the emergence of life and the conditions that led to the diversity and complexity of life on earth [Futuyma98].

It is a long held topic of scientific debate whether there are any biological principles of life and other complex biological phenomena, which are not directly reducible to physical and chemical laws. Living beings, however small and consisting of the same

molecular components as non living, nonetheless exhibit qualitatively different characteristics (ref. Section 1.1). This may be in part due to the complex organizational structure which distinguishes them or it could be because of their quantitatively large number of subcomponents, which give rise to difficulty in analyzing properties using currently available tools.

The direct ways to understand this complex biological phenomena are usually difficult and error prone because living systems are by default complex and hard to manipulate. Even cellular level experiments are difficult to carry out and their simulations are usually cumbersome.

Artificial life is a tool to study the principles explaining this complex phenomena of life without directly getting involved with the real biological systems. The fundamental assumption here is that principles of life are independent of the medium and the carbon based biological life on earth is just one example of the possible forms of life. This implies that even artificial environments like digital media can also exhibit life-like behavior under certain conditions. The studies of ALife, thus complement the main stream biological studies by synthesizing life-like systems using alternative synthetic structure and simulating such artificial models using digital media. There are several such examples where these artificial life forms have been shown to be exhibiting properties remarkably close to higher forms of life, e.g., "parasitism", "hyper-parasitism" and other ecological forms in artificial systems like Tierra [Tierra-Webpage] and Avida [Avida-Webpage].

Artificial life research is, thus, an attempt to study possible generic principles of life by synthesizing life-forms as *they could be* rather than what *they are* [Langton89, Langton95, Adami98]. We need these synthetic alternative structures in ALife studies because studies with real life forms to answer most of fundamental questions on the nature of life are enormously complex and sometimes practically impossible. We can consider for example the question as to whether there can be alternative forms of life possible? Supposing that we agree at a specific definition of life, which can be observed if exists in some system, can we answer at least whether life might exist with some other chemical basis rather than carbon based structures using nucleotide and amino acid bases? Can we know how a possible form of life consisting of alternative chemical structures might have evolved over? All these questions require fundamentally complex

and enormously difficult tasks because of the limitations we have in current experimental set up for working with chemistry as well as the enormously complex simulations required while using computing tools and techniques. Even complete simulations to realize folding of simple protein molecules, which determine their chemical properties, is itself an computationally infeasible task with currently known tools. Therefore the only plausible answer for all such questions seems coming from ALife studies where alternative molecular structures are defined abstractly along with their novel reaction dynamics and are allowed to evolve sufficiently long enough in a simulated environment. Once we observe the emergence of reasonably satisfactory life-forms in such synthetic ALife model, it will be another question to establish proper correspondence with that in real-life to draw useful inferences.

We must note, however, that these artificial synthetic structures are not the simulations of real carbon based life but are indeed the realizations of alternative forms of life under different settings. These ALife models use novel structural building blocks as well as different environmental constraints with novel reaction dynamics. The examples include Tierra [Tierra-Webpage] and Avida [Avida-Webpage], where structures are simple executable programs instead of carbon based molecules and environmental resources analogous to "energy" and "matter" are replaced by "processor execution cycles" and the "hardware memory".

The basic guiding principle of ALife research behind all these novel class of synthetic structures can be expressed as - "life is a property of form and organization rather than the matter which is used to build it" [Langton95]. Validity of this principle follows from the observation that had it not been the case then almost all kinds of material structures could be well considered as living, which is surely not the case.

This criterion to identify life as a property of form and organization immediately poses further questions as to which kind of organizational structures possess life? Which properties should we be looking at in those organizational structures? and so on. We answer these questions in one of the possible ways by identifying properties of living structures with those which undergo evolution as described in previous chapter. Thus for our purpose any system of structures dynamically changing in any ALife universe will be termed alive if that can be observed to be evolving in the sense of neo - Darwinistic

evolution consisting of reproduction, mutation, heredity, and variation and reproductive fitness based natural selection as discussed in Section 1.3.1.

Cellular Automata based models are one of the earliest attempts of synthesis to understand the underlying logic of self replication [Sipper98]. Von Neumann defined universal constructor which in turn could construct itself [Neumann66]. His initial constructor was both capable of constructing the copies for other structures (configurations) and had capability of universal computation as well. This resulted into the machine becoming enormously large in its possible configurations and self replication could be demonstrated only by analytic means [Burks70]. Langton and others later simplified that by only demanding construction of very specific (self) configurations without any computation or universal construction capabilities [Langton84]. We will consider these self replicating structures as an example case study for our framework presented in Chapter 8. The major problem with these models seems to be their limitation with self replication and they seem to be quite fragile against any kind of (mutational) changes.

Later attempts in the field of ALife studies have considered several new kinds of synthetic structures including programs, lambda terms, strings, graphs, automata's etc ([DZB01]) and have demonstrated that one or the other observable properties of life are shared by all of these new class of models, though the parallel diversity and robust evolving structures which we find in real-life are yet to be discovered.

2.2 Goals and Mechanism of Artificial Life Studies

As discussed in the previous section, major aim of ALife studies is to develop a coherent theory of life in all its manifestations, rather than the historically contingent documentation as is the existing situation to some degree with real-life theories [BMRAGIKT00]. The major steps of ALife studies to achieve such goal can be to construct novel synthetic structures under unfamiliar settings which can evolve with novel emergent properties in order to complement theoretical biology by uncovering detailed dynamics of evolution where real life experiments are not possible, and to develop generalized formal models for life, which can be used to define life in more generic setting and to determine criterions so that life in any arbitrary model can be observed.

Goals: [BMRAGIKT00] has presented a succinct list of major challenges for the

short term and the long term research focus in the ALife studies. These challenges can be classified into three main directions: First, how does life arise from the non-living? This is further expected to be answered by synthesizing a molecular proto-organism in vitro and by demonstrating the transitions to life in an artificial chemistry in silico. These studies will therefore help to determine whether fundamentally novel living organizations can exist. Second direction is to develop theories to explain potentials and limits of living systems by asking how rules and symbols are generated from physical dynamics in living systems and to determine what is inevitable in the open-ended evolution of life. Another question in this direction pertains to the development of a theory of information processing, information flow, and information generation for evolving systems. Third direction relates the field of ALife with other fields of the mind, machines, consciousness, and culture. This is to be progressed by demonstrating the emergence of intelligence and 'mind' in an artificial living system.

Mechanism: ALife uses informational concepts and computer modeling to study life in general. The basic design mechanism of all ALife models proceeds in more or the less same way: The designer of the model has to define basic units of the ALife universe such that new functional properties or the new kinds of structures should appear over time. In order for this synthetic universe to progress under simulated environment, rules for progression or updation dynamics are also suitably defined by the designer of the model. Once definitional phase is over, instantiation of the model with specific initial settings is carried out by simulating that model under the repeated application of progression rules. If the model has been set up "appropriately" then over time some novel structures not present in the universe from the beginning or novel class of functional properties not exhibited by initial entities emerge. The simulation experiments might be repeated with several different initial settings as well as new progression rules. In the concluding chapter (13) of the thesis we will discuss some generic design suggestions based upon our work for the ALife researchers.

2.3 The Philosophy of Artificial Life

Functionalism plays very fundamental role in underlying philosophy of ALife. Functionalism has its roots in mind-body duality, though that does not concern us in ALife

studies directly. Functionalism in ALife appears because of the very assumption that life can be dissociated with its material basis and can be reduced to certain kinds of functional behaviors, including reproduction, organization, evolution [Langton95]. Conventional biologists are though opposed to such claims who see life as an integral property of carbon based organic structures. Another reason of skepticism to ALife arises because of the fact that often there is no easy way to translate ALife results into the domain of real-life because the ALife models are purely computational and they can generate systems exhibiting life-like properties that would in real world require complex underlying levels of organizations, which may or may not be feasible due to physic-chemical constraints imposed by laws of physical universe.

The Unity of Life: The underlying philosophy of ALife studies is that life can be present in similar forms independent of the underlying syntactical structures. Thus the focus of ALife studies is to suitably define the semantical aspects on these syntactical forms, which can give rise to life-like phenomena. This is also known as 'strong ALife hypothesis'. The weaker version of it asserts that though life is an exclusive property of real universe, it can be studied by working with similar synthetic structures, which have some correspondence with the real universe.

The Problem of Epistemological Cut: The major philosophical problem with ALife studies is of the epistemological cut. Epistemology is the branch of philosophy which deals with the methodology of asking questions. The problem of epistemological cut is to determine where should one define the boundary for a particular field. This is a major problem in ALife studies as discussed in detail by Pattee in [Pattee95] and is the result of the (unfortunate) fact that there is only one instance known for life—the real-life—which itself is not properly understood and for which it is not known what are the universal properties and what are the contingent ones.

Emergence: Another major philosophical and technically important concept in ALife studies is that of "*emergence*". Intuitively emergence implies to the appearance of the non trivial properties (of life), which were not present in the model by design, during the simulation of the model. The strong form of emergence holds that emergence is indeed a property of consciousness and equates it with some kind 'element of surprise' for the observer of the model [BE97]. Cariani [Cariani91] proposed a methodology to

detect this strong form of emergence using the metaphor of 'emergence relative to a model' (role of conscious an observer is replaced by the 'modeler'). His methodology was discussed and applied in the Tierra model in [Ray91] on all the levels: the emergence of new *syntactic* structures, the emergence of new *semantic* functions in those syntactic structures with respect to each other giving rise to new 'meaning' in the system, and the emergence of new *pragmatic* fitness landscapes in the system. Another notion of emergence [Bedau97] is referred to as 'weak emergence', which on the other hand maintains that living system which were not present in the chemistry by design from the beginning are generated in non trivial manner such that their presence cannot be predicted using analytical means and therefore simulation is fundamental to discover those non-trivial living structures. In our work presented in this thesis, we work with this weaker notion of emergence and necessitate that simulations must be carried out to demonstrate presence of any form of life in the ALife studies.

2.4 Artificial Life versus Real Life

The differences between ALife and real life are very fundamental. Whereas real life is the only known instance of life, that too with enormous variety and diversity, ALife models are potentially infinite and knowing that life might exist in some of these models is a non-trivial task. One technical point where this difference becomes very explicit is nature of genotype and phenotype relationship in case of real life and its counterpart in ALife models. In real life we do not yet know exactly how genotype and phenotype relationship is defined and we also do not know how did it evolve in its current form where a codon triplet consisting of three nucleotide bases codes for an amino acid [1]. All phenotypic meaning of underlying genotype is, therefore, reduced to the variance in the reproductive success or fitness. On the other hand, in case of ALife models there is always an explicit experimenter, who defines phenotypic meaning (expressed in terms of reaction semantics) of underlying genotypes (syntactical units) - for example the beta reduction reaction semantics is defined for λ terms of Algorithmic chemistry [Fontana92].

[1] A related interesting key assumption of molecular biology [Crick70, Werner05] associated with this coding scheme is known as the *central dogma of molecular biology*, which states that information does not transfer from proteins (which finally define all metabolic and other functional properties of an organism) to DNA/RNA. This translates into saying that phenotype changes are probably not inheritable because phenotype is a property of protein structure of an organism and thus changes therein will (possibly) not change the underlying genotype and its only the genotype which is inherited finally.

Another major difference comes because of the technical difficulties in studying certain aspects of real-life because of its historical nature. Because the real life what we study now is the result of more than 3.5 billion years of work by nature, it becomes very difficult to distinguish necessary (universal) aspects from the contingent ones. On the other hand we have enormous flexibility to work with ALife models where we can control almost every aspect of ALife study - the way basic building blocks are defined, the nature of progression rules, the initial settings for the simulation, and the observation process. While only limited kind of experiments can be carried out in real-life studies, with ALife models unlimited variety of experiments can be carried out.

The notions of complexity and adaptedness: While both of these concepts are relatively easy to measure in case of ALife models [Adami98], they are very difficult to even define for real life and the measurement is also problematic. There is no accepted technical notion of complexity in real life (apart from some measures of its diverse and varied forms), we have several well defined notions of complexity in terms of Shannon's information, or in terms and Kolmogorove or algorithmic complexity [Adami98], for studying ALife models. Adaptedness is again not so well defined concept in case of real life, though it is supposed to be measured by reproductive success. But in practice such reproductive success is very hard to estimate. On the other hand though precise notion of adaptedness might be difficult to define for all ALife models, for given specific models it can be concretely defined (sometimes) as fitness function or using fitness landscape.

The third major difference comes because of the nature of inferences we can draw from real life and ALife studies. While most of the real life experiments are supposed to shed some clear insight into the nature of real-life evolution, same is not true always for ALife studies, where it is not easy to infer the relevance of the observations made on the ALife model. The difficulty arises due to often conflicting observations reported in different ALife models.

The forth difference which has the greatest relevance for the purpose of this thesis is that of the role of observations in real-life as compared to that in the case of ALife. There is no doubt that experimental observations are fundamental for understanding real-life as with any other branch of science but because of the kind of uniqueness we encounter in real-life systems (universal genetic structure etc), the role of observations

becomes limited to only uncovering the underlying dynamics or the processes. On the other hand there is no such uniqueness in ALife studies and indeed the observations to be made upon one ALife model might not correspond to what are to be made on the other ALife models. Another point is that in case of ALife studies, in general there is no way to predict which kind of structures might demonstrate life-like behavior without observing closely the simulations. Thus identification of entities which might probably demonstrate evolution and other life-like behavior is a fundamental problem in ALife studies, while in case of real life we expect that all life forms must be following same underlying universal molecular structures.

Finally, as discussed in Section 1.4, evolutionary theory, which is considered to be most plausible theory of real life is not believed to be refutable in strict Popperian sense. On the other hand any theory of artificial life has to comply with the Popperian refutability principle because it must always be possible to cross verify the results of experiments as well as there must always be space in the theory of artificial life for possible refutations for claimed predictions. This places observation process used to study evolution in ALife studies at the center stage and we rigorously formalize that in this thesis.

2.5 Artificial Chemistries

Living phenomena has several aspects to study, one such is the origin of life or the *biogenesis*. Here the problem is to understand how first primitive forms of life such as metabolism and the self replicating structures could have come into existence starting from the non-living chemical compounds. *Artificial chemistries* (AC) are the primary tools in ALife studies aimed at understanding this origin of life and the other complex emergent phenomena. ACs follow the chemical metaphor—like real chemical reactions between molecules, which give rise to new molecules, ACs as well define abstract molecules and reactions and study what emerges during the course of reactions.

An AC has three main components, a set of objects or *molecules*, a set of *reaction rules* or collision rules, and a definition of *population dynamics* [DZB01].

Objects can be abstract symbols, numbers, lambda expressions, binary strings, character sequences, abstract data structures etc. Reaction rules might be string matching

rules, string concatenation, reduction rules, abstract finite state machines, Turing machines, matrix multiplication, simple arithmetic operation, cellular automata, boolean networks etc. Dynamics can be specified in terms of ordinary differential equations, difference equations, meta dynamics, explicit collision simulation, well stirred reactor, self organizing topology, etc.

In AC, we primarily consider the qualitative aspects of emerging structures, before considering the quantitative relations between its components. The quantitative aspect is usually analyzed using reactor flow equations [DZB01]. The stable structures generated by artificial chemistries, the stable sets of molecules, are usually referred to as *organizations*. Understanding which organization will appear is an example of qualitative solution of an AC.

Some of the aspects very commonly studied in AC are - given an AC, the organizations which are possible and which are not. Knowing which organizations are probable and which are improbable. To define an AC to generate a particular organization. Determining the stability of the organizations. Defining the complexity of an organization. To answer whether it is possible to generate an AC which moves from organization to organization in a never ending growth of complexity. Quantitative questions can also be asked, for example, given an AC, in a particular organization how many stable (attractive) states are present inside it?

A survey of ACs appears in [DZB01], which also has some broad classification of ACs based upon the kind of molecular abstractions (explicit or implicit), the types of reaction rules (constructive or non constructive), and the population dynamics. There is also some discussion on several interesting common phenomena which are observed in different kinds of AC systems such as reduction of diversity, formation of densely coupled stabled networks, and the syntactic and semantic closure in these networks.

2.6 Synthesis of Artificial Life with Real Life

It is interesting for any ALife study to discover in what sense there exist a parallel between the virtual universe of ALife model and the real life. In order to understand the biology of real life on molecular level, one needs to know organic chemistry as well as chemistry of macro molecules. Similarly, we need to understand the underlying structure

of ALife model before the parallels can be made. Sometimes ALife studies, unfortunately, only demonstrate 'what we already know' rather than 'what we can actually know' about life [Ray91]. Thus not only it is important to interpret ALife results in real life terms but it is also imperative to shed some additional light on the generic mechanisms behind that paralleling phenomena.

Thomas Ray writes more on this synthesis in [Ray91]:

> "In order to define an approach to the synthesis of life paralleling this historical state of organic life, we must examine each of the fundamental hierarchical levels, abstract the principal biological properties from their physical representation, and determine how they can be represented in our artificial media ... to generate a spontaneously increasing diversity and complexity."

Manyard Smith has also discussed in brief about the role of ALife studies in understanding real-life problems especially related to nature of the fitness landscapes [MSmith91].

In the next chapter we will discuss the conceptual foundations for the work presented in the thesis in the light of the discussion presented in these two chapters on the nature of real-life as well as ALife studies.

Chapter 3

The Background for the Framework

Having surveyed in brief the evolution in case real biological life on earth and the components of the synthetic theory of evolution, and the field of ALife in previous chapters, we will discuss in this chapter the contextual foundation for the work presented in later parts of the thesis. The aim of this chapter is to present a discussion on the problem of observation - its mechanism and goal of identifying life-like phenomena in ALife studies in the light of evolution.

3.1 Role of Observation in Artificial Life Studies

As discussed before in Section 2.4, observations play a fundamental role both in real life studies as well as in ALife research, though with differences. The differences come because of the nature of unique structures we encounter in real life forms on every organizational level (e.g., the universal genetic code), which might well be the result of historical contingencies. The role of observations, thus gets limited to uncover 'somehow' the underlying specific dynamics or processes for such real life forms. These observations though necessary, are therefore restricted by having only a single instance of the life available on earth. Therefore any attempt for generalization to extract generic principles of life based upon single example needs extrapolations and remains debatable because of the element of exercise to remove contingent features, which might be universal at times, from the necessary ones. On the other hand there are several different ALife models

and indeed the observations to be made upon one ALife model might be different from another ALife model.

Next, in case of ALife studies in general there is no way to predict which kind of structures might demonstrate life-like behavior without observing closely the simulations with an eye to identify (detect) life, which is in direct contrast to real life observations which start with the assumption that certain systems are indeed living and then proceed to study their features [1].

3.2 Need for a Formal Framework

The very identification of life thus is an existential problem for all ALife models and we need some workable framework to address this problem. In absence of a formal models, often intuitive arguments remain locally useful to one specific model and thus cannot be seen in generic perspective. Example of self reproduction is one such case where quite often entities are claimed to be self-reproducing on the basis of intuitive justifications based upon the observations, which are again remain imprecise in nature. This thesis is an attempt to answer this fundamental existential problem of observation to be made upon ALife models to uncover life-like structures. This we do by defining an abstract formal framework for identification of (living) entities and their progressive properties and relations with other entities, which might demonstrate evolution and other life-like behavior. In contrast, in the case of real life we usually expect that all life forms must be following same underlying universal molecular structures and the observation do not start with so much uncertainty of identification.

In practice, the relevance of ALife research to the phenomenon of life in general hinges on two steps. The first step is the assumption that fundamental characteristics of life can be captured by computational experiments outside of the physicochemical environment that life on earth is built upon. This assumption is usually made explicit by definitions of life that equate life with evolutionary processes. The second step is the claim that a particular ALife experiment indeed exhibits evolutionary behavior. This claim then often leads to analogies with and suggestions about the nature of life. Usually, evolutionary behavior is claimed with informal arguments. Artificial evolv-

[1] Some philosophers of biology, even have hypothesized that all living systems have an innate ability to detect life [Pinker02].

ing system, which follow this line of argument include, Echo [FJ94], [HJF97], Tierra [Tierra-Webpage], and Avida [Avida-Webpage].

We question whether these informal arguments are sufficient to support the presence of an extremely complex phenomenon such as evolution. Without formal foundations to ascertain these informally presented claims, there is always a danger to run into conflicting arguments, which might be based upon observations of the chemistry on different levels. For example, observations can be made upon the real life universe on several levels including on atomic level, on molecular level, on genetic level, on cellular level, on organismic level and so on. In such a situation informal arguments to conclude the levels at which evolution is effective will remain inconclusive. We encounter similar examples in ALife studies, for example in case of Langton loops (discussed in Chapter 8), where observations can be made upon the level of individual cells of the lattice as well as on the level of very specific subsets of cells. We note that even though individual cells increase in number, the accepted entities for self replication are loops consisting of several consecutive cells. In order to provide a more solid foundation for ALife experiments, we propose in this thesis an observation-based framework as a high level abstraction mechanism for the ALife models. The work aims to provide a formal algebraic framework for characterizing the observations needed to establish evolutionary behavior in ALife studies.

The central concept of the framework is the formalization of the observation process, which we believe is essential, but most often remains implicit in ALife studies. These observations lead to abstractions on the model universe and are consequently used for establishing the necessary elements and the level of evolution of life in the model. The framework is sufficiently generic to be applicable to a wide variety of ALife studies, as exemplified with known examples of Cellular Automata based Langton Loops in Chapter 8 and Lambda calculus based Algorithmic Chemistry in Chapter 9, as well as new experimental chemistries described in Chapter 10 and 11. The framework does not depend upon the low-level dynamics and structure of the underlying model universe of the particular ALife study at hand, which permits the study of higher-level emergent phenomena as the basis of evolutionary processes. We will refer these ALife models as *chemistries* in our discussion; a survey of Artificial Chemistries also appears in [DZB01].

3.3 Contributions

The main contribution of this thesis is to bring insights from evolutionary theory for real life into the realm of artificial life by defining a formal algebraic framework for observational processes, which are needed for the identification of life-like phenomena in any ALife model.

The Concept of Observation

We bring the implicitly assumed notion of *observations to be carried out independent of the underlying chemistry structure* into main focus of ALife studies. It was not clear before that observational processes can be independently studied in their own right and the work presented in this thesis makes it clear by placing observations into distinct formal algebraic platform. By independence we mean that formalism to study observations is generic enough to be applicable to arbitrary models and specific differences among the models do not affect the applicability and analysis. One important property of such study is to make multi-level observations clearly distinct concept, which will be elaborated using the example case study presented in Chapter 10. Thus the work can also be seen as an attempt to fulfill the need for explicitly separating the chemistry from the abstractions used to describe it.

The Observation Process

We formally elaborate in algebraic terms the necessary and sufficient steps for an observational process, to be employed by an ALife researcher upon the time progressive model of his model universe, to uncover (hidden) life-like phenomena in the light of Darwinian evolution as defining characteristics of life (Part II, Chapters 4–6). The observation process as specified in our framework may be carried out manually or can be alternatively algorithmically programmed and integrated with the model.

The Inference Process

We specify necessary conditions, as axioms, which must be satisfied by the outcomes of observations made upon the model universe in order to infer whether life-like phenomena is present in the model universe of study (Chapters 6). These axioms also specify the

experimental work necessary in order to observe and lay claims for the presence of life in the model universe. The axioms are defined in algebraic as well as statistical terms.

The Case Studies

The generic formal framework is instantiated by different kinds of concrete ALife models with different levels of evolutionary processes at work. Two of the case studies presented in Chapters 10 and 11 are entirely new, which demonstrate the wider applicability of the framework.

Finally these case studies are used to draw useful design suggestions (Chapter 13) for the ALife researchers so that interesting evolutionary phenomena involving life-like entities can be observed in the model and one can understand possible evolutionary dynamics better.

3.4 Thesis Outline

The thesis is organized as follows: In Chapters 4–6 in Part II, we will formally elaborate the framework with an intuitive running example of a hypothetical binary string based chemistry, which will be used to explain and elaborate the concepts and design of the framework. Case studies will follow in Part III of the thesis. Chapter 8 applies the framework to cellular automata based Langton's Loops and Chapter 9 on Lambda calculus based Algorithmic chemistry [Fontana92]. Chapters 10 and 11 discuss how we can instantiate the framework on new experimental chemistries - one the Reduced Instruction Set Artificial Chemistry and another the Artificial Graph Chemistry. The concluding part IV provides a final discussion of the work. Chapter 12 presents a discussion of related work, and is followed by concluding remarks in Chapter 13, which presents a discussion on design suggestions for the ALife researchers together with the limitations of the work as well as pointers for further work.

Part II

The Framework

Chapter 4

Fundamental Components of the Framework

Before we formally present the framework in forthcoming chapters, let us define some basic terminological concepts, which are used to elaborate the framework. These concepts form the basis of the underlying mechanism of the framework.

4.1 The Chemistry Structure

In the formal discussion of the framework we will use the term *Chemistry Structure* to refer to an underlying (fundamental) element of an ALife model (chemistry). In general there is no rule to characterize as to what should be treated as a fundamental element of a given chemistry and quite often it is entirely dependent upon the underlying chemistry structure and its design. Nonetheless in the current scope of the framework we will focus our attention to the dynamic progression of chemistry during its simulations such that any suitable definition which characterizes different states of the chemistry with time progression built into it will suffice the purpose. Thus a sequence of states of the chemistry during its simulation will give us the required "chemistry structures".

4.2 Observer Decisions

Observer Decisions will be used to refer to specific observations and corresponding abstractions made by the observation process on the chemistry during its simulation.

Under the scope of the current framework these observations need to be specifically related to possible evolutionary phenomena effective in the chemistry. These observations, in the current context are made upon the "entities" and their "evolutionary relationships". An interesting aspect of these observations is that different observers (observational processes) might well observe the same chemistry but abstract differently from their observations. Though we will not discuss such multi level observation in the current scope of the thesis, it provides interesting direction for further work.

In this thesis, we limit our attention to only those observations having evolutionary significance, though many other observations can be made upon the chemistry (see Chapter 13). As explained in the previous chapter, the main conceptual foundation of our framework is the population centric evolutionist approach of defining life as usually discussed in 'neo-Darwinistic' models for real life evolution [SS97, SS99]. This approach usually hinges upon the observations made only upon the evolutionarily active entities in the model and considers their reproductive relationship with each other. In essence we favor Bedau's view that the property of "life" is to be attributed to the system, not individual entities.

4.3 Auxiliary Formal Structures

We reserve the term *Definition* to refer to an auxiliary mathematical structure to be used by the observer in the intermediate stages of analysis and defined in terms of the Observer Decisions and other such Definitions. Note that the concepts defined as Definitions are not the observations made by the observer upon a chemistry but are aimed to add mathematical convenience for the purpose of analysis of the observations.

As an example of such auxiliary structures, we can consider the similarity measure which an observer defines to determine the limit under which two different entities in the same state or a single entity changing across states can be considered alike. The immediate implication of this is that the framework limits the amount of changes entities can have in order to keep an observer making suitable associations between them. This is significant for an observer who is witnessing the progression of the chemistry only from a macro level perspective without getting into micro level details of the underlying reactions, since this enables him to recognize the continuation of entities across states

under changes as well as to make plausible associations between entities for reproductive descendence based upon the similarities through observations.

4.4 Fundamental Axioms

Axioms will be used to define the fundamental conditions which need to be established by any observer (using the Observer Decisions and Auxiliary Definitions) in order to infer various components of the evolution. Thus for each fundamental component of evolution: self reproduction, mutation, heredity, and natural selection, we have underlined certain Axioms which formally specify what is needed to be observed and consequently inferred in a formal way if any claim towards presence of any of these evolutionary components has to be substantiated. The aim is to define these formal Axioms such that only valid claims for evolutionary processes in a chemistry can be entertained.

Consider the case of natural selection as an example where we believe any claim for establishing natural selection has to require observations across statistically significant number of generations of populations. Therefore we define a corresponding axiom to guide the designer of a chemistry to run the simulations for a sufficiently long time in order to be able to justify that evolutionary processes of natural selection are in effect. As can be seen, the underlying motivation is to render the claims sufficiently general and not based upon specific selected observations.

Having defined the main fundamental conceptual components of the framework, we will now present the formal structure of the framework in the next two chapters.

Chapter 5

The Formal Structure of the Framework

In this chapter we will present the formal algebraic analysis of the chemistry and the entities to be observed. We will consider a simple running example of binary string based chemistry throughout the discussion to assist the intuition behind the formalism. We will use **CBS** (Chemistry of Binary Strings) to refer to this chemistry.

5.1 The Observation Process

We define the observation process operating during the simulation of the chemistry as a transformation from the underlying universe of the chemistry to a set of observer decisions as follows:

Definition 1. *An observation process Obj is defined as a computable transformation from the underlying chemistry structure* $\Gamma = (\Sigma, \mathcal{T})$ *to observer decisions* $\Pi = (E, F, \Upsilon, D, \delta_{mut}, \delta_{rep\_mut}, C)$ *and represented as* $\Gamma \mapsto_{Obj} \Pi$. Γ, *and* Π *are defined below.*

The condition of computability is used to serve two distinct goals - it is to ensure that the framework is tractable, that is, the observation process involves computable steps and can be algorithmically programmed by the designer of the chemistry and also to ensure that non computable observation processes defined in terms of non verifiable claims, for example, information based claims, can be avoided. To see the relevance of the computability, consider the case of (Turing computable) chemistries like Avida

[Avida-Webpage], Tierra [Tierra-Webpage], and Coreworld [RKFH90, RKF91], where an (non computable) observation process might define similarity measure between two "program entities" based upon their "functional equivalence", which is a known undecidable problem [HMU00] rendering the observation process undecidable and cannot be implemented in any practical sense.

We will use the term *observer* instead of *observation process* in personified mode in the rest of the thesis. Although we will refer to observers and their decisions in such a personified mode in the following discussions, it should be clear that such use should not hide the mathematical meaning of above definition. This is done only to add convenience in the presentation.

5.2 The Chemistry

Chemistry Structure 1. Σ*: set of observed states of the chemistry across all executions.*

As discussed in the previous chapter, exact definition of a "state" varies from one chemistry to the other due to their irreducible design differences. For example, an observed state in the case of our example chemistry of binary strings, **CBS**, might be a multiset of binary strings - such that a specific state - could consist of the multiset

$$\{00101, 00101, 10101, 010, 1110\}.$$

Quite often a multiset [1] representation can be adequately used to describe a state of a chemistry by defining it as collection of entities and their corresponding multiplicities present in the chemistry at any instance during its simulation. Further illustrative examples can be seen in the chemistries discussed in case studies appearing in Chapters 8, 9.

Chemistry Structure 2. $\mathcal{T}$*: set of observed sequences of states, ordered with respect to the temporal progression of the chemistry. Each such sequence represents one* execution *or* run *of the chemistry.*

[1] A *multiset* M on a set E is a mapping associating nonnegative integers with each element of E, $M : E \rightarrow \mathcal{N}$. For $e \in E$, $M(e)$ is called its *multiplicity* in the multiset.

A temporally ordered state sequence is one of the basic building blocks in the framework upon which all other observed abstractions are made. Such a definition of a run of chemistry implicitly implies that the framework is fundamentally based upon the dynamic simulations of the chemistry and not upon static analytical inferences. This is in accord with the weaker notion of "emergence" as discussed in Chapter 1 which is generic characteristic of most of the ALife studies.

Because of the temporal ordering of states in a state sequence in $\mathcal{T}$, we can define a natural isomorphism from the a state sequence to an ordered sequence of integers. And so we override and use arithmetic operators with corresponding meaning for state sequences. For example, in a given state sequence $T \in \mathcal{T}$, containing states S and S', we write $S < S'$ to indicate that S appears before S'. Similarly for a state S, $S-1$, and $S+1$ denote the the states just before and after S in the state sequence.

In case of our hypothetical binary string based chemistry, **CBS**, we allow the binary strings to react to each other and give rise to new binary strings, though the underlying reaction mechanism need not to be specified in the spirit of high level observations to be made by the observer by which he can notice the appearance and disappearance of these strings and can thus can define changes in the state of the chemistry and the corresponding state sequences.

Once the fundamental underlying structure of the chemistry $\Gamma = (\Sigma, \mathcal{T})$ is precisely defined, we require the observer to identify the entities of interest.

5.3 Observations and Abstractions

Observer Decision 1. *E: set of tagged entities observed and uniquely identified by the observer within a state and across the states of the chemistry.*

An particular observer is free to select his own set of entities in a given state of the chemistry, though he is not allowed to identify different sets of entities in two states which are the same. Thus two different observers might well choose to identify two different sets of entities which might then lead them to draw different conclusions in subsequent stages. This arbitration in defining E is both a strength as well as weakness of the framework. The strength lies in the fact that there is no single well defined computable criterion as to what should be the set of entities in any given arbitrary

chemistry, which can be used to establish the evolutionary processes. Therefore at best it can be kept open till the actual instantiation of the framework on a given chemistry. On the downside by keeping it arbitrary we are allowing potentially infinite observers for a single chemistry while possibly only one of them might be sufficient to establish evolutionary processes under the current scope of the framework.

Next we use "tagging" as an basic mechanism to be employed by every observer for the identification of individual entities. We require that the observer associates and correspondingly identifies each and every entity in a state using a different tag. We even demand different tags in different states for the "same" entity. This is primarily for the mathematical convenience in the framework, and is not an stringent requirement and can always be satisfied in each such case where the observer employs reliable tagging/identification mechanism. In the following discussions every entity will be assumed to be associated and identified with a tag. We will use for these *tagged entities* only term *entities* in the rest of the discussion.

In case of **CBS**, we associate with every binary string an integer tag such that with tag i, an entity corresponding to the binary string s will be represented as $[s, i]$. A possible set of entities corresponding to the example state given above is

$$\{[00101, 1], [00101, 2], [10101, 3], [010, 4], [1110, 5]\}.$$

Observer Decision 2. $F : E \rightarrow \Sigma$. *The function F returns the state in the state sequence in which a particular entity is observed.*

An observer will use, as discussed later, the state information provided by F for entities to define valid evolutionary relationships between them. In general the observers may use different mechanisms based upon the nature of chemistry as well as the entities defined, to determine the state for a given entity. For example, as a simple mechanism, in case of **CBS**, the observer can maintain a table mapping entities to their corresponding states in order to define F. Other alternatives for defining F can be considered where the tag i in $[s, i]$ is selected such that the observer can determine the state of the entity by the tag i alone.

Having defined the sequence of states in a chemistry with temporal ordering and the entities identified by their tags, we will now proceed to discuss how an observer might

define the detailed observable characteristics for such entities. Using these characteristics he will draw evolutionary relationships, e.g., descendent relationship, as well as establish presence of other components of evolution, e.g., heredity and variation.

To this aim, we will employ the concept of 'nomological space' [MB97], which is a simple generalization of $\mathcal{R}^n$, ($\mathcal{R}$ is the set of real numbers), where one can alternately use *symbolic* values to determine specific characteristics without quantifying them.

Observer Decision 3. *An observer defines the set of all possible orthogonal measurable characteristics for every possible entity in the chemistry. Formally, the observer defines a multi dimensional nomological space* $\Upsilon = Char_1 \times Char_2 \times \ldots \times Char_n$*, where each of* $Char_i$ *is the set describing values for* i^{th} *characteristics. Each of* $Char_i$ *make one dimension in the nomological space* Υ *and each entity* $e \in E$ *is a point, say* $(v_1, v_2, \ldots v_n)$*, where* $v_i \in Char_i$*, in* Υ*.*

For a vector $x = (a_1, a_2, \ldots, a_r), i^{th}$ element (a_i) will be denoted as $x[i]$. Associated with each characteristics $char_i \in \Upsilon$ is a total ordering relation $\leq_i$ such that the absence of any characteristics in an entity is represented by 0_{char_i} $(\forall v_{char_i} \in Char_i.\ 0_{char_i} \leq_i v_{char_i})$

In case of **CBS**, for simplicity we may assume that chemistry consists of binary strings of size n, which can be some large finite positive integer. In that case each position of the string can represent one orthogonal dimension and we have only two binary values ($\{0, 1\}$) at any position in a string for corresponding dimension. Thus nomological space Υ in **CBS** is n dimensional binary hypercube with each string occupying a possible diagonal end point. We will represent this hypercube as $\{0, 1\}^n$. The ordering relation $\leq$ for all dimensions is the same and defined as $0 < 1$, $0 = 0$, and $1 = 1$.

5.4 The Clustering Distance Measure

Another important structure in the framework is the "dissimilarity measure" (D) to define the "observable differences" (**Diff**) between the characteristics of the entities in a population. The distance measure defined below has an important property that it can be used by the observer to distribute entities into separate clusters such that all the entities in the same cluster are sufficiently similar while entities from different clusters are distinguishably different in their characters. Again exact definition of distance function is chemistry dependent.

Observer Decision 4. *An observer defines a decidable clustering distance measure* $D : E \times E \rightarrow$ **Diff**, *where* **Diff** *is the set of values to characterize the observable "differences" between entities in* E.

The range of distance measure D, **Diff** is indeed a *vector* of values. Let **Diff** $= (\mathit{diff}_0, \mathit{diff}_1, \ldots, \mathit{diff}_n)$, such that each element diff_i represents the difference in the values for one particular characteristics Char_i between entities.

The partial ordering relation $\leq_i$ defined for the values in Char_i also induces a partial ordering relation to compare the "differences" among the entities for each of the characteristics $\mathit{Char}_i \in \Upsilon$. Formally, $\preceq_i \subseteq \mathit{diff}_i \times \mathit{diff}_i$ will be be used to compare the relative differences among the entities for Char_i. The least element 0_{diff_i} is used when there is no observed difference between two entities for Char_i. $\preceq_i$ and $\leq_i$ are related in a consistent way, that is, $\forall e_1, e_2, e_3 \in E$ we have the following - $D(e_1, e_2)[i] = 0_{\mathit{diff}_i} \Leftrightarrow e_1[i] = e_2[i]$, $e_1[i] \leq_i e_2[i] \leq_i e_3[i] \Rightarrow D(e_1, e_2)[i] \preceq_i D(e_1, e_3)[i] \wedge D(e_2, e_3)[i] \preceq_i D(e_1, e_3)[i]$.

In some cases, it might be convenient to choose **Diff** as a vector of values from the set of non negative integers N^+. Hamming distance is used to define the distance between genomic strings in the Eigen's model of molecular evolution [Schuster01]. Set of points where two computable functions differ in their function graphs, or the set of instructions where two programs may differ, etc are some other examples of such distance measures, which can be suitably used with ALife models. One of the known criterions to define the concept of species is "phenotype similarity" [Ridley96], which can be seen as another example for distance measure.

In case of **CBS**, we define an auxiliary function $\oplus : \{0,1\} \times \{0,1\} \rightarrow \{0,1\}$ as a binary XOR such that we have $0 \oplus 0 = 1 \oplus 1 = 0$, and $1 \oplus 0 = 0 \oplus 1 = 1$. Thus the clustering distance measure $D : E \times E \rightarrow \{0,1\}^n$ is defined such that $\forall i.D(e_1, e_2)[i] = e_1[i] \oplus e_2[i]$, which implies that **Diff** $= \{0,1\}^n$. For example in case of two $n = 3$ bit binary strings (with tags) $e_1 = [001, 1]$ and $e_2 = [101, 2]$, $D(e_1, e_2) = 100$. Other alternatives may include Hamming distance measure $D(e_1, e_2) = \sum_{i=1}^{n}(e_1[i] \oplus e_2[i])$ with **Diff** $= \{0, 1, \ldots, n\}$.

5.5 Observable Limits on Mutational Changes

As briefly discussed previously, the observer needs to specify the limits (δ_{mut}) under which it can recognize an entity across states even in the presence of mutational changes due to the interactions with the environment. This is an inherent limiting property on the part of the observer and can be different for different observers. Based upon the limit, an observer can establish whether two entities in two different states (successive) are indeed the same with differences owning to mutations or not. The smaller the limit, the harder it will be for an observer to keep on recognizing entities across states and he would be counting mutated entities as the new entities. Also note that as we observe entities in more and more refined levels of details, their apparent similarities melt away and differences become sharply noticeable.

Similarly another source of mutations arises during reproduction, in which case an observer has to identify whether an entity is indeed an descendent of another entity in previous state even though they might not be exactly the same. This limit on observable reproductive mutations (δ_{rep_mut}) indeed becomes significant in chemistries where epigenetic development plays crucial role [MB97]. This is because in such chemistries including examples from real life, the "child" entity and the "parent" entities do not resemble with each other at the beginning and observer has to wait until whole epigenetic developmental process has unfounded and then compare these entities for their characteristics. Here the observable limits set by an observer assist him to establish whether a particular entity should be treated as "descendent" of another entity or not. Another major reason for introducing the concept of "limit on the observable reproductive mutations" is that from the view point if an high level observation process not recording every micro level details, it is quite essential to distinguish a parent entity with other secondary entities involved in reproductive process. Consider, for example, a chemistry where entity A reproduces according to reaction $A + B \rightarrow 2A' + C$, where A' is mutant child entity of A, which can be determined by an observer only when he can establish that A and A' are sufficiently similar with respect to their characteristics, while A' and B are not. These limits on observable differences are formally defined as follows:

Observer Decision 5. *Based upon the choice of clustering distance measure D, the observer selects some suitable $\delta_{mut}, \delta_{rep\_mut} \in$* **Diff***, which will be used later to bound*

mutational changes (both reproductive and otherwise) for proper recognition. δ_{mut} *and* δ_{rep_mut} *are vectors such that each element specifies an observer-defined threshold on the recognizable mutational changes for corresponding characteristics.*

In case of **CBS**, we can choose any vector in $\{0,1\}^n$ for δ_{mut} and δ_{rep_mut}. Before we select such a vector, let us see what is the intuitive meaning of 0 and 1 for the observer. In case, the observer selects a δ_{mut} or δ_{rep_mut} such that i^{th} bit is 0, then that means for the observer every difference in the values for i^{th} dimension between two entities is observable, whereas 1 will be case where the observer cannot distinguish any differences. Thus $\delta_{mut} = (00\ldots0)$ distinguishes every single string in 2^n from other strings and therefore cannot recognize entities in next state if they mutate at any position, whereas $\delta_{mut} = (11\ldots1)$ means the observer will ignore all the differences between the entities and thus can even recognize an entity in next state with the presence of mutational changes at any position. δ_{rep_mut} is similarly used to specify the observable limit on reproductive mutations.

Let us consider an intermediate situation where for example the $\delta_{mut} \in 2^n$ is selected such that $\forall i \leq n$, $\delta_{mut}[i] = 1$ if i is a multiple of 3 , otherwise $\delta_{mut}[i] = 0$. Thus, the observer ignores all the differences between values at positions $0, 3, 6, \ldots$ while for other positions it can notice the differences. Let us consider that δ_{rep_mut} is chosen such that $\forall i \leq n$, $\delta_{rep\_mut}[i] = 1$ if i is a multiple of 2 , otherwise $\delta_{mut}[i] = 0$. Other similar examples can be where an observer with some memory limitations is forced to ignore mutational changes beyond certain number of bits etc.

Having defined the observation process as a computable transformation from the underlying states of the chemistry to the set of components involving entities, their tags and their observable characteristics with measurable differences as well as limits such differences, we will proceed in the next chapter with formalization of the evolutionary components as discussed in Section 1.3.1.

Chapter 6

Evolutionary Components

Aim of this chapter is to formalize the evolution for ALife models based upon the background structure presented in the last chapter. The evolution is formalized in terms of the four basic conditions described in the Section 1.3.1: the variations caused by the mutations, the reproduction, the heredity, and the natural selection.

6.1 Mutations

Entities can change (mutate) over the course of their life times because of their interaction with the environment (other entities.) Moreover, there can be observable differences among the child and the parent entities after reproduction (and further developments.) Some of these mutation may or may not be inheritable. For example in case of higher order organisms in real biology, phenotypic changes, which continuously occur over the course of life time of an organism, are not transferred to the next generations owing to the molecular barrier (known as the central dogma of molecular biology), while on the other hand in case of bacteria, mutations which are the result of horizontal gene transfer among bacteria are also vertically transferred to the future generations.

The observer uses the following formalizations to establish these mutational changes as well as reproductive mutations.

Definition 2. *The observer establishes recognition of entities across states of the chemistry with (or without) mutations by defining the function* $\mathbf{R}_{\delta_{\mathbf{mut}}}$*:* $E \rightsquigarrow E$*, which is a partial function and satisfies the following axioms:*

Axiom 1. $\forall e, e' \in E \;.\; \mathbf{R}_{\delta_{\mathbf{mut}}}(e) = e' \Rightarrow F(e') = F(e) + 1.$

Informally, the axiom states that entities to be recognized as the same even with mutational changes have be observed in successive states. $\mathbf{R}_{\delta_{\mathbf{mut}}}$ is anti symmetric to ensure that entities are recognized based upon the time progression of the chemistry not in any other arbitrary order..

Axiom 2. $\mathbf{R}_{\delta_{\mathbf{mut}}}$ *is an injective function, that is,* $\forall e, e' \in E.\ \mathbf{R}_{\delta_{\mathbf{mut}}}(e) = \mathbf{R}_{\delta_{\mathbf{mut}}}(e') \Rightarrow e = e'$

Informally, the axiom states that no two different entities in one state can be recognized as the same in the next state. ${\mathbf{R}_{\delta_{\mathbf{mut}}}}^{+}$ is the transitive closure of $\mathbf{R}_{\delta_{\mathbf{mut}}}$.

Axiom 3. $\forall e, e' \in E.\ \forall Char_i \in \Upsilon.\ \mathbf{R}_{\delta_{\mathbf{mut}}}(e) = e' \Rightarrow 0_{diff_i} \preceq_i D(e, e')[i] \preceq_i \delta_{mut}[i]$

Informally, $\forall e \in E$, $\mathbf{R}_{\delta_{\mathbf{mut}}}(e)$ is that $e' \in E$, which is recognized in the next state by the observer as e in the previous state with possible mutations bounded by δ_{mut}. In other words if entity e mutates and changes in the next states and identified as e', then if these changes (between e and e') are bounded by δ_{mut}, then the observer will be able to recognize e and e' as the same.

The above three axioms formally characterize mutations in the spirit of the framework based upon observations. Mutational changes are primary sources of variation in characteristics among entities the population (see Section 1.3.1). In case of our example chemistry **CBS**, let us assume that random one bit mutations take place. Thus at every state of the chemistry some of the binary strings in the population may undergo one bit mutations. For example for $n = 5$ bit chemistry, say string 10010 mutates to 10110 in the next state. Since δ_{mut} has been defined such that the observer can recognize entities in the next state even with mutations at position 3 so in this case the observer will be able to recognize the entities (with tags) $[10010, 1]$ and $[10110, 2]$ as the same. On the other hand, if mutation takes place at position 1 or 2, then the observer will consider those mutated entities as different entities in the next state and will not recognize them as same with those in previous state.

6.2 Reproduction

Reproduction is one of the fundamental components of evolution as discussed in Section 1.3.1. Through reproduction, entities pass on their characteristics to the next

generation and increase the size of the population. In our framework, the way an observer establishes reproduction is by providing observed evidence for it. He does so by defining the descendence relationships among the entities. The parent and the child entities are recognized by the observer as being "causally" connected across the states. Using Δ we formally define that their differences due to reproductive mutations are also bounded by δ_{rep_mut}.

Definition 3. $\Delta \subseteq E \times E$ *such that* $\forall e, e' \in E \,.\, (e, e') \in \Delta \Leftrightarrow \forall Char_i \in \Upsilon \,.\, D(e, e')[i] \preceq_i \delta_{rep\_mut}[i]$.

Informally for (e, e') to be in Δ, their differences for each single characteristic $Char_i$ must be bounded by $\delta_{rep\_mut}[i]$. Next, the observed causality between entities is captured using the following relation:

Observer Decision 6. *$C \subseteq E \times E$. C establishes the observed causality among the entities appearing in the successive states. C satisfies the following axiom:*

Axiom 4. $\forall e, e' \in E \,.\, (e, e') \in C \Rightarrow [F(e') = F(e)+1] \wedge [\nexists e'' \in E \,.\, F(e'') = F(e) \wedge \mathbf{R}_{\delta_{\mathbf{mut}}}(e'') = e']$

Informally, the above axiom on causal relationship C states that, if an entity e is causally connected to another entity e', then the observer must observe e' in the next state of e and never before. This is to ensure that mutations are not confused by the observer with reproductions. Notice that in order to establish causal relation between entities, the observer need not to necessarily know the underlying reaction rules of the chemistry which determine the result of the reactions between the entities in the chemistry. We only require that the observer's claimed causality conforms with the stated axiom.

Based on the thus established notion of "causal" relationships between entities, we define an **AncestorOf** relation, which connects entities for which an observer can establish descendence relation across generations.

Definition 4. AncestorOf $= (\, (C \cup \mathbf{R}_{\delta_{\mathbf{mut}}})^+ \cap \Delta)^+$

In this definition the (inner) transitive closure of $(C \cup \mathbf{R}_{\delta_{\mathbf{mut}}})$ captures the observed causality (C) across multiple states even in cases when "parent" entities might undergo

mutational changes ($\mathbf{R}_{\delta_{\mathbf{mut}}}$) before "child" entities complete their "epigenetic" maturation with possible reproductive mutations. Intersection with Δ ensures that causally related parent and child entities are not too different from each other, that is, reproductive mutational changes are under observable limit. Outer transitive closure is to make **AncestorOf** relationship transitive in nature so that entities in the same lineage can be related with each other.

For $e, e' \in E$, $(e, e') \in$ **AncestorOf**, describes that e is observed as an ancestor of e'. Based upon the **AncestorOf** relation, we now can consider the cases of *sufficient self reproduction* and *Fecundity*:

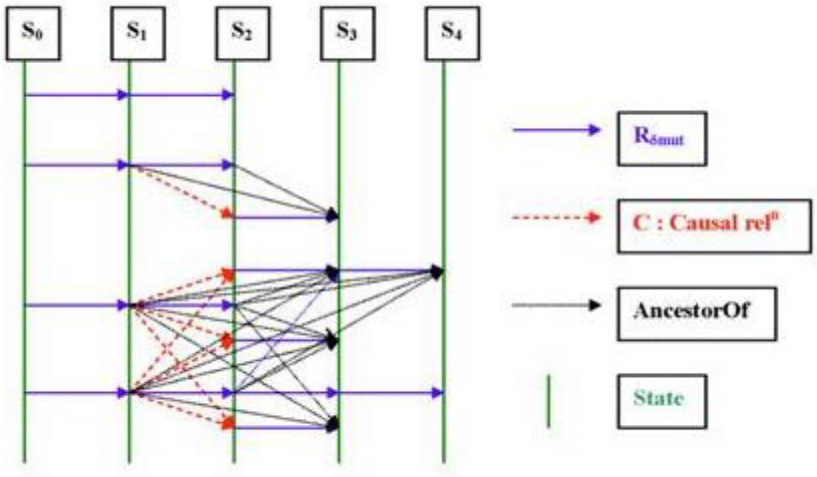

Figure 6.1: Graphical view of the relationships between entities in successive states. Recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$, Causal relation C, and **AncestorOf**.

Figure 6.1 depicts graphically the relationships between entities in successive states. Vertical lines represent the states $(S_0, S_1, S_2, S_3, S_4)$. Various kinds of arrows represent different relationships: recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$, causal relation C, and **AncestorOf**. The end points of the arrows on state lines represent entities.

6.2.1 Case 1: Sufficient Self Reproduction

We consider the case where parent entity reproduces child entities with or without reproductive mutations. For a given run $T \in \mathcal{T}$ of the chemistry, an observer defines the following $\mathbf{Parent}_{\mathbf{\Delta}}$ relation:

Definition 5.

$$\mathbf{Parent_{\Delta}} = \{(p,c) \in \mathbf{AncestorOf} \mid \nexists e \in E \;.\; [(p,e) \in \mathbf{AncestorOf} \wedge (e,c) \in \mathbf{AncestorOf}]\}$$

The condition in defining $\mathbf{Parent_{\Delta}}$ is used to ensure that p is the immediate parent of c and thus there is no intermediate ancestor e between p and c. Using $\mathbf{Parent_{\Delta}}$ relation, in order for the observer to establish reproduction in the chemistry, the following axiom should be satisfied:

Axiom 5 (Reproduction)**.** $\exists$*state sequence* $T \in \mathcal{T}$. $\mathbf{Parent_{\Delta}} \neq \emptyset$

This means, if there is reproduction in the chemistry, then there exists some run $T \in \mathcal{T}$ of the chemistry, where at least one instance of reproduction is observed.

In case of **CBS**, we consider a very simple model of reproduction, where at any state of the chemistry some of the strings are randomly chosen and are copied with some random errors. How it is done remains hidden from the observer but the observer can observe which parent entities are chosen for copying and can establish causal relation between these parent and their copied child entities if the random errors occur only at even positions as the way δ_{rep_mut} has been defined in Section 5.5. It is obvious that under such construction scheme *Axiom of Reproduction* will be satisfied.

6.2.2 Case 2: Fecundity

In case of fecundity, an observer need not to observe all the parents in the same state, nor do children need to be observed in the same states of the chemistry. We require the observer to establish Fecundity by satisfying the following axiom

Axiom 6 (Fecundity)**.** $\exists G \subseteq E \;.\; |G| > |\{p \in E | \exists c \in G \;.\; (p,c) \in \mathbf{Parent}\}|$

This axiom ensures that in case of fecundity, the number of child entities in successive generations are effectively more than the number of parent entities.

The axioms of Reproduction and Fecundity formally characterize the first essential component of evolution, that is, reproduction as discussed in Section 1.3.1.

To illustrate fecundity in our example chemistry **CBS**, let us consider another copying process which makes several copies of certain type of entity if at least r copies of

that type of entities are present in the chemistry at any state. Here two entities are said to be of the same type only if they are exactly the same (under string comparison) irrespective of their tags. Therefore an observer can notice the fecundity at any state of the chemistry when at least one of the entity types had population size more than r. Because there are no errors introduced during this copying process, parent-child relationship will be established correctly. *Axiom of Fecundity* thus naturally holds under this set up.

We can now formulate an important axiom from evolutionary perspective, which asserts that reproduction in the chemistry does not cease because of (harmful) mutations.

Axiom 7 (Continuity of Reproduction under Mutations)**.** *Some mutations do preserve reproduction.*

Formally, $\exists e \in E \;.\; Ch_e = \{e' \in E : (e, e') \in \mathbf{Parent_{\Delta}} \cup \mathbf{R}_{\delta_{\mathbf{mut}}} \} \neq \emptyset \Rightarrow \exists e'' \in Ch_e \;.\; \{e' \in E : (e'', e') \in \mathbf{Parent_{\Delta}}\} \neq \emptyset$

Informally, this means, there exists entity $e \in E$, which reproduces (with mutations) and one of those (mutant) children of e can also further reproduce. Ch_e denotes the set of children of e. *This axiom formally characterizes the essential condition of persistence of reproduction (Section 1.3.1.)*

In case of **CBS**, since copying mechanisms do not work differently based upon selected entities, hence the errors during copying process do preserve the above axiom of *Continuity of Reproduction under Mutations.*

6.3 Heredity

The continuity axiom defined above does not specify anything about the inheritance of mutational changes into the future generations. This is partly because of the fact that heredity in the presence of mutations necessarily requires further mechanisms to reduce possible undoing of current mutations in future generations owing to other new mutations (Section 1.3.3). In the real biology, this is achieved with the effect of several mechanisms including error correction mechanisms, relatively low mutation rates and short reproduction cycles (high fecundity.) The long term end result of mutations and their inheritance along with other environmental factors (e.g., geographical isolation etc.)

is observed in "speciation", where new "kind of" (reproductively isolated) population of entities emerge over the course of evolution [Ridley97]. Nonetheless we may not observe heredity in every case, e.g., heredity is not very strong and possibly does not lead to speciation in most of the microbiological organisms involving bacteria.

Each mutation can be thought of carrying two kinds of changes in the entities: one where the mutant entity (after mutations) has changed values for the characteristics, e.g., mutation changing the height and/or color character of the entities in real life organisms, secondly where mutations alter the characteristics entirely such that mutant entity has at least one new character not present before or when certain characteristics are lost. In terms of values in nomological space Υ, these effects of mutations can be seen as "distance vectors" joining the entities with their mutants in the same dimension in the first case and across dimensions in the second case.

Thus heredity becomes the property of the reproduction process where some of the variations caused by mutations in the population are carried on to next generation as well.

More precisely, the observer observes sufficiently many generations of reproducing entities and determines that the number of parent - child pairs where certain characteristics were inherited by child entities without further mutations is significantly larger than those cases where mutations altered the characteristics in the child. We can express it as the following axiom:

Axiom 8 (Heredity)**.** *Let a statistically large observed subsequence of a run T:*

$$\Omega = lim_{N\to\infty}\langle S_n, \ldots S_N\rangle, n \ll N$$

Consider $\mathbf{Parent}^{\mathbf{\Omega}}_{\mathbf{\Delta}} = \{(e, e') \in \mathbf{Parent}_{\mathbf{\Delta}} \mid F(e) \in \Omega \wedge F(e') \in \Omega\}$ *to be the set of all parent - child pairs observed in* Ω*. Again let* $\mathbf{Inherited}^{\mathbf{i}}_{\mathbf{\Omega}} = \{(e, e') \in \mathbf{Parent}^{\mathbf{\Omega}}_{\mathbf{\Delta}} \mid \exists Char_i \in \Upsilon \ . \ D(e, e')[i] = 0_{diff_i}\}$ *be the set of those cases of reproduction where* i^{th} *characteristics were inherited without (further) mutation. Then high degree of inheritance for* i^{th} *characteristics* $Char_i$ *implies that* $|\mathbf{Parent}^{\mathbf{\Omega}}_{\mathbf{\Delta}}|/|\mathbf{Inherited}^{\mathbf{i}}_{\mathbf{\Omega}}| \simeq 1$*. For inheritance to be observed in a population of entities, we should have some such characteristics which satisfy this condition.*

The axiom of heredity characterizes the third essential component of evolution as discussed in Section 1.3.1. The axiom of heredity together with the axiom of continuity of reproduction under mutation ensures that reproductive variation is maintained.

It is interesting to see how heredity can be observed in case of **CBS**, because heredity is not directly implied for any characteristics (bit position) if mutation rates are relatively high. Therefore in order to preserve the mutations system need to keep copying errors very low. Since in our chemistry there are two kinds of copying processes taking place - one which copies individual entities but might introduce errors and second copying process which requires certain minimum number of entities of the same type to be present in the population but produces several copies without errors. Therefore in the beginning when there will be more diversity in the chemistry but low density of any particular kind of entity, first copying process will be rather effective and heredity might not be visible or possible. But as soon as diversity reduces and density of similar strings starts increasing, second kind of faithful copying process becomes effective and then the observer can indeed observe heredity for certain characteristics. Thus we can conclude that faithful reproduction when fecundity takes place can effectively give rise to heredity. Notice that in our chemistry, fecundity can be loosely interpreted as some kind of cooperative organization of similar entities, which reduce the copying errors collectively during reproduction (or we might say that it roughly acts like a "*hypercycle*" [ES79]).

Apart from mutations, reproduction and heredity, next most important element of evolution is natural selection, which is described next.

6.4 Natural Selection

There are several existing notions of selection in the literature on evolutionary theory [Futuyma98], [Ridley96], [Ridley97], [Stephen00], [MB97], [Kimura83]. In case of our observation based framework we choose to define natural selection as a *statistical inference* of *average reproductive success* (also known as *Darwinian fitness*), which should be established by an observer on the population of self reproducing entities over an evolutionary time scale i.e., over statistically large number of states in a state sequence. Other notions of selection using fitness, adaptedness, or traits etc. are rather intricate

in nature because these concepts are relative to the specific abstraction of "common environment" shared by entities and "the environment-entity interactions", which are the most basic processes of selection. Nonetheless selecting appropriate generic abstraction for these from the point of view of an observation process is not so simple. Therefore we consider more straightforward approach based upon the idea that on evolutionary scale the relative reproductive success is an effective measure, which is also an indicator of better adaptedness or fitness. We thus define the following (necessary) axioms for the natural selection:

Axiom 9 (Observation on Evolutionary Time Scale)**.** *An Observer must observe statistically significant population of different reproducing entities, say* Λ *(*$|\Lambda| \gg 1$*), for statistically large number of states in a state sequence* $T \in \mathcal{T}$*. That is, for a statistically large subsequence* Ω *of* T*,* $\Omega = lim_{N\to\infty}\langle S_n, \ldots S_N\rangle, n \ll N$*, the observer defines the set of reproducing entities* $\Lambda \subseteq \bigcup_{S_j\in\Omega} SR(S_j)$*, where* $SR(S_j) = \{e \in E| [F(e) = S_j] \wedge [\exists e' \in E \,.\, (e, e') \in \mathbf{Parent_\Delta}]\}$ *is the set of all reproducing entities in state* $S_j \in \Omega$*.*

Axiom 10 (Sorting)**.** *Entities in* Λ *should be different with respect to characteristics in* Υ *and there should exist differential rate of reproduction among these reproducing entities. Rate of reproduction for an entity is the number of child entities it reproduces before undergoing any mutations beyond observable limit.*

In other words, $Rate_{rep} : E \to N^+$ *defined as* $\forall e \in E \,.\, Rate_{rep}(e) = |Child_e|$ *where* $Child_e = \{e' \in E|\exists e'' \in E \,.\, (e'', e') \in \mathbf{Parent_\Delta}$ *and* $[\mathbf{R}^+_{\delta_{\mathbf{mut}}}(e) = e'' \wedge \forall Char_i \in \Upsilon \,.\, D(e, e'') = 0_{diff_i}]\}$*.*

The above two axioms though necessary are not sufficient to establish natural selection since these cannot be use as such to distinguish between natural selection with neutral selection [Stephen00]. The following axioms are therefore needed to sufficiently establish natural selection.

Axiom 11 (Heritable Variation)**.** *There must be variation in heritable mutations in population of* Λ*. Formally, let*

$$Child_{mut} = \{e \in \Lambda|\exists e' \in \Lambda \,.\, (e, e') \in \mathbf{Parent_\Delta} \wedge [\exists Char_i \in \Upsilon \,.\, 0_{diff_i} \prec D(e, e')[i]]\}$$

be the set of child entities carrying reproductive mutations. Let $Var\_Child_{mut} \subseteq Child_{mut}$ *be the set of those child entities which carry different mutations with respect to characteristics in* Υ*, that is,*

$$\forall e, e' \in Var\_Child_{mut} \text{ we have } \exists Char_i \in \Upsilon \ . \ 0_{diff_i} \prec D(e, e')[i]$$

Then axiom of heritable mutation demands that $|Var\_Child_{mut}| \gg 1$*, that is, there are significantly many child entities carrying different mutations.*

Axiom 12 (Correlation)**.** *There must be non zero correlation between heritable variation and differential rate of reproduction. Formally,*

$$\forall Char_i \in \Upsilon \ . \ \forall e, e' \in Var\_Child_{mut} \ . \text{ the following two conditions should hold:}$$

$$i) \ e[i] <_i e'[i] \Leftrightarrow [Rate_{rep}(e) < Rate_{rep}(e')] \vee [Rate_{rep}(e) > Rate_{rep}(e')]$$

$$ii) \ e[i] =_i e'[i] \Leftrightarrow Rate_{rep}(e) = Rate_{rep}(e')$$

Informally, this means as the value of characteristics inherited by the child entity changes, rate of reproduction also changes. Based upon the environmental pressures with respect to a particular characteristics, rate of reproduction might either increase or decrease as the characteristic changes.

The last two axioms state that there must be significant variation in population (in characters) of entities which must be maintained for evolutionarily significant periods and that this variation must be caused by the differences in inheriting mutations from the parent entities, which in turn directly affect the rate of reproduction (refer to the conditions 2 and 4 discussed in Section 1.3.1). In case of real life natural populations show variation in the characters on all levels, from gross morphology to genetic sequences. These variations in characters are generated randomly by mutation and recombination [Ridley97] on genetics material. Based upon the variation in these characters, entities in natural population also vary in their reproductive successes.

In order to illustrate the process of natural selection formalized by the above axioms, we consider very simple model for our example chemistry **CBS** based on *dilution flux*. System imposes a dilution flux which in a uniform way selects as many entities for

deletion from the population at any state, as were introduced anew by copying steps. This simple model of uniform elimination captures each of the above axioms. To see this, notice that number of entities of particular type to be eliminated will be selected based upon their relative frequency or density in the population. Therefore those entity types which can reproduce more, will also survive relatively more in number after the elimination step. Therefore an observer can observe over long periods that there are entities with different rates of reproduction and that these different rates are directly correlated with the similarity based relative density of entities. This supports necessary *axioms of evolutionary scale observation* and *of sorting* with respect to the characteristics. Again notice that fecundity copying process requires certain minimum number of (cooperative) similar entities of same type to be present in the population. Therefore when some mutations occur and new entity type emerges then once this entity type has sufficiently many copies because of subsequent low rate of mutations or due to some other entities mutating and giving rise to the more copies of same entity type, it can start reproducing at relatively higher rate. This is exactly what is required by the *axioms of heritable variation* and *of correlation*. Therefore the resultant effect of dilution flux and both kind of copying processes (with and without errors) induces selection in the chemistry, though in a weak form.

Having formalized the fundamental component of evolutionary processes to be observed in a chemistry, we will illustrate the framework on four different ALife models in the following part of the thesis. These illustrations will later be used in concluding Chapter 13 to extract generic design principles for ALife researchers.

Part III

Case Studies

Chapter 7

General Considerations

Having described the generic formal framework in Part II, which is based upon the concept of observations and axiomatic inferences to establish the level of evolution for ALife studies, in the following chapters, we will apply the formalism to four different artificial life models as case studies. These case studies include Cellular Automata based Langton Loops [Langton87] and Lambda Calculus based Algorithmic Chemistry [Fontana92], as well as new experimental models - the Reduced Instruction Set Artificial Chemistry [Vedvyas05], and the Artificial Graph Chemistry[Peng05]. The case studies elaborate technical details specific to the example universe of the model, which remained implicitly defined in the generalized description of the framework.

For any such arbitrary model, the steps to instantiate the framework can be described as follows: The observation process works on the simulation model which iteratively changes the state based upon the application of the underlying rules of the chemistry. Therefore the observation process starts with the identification of states of the chemistry (Σ) during its simulation (i.e., state sequences $\mathcal{T}$)). Usually any change in the chemistry (i.e. the changes in the set of basic units or the molecules) gives rise to a change of the state. It is important to note that in some cases the chemistry might not change its states even tough there is some underlying activity in the chemistry, that is, when chemistry reaches, for example, a fix point.

For every state in the state sequence, the observation process (or the observer) needs to identify a set of well defined entities with suitable tagging for individual identification (E). These entities need to be described in terms of their characteristics (Υ). Next

important task is to define the limits on the mutational changes in individual characteristics of these possible entities (δ_{mut}, δ_{rep_mut}), which will define the recognition ($\mathbf{R}_{\delta_{\mathbf{mut}}}$) of entities persisting across states of the chemistry as well to define whether two entities might be considered related under descendent relationships.

Once the sets of entities in various successive states of the chemistry as well as their characteristics are known, important evolutionary relationships need to be established between them. These evolutionary relationship depend upon the intermediate causal relation (C) between the entities as observed under the mechanics of observation process. Using the limits on mutational changes as well as causal relationship between entities, we proceed to define the Ancestor (**AncestorOf**) and the Parent sets ($\mathbf{Parent}_{\boldsymbol{\Delta}}$). These sets determine whether there are entities which might be potentially reproducing in the chemistry, even with some changes between parent and child entities (Δ).

Next stage of the observation process in inference proceed to uncover certain statistically significant patterns present in the chemistry to ascertain the effectiveness of natural selection. Based upon the long term observations on the chemistry for statistically large number of generations, one can infer some statistical patterns for degree of heredity and variation. The statistically significant population of entities observed across such large number of generations may provide certain conclusion regarding effectiveness of natural selection in the chemistry. If natural selection has to work then there should exist large number of reproducing entities with significant variation in their characteristics such that there must be some direct correlation of this variation in characters for the reproductive success of the entities.

This process at the end establishes the validity of all or some axioms of the framework for the given chemistry which provides clues to the degree upto which evolutionary processes might be effective in that model universe.

The examples in following chapters will illustrate this process in detail.

Chapter 8

Case Study 1: The Langton Loops

This chapter aims to illustrate the formal framework presented in Chapters 4–6 as a case study on Cellular Automata based Langton Loops [Langton84]. The cellular automata based ALife models offer an ideal example for our observer (observation process) based framework since the replicating structures and their variations can be explained to have evolved only with respect to some high level observation process. The observation process can be used to define reproducing entities (loops) and their evolution. We will illustrate our formal framework by instantiating it on the Cellular automata based Langton loop chemistry in the Section 8.2 after presenting a brief survey of the history of CA based research on self reproduction and other aspects of life.

8.1 History of Cellular Automata based Self Reproduction

Research on the self reproduction has a long cherished history starting in early fifties [Burks70, Sipper98]. Though the current focus of research is understanding the nature of self reproduction in broader sense in any arbitrary model universe, the focus of early research was bit different. That was the time when machine based thinking was mainly predominant and discoveries on the molecular basis of living systems had not come up. That is why the nature of early research was focussed on answering the questions like - "Can a 'nonliving' machine self replicate?" These kinds of questions were based upon some fundamental conceptions on the nature of self reproduction. The most difficult

of these was related to precise meaning of 'information or knowledge of the self'. The analysis of 'the self' would be usually made in terms of 'information' or 'knowledge' of a system about itself and then naturally it could lead to infinite regression. After the pioneering work of Alan Turing in early 40s to define in very fundamental terms the mechanical meaning of 'computation' as a Turing machine transitions, John von Neumann defined Cellular Automata [Neumann66] to explain the generic logic of self reproduction in mechanical terms. His Cellular Automata model was a two dimensional grid divided into cells, where each cell would change state based upon past states of its neighborhood cells, its own state and its transition rule. All cells change their states in parallel. For such CA model, von Neumann defined a virtual configuration space where he demonstrated analytically that there existed some universal replicator configuration which could replicate other configurations as well as itself. Though universal replicators are not found in nature and the such self replicator was extremely large in its size, nonetheless the underlying logic of treating states of cells in the grid both as 'data' as well as 'instruction' was very fundamental contribution of this model and that was exactly was was discovered later in case of real life where DNA sequences specify both transcription as well as translation for their own replication in a cell. Another strength of von Neumann's formulation was its ability to give rise to unlimited variety of self replicators [McMullin00a, McMullin00b]. Over the years this model was simplified and reduced in size considerably [Codd68, 81 - 105].

Finally Langton introduced loop like self replicating structures in [Langton84], which retained the 'transcription - translation' property of von Neumann's model but removed the capabilities of universal replication and symbolic computation. These loops were extremely simplified self replicating configurations, which have since then, been extended into several interesting directions including evolving Evoloops in [Sayama98a, Samaya99]. These cellular automata based ALife models offer the ideal example for our observer (observation process) based framework since these replicating loops and their variations evolve only with respect to some high level observation process, which can be used to define entities (loops) and their evolution. We will illustrate next the formal framework by instantiating it on the Cellular automata based Langton loop chemistry. Further details on the chemistry itself can be found in the above references.

8.2 Instantiating the Framework

We consider the case of two dimensional CA lattice based chemistry. Langton's self-replicating structure is a 86-cell loop constructed in two-dimensional, 8-state, 5-neighbor cellular space consisting of a string of core cells in state 1, surrounded by sheath cells in state 2.

We formalize observation process to be defined on a CA chemistry by assuming an underlying coordinate system such that each cell in a two dimensional cellular automata (CA) lattice can be associated with unique coordinates (represented as (x, y).) A cell is then completely represented as $< (x, y), s >$, where $s \in [0..7]$ is the state of the cell. When a cell is in state 0, it is also known as a *quiescent* cell. Let us denote the set of all cells of a CA chemistry as $Cell$, which is a potentially infinite set.

For a given cell $< (x, y), s > \in Cell$, we access its coordinates and state as follows: $co_x(< (x, y), s >) = x$, $co_y(< (x, y), s >) = y$, $co(< (x, y), s >) = (x, y)$, and $st(< (x, y), s >) = s$. We can extend these to sets of cells: $\forall X \subseteq Cell$, $co_x^+(X) = \bigcup_{c \in X} co_x(c)$, $co_y^+(X) = \bigcup_{c \in X} co_y(c)$, and $st^+(X)) = \bigcup_{c \in X} st(c)$.

$Neigh : Cell \rightarrow 2^{Cell}$ gives the coordinate wise non quiescent cells in the surrounding neighborhood of a cell. Formally,

$$
\begin{aligned}
&\forall (c =< (x, y), s >) \in Cell \quad we\ have \\
&Neigh(c) = \{ < (x+1, y), s' >, < (x-1, y), s' >, \\
&\qquad\qquad < (x, y+1), s' >, < (x, y-1), s' > \\
&\qquad\qquad | s' \neq 0\}
\end{aligned}
$$

The Chemistry Structure

A CA-based chemistry is usually initialized by setting some finite number of selected cells to non-quiescent states. At each step, state of every cell of the chemistry is changed as per the state transition rules. Therefore we define for an observer *state* of the Langton's chemistry as the set of all non quiescent cells, such that for the observer state changes only if there is any change in the set of non quiescent cells. The *state* of the chemistry for the observer will also be referred to as $configuration$. Thus Σ denotes the set of all possible different configurations and a state sequence in $\mathcal{T}$ is a sequence of configurations

observed in temporal order by the observer starting from some specific configuration. The observer focuses on observing only one state sequence at a time, therefore in the following discussion we will assume the underlying state sequence given as $T \in \mathcal{T}$.

Entities

With the above structure of Langton's CA chemistry, the observer takes the following decisions. Each entity in some state is characterized by two values - the connected set of non quiescent cells and the associated *pivot*. Two cells are connected only if there exists a consecutive sequence of neighboring non quiescent cells joining them in the lattice. The (function) *pivot* gives the coordinates for a cell uniquely associated with an entity in CA lattice in a particular state. Formally, the set of entities (loops) in the chemistry is defined as follows:

$$\begin{aligned} E = \{[X, pivot(X), i] \mid & \exists \textit{ a configuration } S \in T \ . \\ & [X \subseteq S \wedge X \neq \emptyset] \wedge [\forall c \in X \ . \ \exists c' \in Neigh(c) \ . \ c' \in X], \\ & i \in Tag\} \end{aligned}$$

where *Tag* is the set of tags uniquely associated with entities. To define the function *pivot*, the observer may choose the coordinates of top left hand corner cell of an entity as the pivot for it. Formally

$$\forall (e = [X, pivot(X), i]) \in E \ . \ pivot(X) = (min\{co_x^+(X)\}, max\{co_y^+(X)\})$$

This gives obvious characterization for a two dimensional nomological space $\Upsilon = Char_1 \times Char_2$ with $Char_1$ being the set of all non quiescent connected set of cells and $Char_2$ being the set of corresponding pivots. We do not include tags as characteristic of the entities.

$\forall e \in E, F(e)$ is the state in which an entity is observed.

Now we will proceed to define important relations on the chemistry which will be used by our observer to establish reproduction and fecundity.

Distance Measure

Distance function $D : E \times E \rightarrow \{0,1\} \times \{0,1\}$ is defined such that $\forall e_1, e_2 \in E \,.\, D(e_1, e_2) = [d_1(e_1[1], e_2[1]), d_2(e_1[2], e_2[2])]$, where auxiliary functions d_1 and d_2 are defined as follows:

$d_1 : Char_1 \times Char_1 \rightarrow \{0,1\}$ such that d_1 returns 0 only if X and Y have same number of cells arranged in same geometric correspondence or else it returns 1. Formally, $\forall X, Y \in Char_1 \,.\, d_1(X,Y) = 0$ *iff* there exists a *one to one correspondence* h between X and Y such that $\forall x \in X \,.\, \forall x' \in Neigh(x) \Leftrightarrow h(x') \in Neigh(h(x))$; $d_1(X,Y) = 1$ otherwise.

$d_2 : Char_2 \times Char_2 \rightarrow \{0,1\}$ such that

$$\forall\ (pivot_1 = (x_1, y_1), pivot_2 = (x_2, y_2)) \in Char_2$$
$$d_2(pivot_1, pivot_2) = 0 \ \textit{if}\ x_1 = x_2 \wedge y_1 = y_2$$
$$d_2(pivot_1, pivot_2) = 1 \ \textit{otherwise}$$

Informally this means d_2 returns 0 to indicate that pivots for both the entities are same and 1 otherwise. This will be used to determine recognition by our observer as defined below.

Limits on Observable Mutations

Next our observer selects $\delta_{mut} = [1, 0]$, which means that observer can recognize entity in future even with mutations provided that pivot remains the same, and $\delta_{rep\_mut} = [0, 1]$ which means with reproduction observer strictly demands same geometrical structure of the parent and child entities, though may have different pivots - this is essential to capture exact replication of the loops.

8.3 Observing Reproduction and Fecundity

Recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}} : E \rightarrow E$ is defined as follows:

$$\begin{aligned}\forall e, e' \in E, \mathbf{R}_{\delta_{\mathbf{mut}}}(e) = e' \Leftrightarrow & [F(e') = F(e) + 1] \wedge \\ & [D(e, e')[1] \leq \delta_{mut}[1]] \wedge \\ & [D(e, e')[2] \leq \delta_{mut}[2]]\end{aligned}$$

Informally this means two entities in consecutive states are recognized same only if they have the same pivots.

Lemma 1. $\mathbf{R}_{\delta_{\mathbf{mut}}}$ *satisfies Axiom* 1*, Axiom* 2*, and Axiom* 3*.*

Proof. Axiom 1 and Axiom 3 are satisfied by definition. Axiom 2, which states that $\mathbf{R}_{\delta_{\mathbf{mut}}}$ is an injective function holds because no two entities in the same state share the same pivot. This is because pivot as defined above is connected to all other cells of the entity and all the non quiescent cells which are connected in any state are taken together as one entity. Thus two different entities in the same state always consist of cells such that cells in one entity are not connected with the cells of second entity, and hence always have different pivots. □

The observer defines $\Delta \subseteq E \times E$ such that $\forall e, e' \in E \; . \; (e, e') \in \Delta \Leftrightarrow \forall Char_i \in \Upsilon \; . \; D(e, e')[i] \leq \delta_{rep\_mut}[i]$.

The causal relation C between entities in consecutive states is defined as follows: $C \subseteq E \times E$ such that $\forall [e = ([X, pivot(X)], i), e' = ([X', pivot(X')], j) \in E]$ we have

$$\begin{aligned}(e, e') \in C \Leftrightarrow & ([co_x^+(X) \supset co_x^+(X')] \wedge [co_y^+(X) \supset co_y^+(X')]) \wedge \\ & [F(e') = F(e) + 1] \wedge \\ & [pivot(X) \neq pivot(X')]\end{aligned}$$

Intuitively what we demand with above definition of causal relation C is that child entity breaks off from the parent entity at certain state, as can be seen in Figure.

Lemma 2. *Causal relation* C *defined above satisfies Axiom* 4*.*

Proof. First condition of Axiom 4 is satisfied by definition since $F(e') = F(e) + 1$. For the second condition $[\nexists e'' \in E \; . \; F(e'') = F(e) \wedge \mathbf{R}_{\delta_{\mathbf{mut}}}(e'') = e']$ notice that otherwise

because of the definition of $\mathbf{R}_{\delta_{\mathbf{mut}}}$, e'' and e' will have the same pivots, which means pivot of e'' will be included in the set of cells in e' (since $[co_x^+(X) \supset co_x^+(X')] \wedge [co_y^+(X) \supset co_y^+(X')]$), which is not possible because e and e'' being different entities in the same state cannot have cells in common as discussed above in the proof of previous lemma. □

Relations **AncestorOf** and **Parent** are defined by our observer same as in the framework.

Lemma 3. ***Axiom of Reproduction*** *and the* ***Axiom of Fecundity*** *can be established by our observer in a given Langton Loop chemistry.*

Proof. These two axioms can be witnessed by our observer in a given state sequence, which is seen by the providing only one such example as depicted in Figure 8.1 and Figure 8.2. □

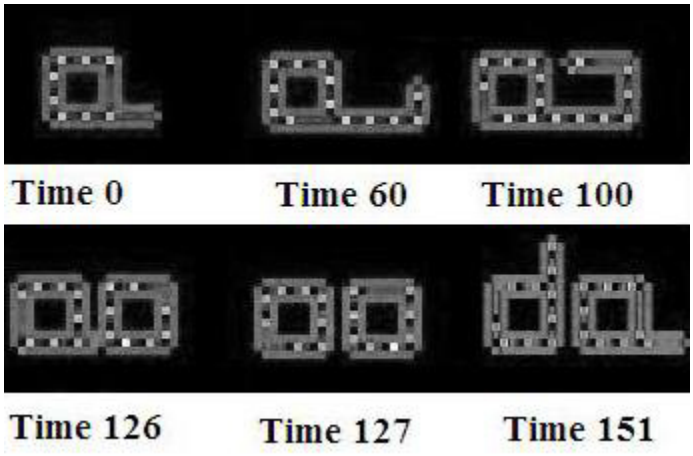

Figure 8.1: Self-Reproduction in Langton loops; screen shots from [Sayama05]

8.4 Discussion on Mutations, Inheritance, and Natural Selection

Primary focus of Langton while defining the CA based replicating loop chemistry was to demonstrate that genotype - phenotype based coding decoding scheme can be captured in CA universe as well [Langton87]. And we have seen that this can be observed by the observer as defined above. Nonetheless, Langton loops, as designed as yet do not exhibit something which can be treated as reproductive and inheritable mutations and indeed

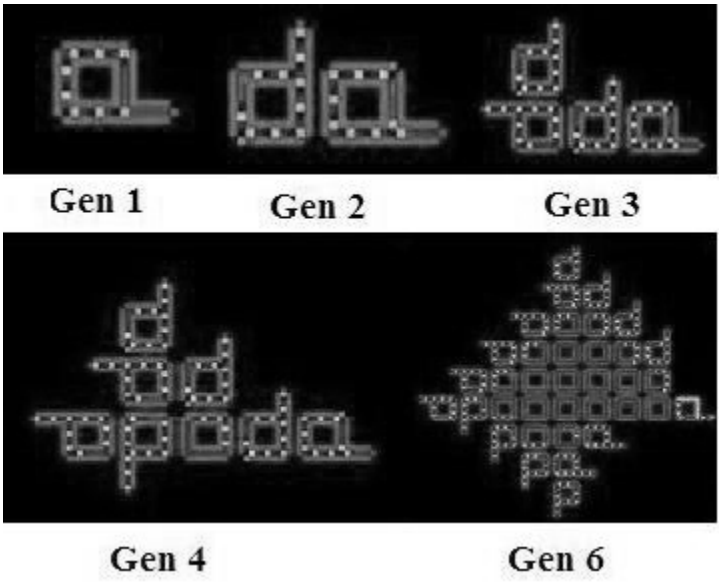

Figure 8.2: Fecundity across generation in a population of Self Replicating Langton Loops; screen shots from [Sayama05]

if we analyze the underlying state transitions defined for the cells in the chemistry, it becomes clear that the transition behavior required for the reproduction changes immediately if any changes are introduced in an entity and resulting entity is no longer capable of reproduction or in other terms, none of the mutations in existing replicating loops preserve reproduction and therefore *Axiom* of *Continuity of Reproduction under Mutations* is not preserved. Because of the enormity of possible configurations and transition dynamics it is not easy to analyze which kind of replicating loops can ever withstand certain mutations and can preserve replicating functionality. Heredity of course is worth considering only when entities mutate and continue reproduction. Thus with existing Langton loops, an observer cannot observe heredity and subsequent natural selection.

Evoloops The major difficulty with CA based chemistries is to define extended state transitions such that at any state of the chemistry cell transitions to next states are not based upon some fixed transition table and constant neighborhood but with more generalized state transitions with multiple possible outputs determined by more generalized neighborhood as well as other conditions as observed for example in natural chemical environment. The extension of Langton loops defined by Sayama as *Evoloops* in [Samaya99] is one such attempt in this direction, where not all the loops in the chemistry are of same type with respect to the number and geometrical arrangement of cells and

final population witnesses variety of different kinds of reproducing loops scattered on the lattice forming colonies. The Evoloops and their evolution can be formulated in our framework by changing the definition of δ_{rep_mut} such that the observer should be able to establish parent child relationship even when the parent and child loops are not the same. Since the evoloops of different types replicate at different rates, where rate of replication is measured in terms of number of state transitions, we can infer that the loops satisfy the axiom of sorting. Indeed in a weak sense with available simulation results it appears that evoloops can be observed demonstrating strong heredity as well as correlation.

We will next present the following chapter, the example of Algorithmic Chemistry, which is of a different nature compared to Langton's loop chemistry.

8.5 Conclusions

This way we have seen that we can define precise observation process on the CA universe which discovers the self replication of so called Langton loops during the simulation of model. The specific observer presented here follows the intuition that Langton implicitly stated when describing the loops. We also noted that mutations, heredity, and selection based axioms are not met in the chemistry where this limitation can be attributed to underlying transition rules of the chemistry. Evoloops, which were designed as extensions of Langton loops with mutations can be seen to be indeed evolving with natural selection.

Chapter 9

Case Study 2: The Algorithmic Chemistry

Lambda calculus based Algorithmic Chemistry was introduced in [Fontana92] and further discussed in [FL94]. The main focus of their Algorithmic Chemistry is to study the principles behind emergence of biological organizations with the approximate abstraction of real chemistry as lambda calculus with finite reductions. In their study they describe the emergence of three different kinds of organizations, starting with random population of λ terms (molecules), with different filtering conditions or constraints imposed upon reactions. The three basic organizations observed are: *Level 0* organization consisting of set of self copying lambda terms or (cooperating) hypercycles with mutually copying lambda terms, *Level 1 self maintaining* organizations consisting of lambda terms such that every term is effectively produced as a result of reaction between some other terms in the same organization and lastly *Level 2* organization where molecules migrate between two or more self maintaining Level 1 sub-organizations. They also provide detailed algebraic characterization of self maintaining and Level 2 organizations without referring to the underlying micro structure of lambda terms (molecules) or the micro dynamics (reduction semantics and filtering conditions) governing result of a reaction.

9.1 Instantiating the Framework

In view of our observer based framework, characterization of self replicating molecules or the hypercycles of mutually copying molecules is quite natural. Because we demand

in our framework that one particular observer focusses on only one level of molecular abstraction, therefore we can define, as discussed next, an observer focussing on hypercycles as a set of individually replicating lambda terms.

However since Level 1 and Level 2 organizations as described to be emerging in case of Algorithmic Chemistry [Fontana92], emerge only when self copying reactions are filtered out, these organizations cannot be analyzed under our current framework because of the primary focus of our framework on reproduction, mutation, inheritance, and selection based evolution and emergence of organizations.

The Chemistry Structure

Consider the chemical soup of Algorithmic Chemistry consisting of lambda terms. The soup is usually initialized with a population of large number of randomly generated lambda terms. Thus *state* of the chemistry can be taken as the collection of all these lambda terms (with multiplicity). Since every non elastic reaction results into introduction of output lambda term into the soup and possible removal of some other randomly chosen terms, therefore it is natural to consider such succession of states after every reaction step as a state sequence $T \in \mathcal{T}$. Notice that the observer only observes the colliding inputs terms, resultant output terms to be added and the randomly deleted terms from the soup, not the actual reaction details between colliding lambda terms.

Entities

We consider an observer who considers each lambda term as one separate entity and associates an unique integer tag with it. We can define E as set of all such entities in the chemistry. Each such entity is represented as $[w, i]$ where i is the tag uniquely associated with λ-term w.

Thus each entity is characterized by only two values - the lambda term and the tag associated with it. This implies a two dimensional nomological space $\Upsilon = Char_1 \times Char_2$ with $Char_1$ being the set of all possible syntactically valid lambda terms and $Char_2$ is the set of tags.

$\forall e \in E, F(e)$ is the state in which an entity is observed.

Tagging: Before we proceed to define important relations on the chemistry which will

be used by our observer to establish self reproduction and fecundity in the chemistry, let us elaborate more on tagging mechanism, which will be used by the observer to recognize whether two lambda terms in successive states (with different tags) are the same. The tagging assumes very important role in the chemistry since there can be multiple copies of syntactically same lambda term in the soup at any state of the chemistry. Our observer for the chemistry associates tags of the form $< i_{size}, i_{lex}, i_{mul} >$ $(i_{size}, i_{lex}, i_{mul} \in N)$ for the molecules in the following way: for the initial population with $n > 0$ terms in the soup it arranges the terms with respect to their sizes and assigns the size of these terms as first component in their tags (i_{size}) and for terms with the same size it arranges them lexicographically and assigns in increasing order second component of the tags (i_{lex}) such that multiple terms with exactly same syntactical form are put together and get same first two components of the tag and then assigns increasing integers to each of the terms in multiplicity as their third component of the tag (i_{mul}). This can be seen the same kind of numbering as usually used in arithmetization of formal theories. For a given tag $tg =< i, j, k >$ we access its components as $i = tg[1], j = tg[2]$, and $k = tg[3]$. Under such tagging scheme a small population of *lambda* terms $\{\lambda x_1.\lambda x_2\ .\ x_2, \lambda x.x, \lambda x.x\}$ can be represented as $\{[\lambda x.x, < 3, 1, 1 >], [\lambda x.x, < 3, 1, 2 >], [\lambda x_1.\lambda x_2.x_2, < 5, 1, 1 >]\}$.

Next comes the question of *updating* these tags after every reaction step and elimination step, which is done as follows: all what is needed is that the observer uses such tag updating mechanism f between two successive states, for which given tag t_g for an entity in a particular state, its tag t'_g in the previous state can be derived from t_g in order to know whether that term is the same present in the previous state? This means if tag for an entity e is updated from t_g to t'_g between two successive states under f, then the observer can deduce t_g from t'_g uniquely. One such simple mechanism can be defined based upon the observer's knowledge of new terms added to the soup in the next state which are the result of some reaction between entities in previous state. So what an observer can do is that it adds 1 to the third component of the tags for every older term from previous state and gives new unique tag to new term with respect to its position in the list of all the terms based upon its size, lexicographic order, such that third component of newly added term is always given value 1. This numbering scheme will robustly maintain recognition of terms across states of the chemistry because only

the new term added in the previous state can have its 3^{rd} component as 1.

Distance Measure

Distance function $D : E \times E \rightarrow \{0,1\} \times \{0,1\}$ is defined such that $\forall e_1, e_2 \in E \,.\, D(e_1, e_2)[1] = 0$ if $e_1[1]$ and $e_2[1]$ are the same with respect to α renaming; otherwise $D(e_1, e_2)[1] = 1$. $\forall e_1, e_2 \in E \,.\, D(e_1, e_2)[2] = 0$ if $|e_1[2][3] - e_2[2][3]| = 1$ otherwise $D(e_1, e_2)[2] = 1$. Notice that we have defined our distance function D keeping in mind the use of the distances in recognition later on.

The Limits on Observable Mutations

Next our observer defines $\delta_{mut} = [0, 0]$ and $\delta_{rep\_mut} = [0, 1]$, which means the observer treats two semantically different lambda terms as different entities. Notice that with this view our defined observer cannot recognize entities which are even very slightly different in their syntactical representations. This is since two entities even with very slight differences in syntactical representations might have quite different reaction semantics.

9.2 Observing Self Replicating Hypercycles

We will now illustrate how our observer can observe the self-replicating elementary hypercycles as sets of self-replicating entities in Level 0 organizations. We define the recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}} : E \rightarrow E$ as follows: $\forall e, e' \in E$, $\mathbf{R}_{\delta_{\mathbf{mut}}}(e) = e' \Leftrightarrow [F(e') = F(e) + 1] \wedge D(e, e') \leq \delta_{mut}$. Informally this means two entities in consecutive states are recognized same only using their tags.

Lemma 4. $\mathbf{R}_{\delta_{\mathbf{mut}}}$ *satisfies Axiom* 1*, Axiom* 2*, and Axiom* 3*.*

Proof. Axiom 1 and Axiom 3 are satisfied by definition. Axiom 2, which states that $\mathbf{R}_{\delta_{\mathbf{mut}}}$ is an injective function holds because of specific construct of tagging mechanism and the definition of Distance function D which is such that two entities in successive states are recognized as same only when the difference between their third components of tags is 1, and we know that the observer selects new tags in such a way that this difference is 1 only when same entity was present in the previous state. □

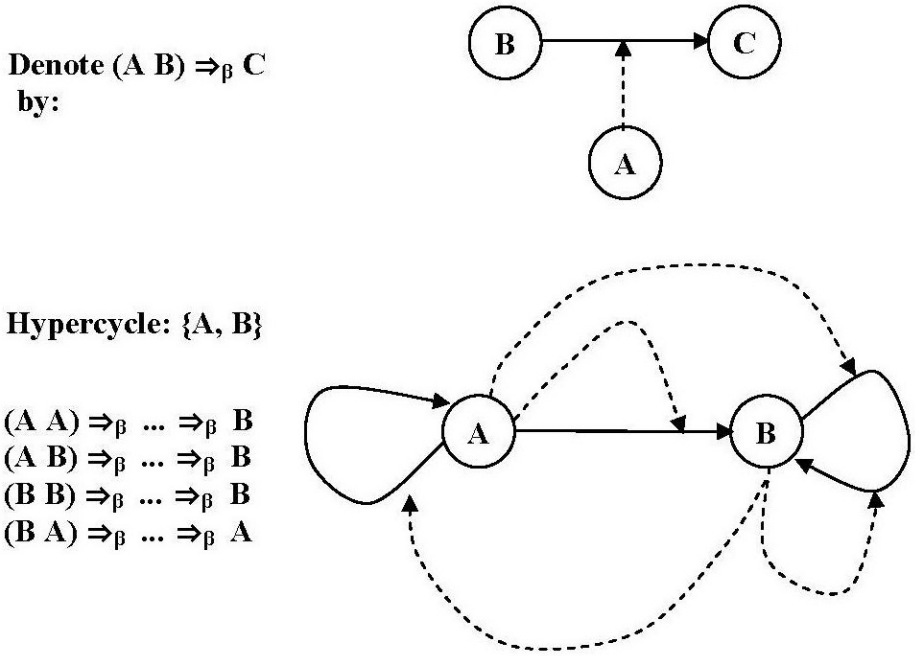

Figure 9.1: Example of self replicating elementary hypercycle organization in AlChemy from [FL94]. $A = \lambda x_1.\lambda x_2.x_2$ and $B = \lambda x.x$. $(AB) \Rightarrow_\beta C$ represents reaction between A and B by applying A on B yielding C under β reduction.

The observer defines $\Delta \subseteq E \times E$ such that $\forall e, e' \in E.(e, e') \in \Delta \Leftrightarrow D(e, e') \leq \delta_{rep\_mut}$. In order to define causal relation between entities in the Algorithmic Chemistry, the observer observes the reacting entities at each reaction step which yield output to be added in the soup. If entities e and e' react in some state and yield e'' then the observer defines causal relation such that $(e, e'') \in C$ and $(e', e'') \in C$ with $F(e) = F(e') = F(e'') - 1$.

Lemma 5. *Causal relation C defined above satisfies Axiom 4.*

Proof. First condition of Axiom 4 is satisfied by definition since $F(e'') = F(e) + 1 = F(e') + 1$. The second condition $[\nexists e_1 \in E.F(e_1) = F(e) \wedge \mathbf{R}_{\delta_{\mathbf{mut}}}(e_1) = e'']$ again follows from the specific construct of tagging as well as the distance function because as per the tagging mechanism explained before e'' being newly added entity in the chemistry will have the 3^{rd} component of its tag as 1 and all previously present entities, including e_1, in the chemistry have their tags in new states updated such that their 3^{rd} components are always greater than 1. □

Relations **AncestorOf** and **Parent** are defined by our observer same as in the framework.

Lemma 6. ***Axiom of Reproduction*** *and the* ***Axiom of Fecundity*** *can be established by our observer in a given Algorithmic Chemistry.*

Proof. These two axioms can be witnessed by our observer in a given state sequence, which is seen by the providing examples of self copying lambda terms or elementary hypercycles. Note that in case of hypercycles, the observer establishes multi step reproduction using transitive closure of causal relation for each of the entities in the hypercycle. Examples of self replicating entities include $\lambda x.x$ and $\lambda x_1.(x_1.\lambda x_2.(x_2 x_1))$, with the usual notation for lambda terms from [Fontana92]. $\lambda x.x$ and $\lambda x_1.\lambda x_2.x_2$ are examples of copiers and self copiers. Example of hypercyclic organization consisting of two entities is $\{\lambda x_1.\lambda x_2.x_2, \lambda x.x\}$ as illustrated in Figure 3. As per the definition of causal relation, entity instances of $\lambda x_1.\lambda x_2.x_2$ and of $\lambda x.x$ are causally related to past instances of each other and of themselves. □

9.3 Discussion on Mutations, Inheritance, and Natural Selection

As emphasized in [FL94], the primary goal of Algorithmic Chemistry is to study alternative pathways in which higher level biological organizations (i.e., hypercycles, self maintaining organizations) can emerge in the chemistry. And therefore there is no explicit notion of mutation present in Algorithmic Chemistry. To see this notice that every new entity in the population is the result of reaction between two other entities. Therefore if one particular observer decides that one of the reacting entities is mutating into the resulting entity, it is still difficult to decide which of the two reacting entities should be considered as mutating into the new one. Even if such a view is adopted, the observer will observe that if a self-copying entity at any reaction step mutates into another entity then most often the new entity can no longer self replicate. Thus *Axiom* 7 (Preservation of Reproduction under Mutation) is most often violated. Finally as discussed at the beginning of the chapter owing to the focus of our framework on the evolutionary processes, self-maintaining organization of the kind that arise in Algorithmic Chemistry are beyond the scope of discussion.

9.4 Conclusions

Thus we have demonstrated that a precise observation process can be defined to work with Algorithmic Chemistry, which can be used to discover the self replicating simple lambda terms as well as simple hypercycles in the model. We also noted that mutations, heredity, and selection based axioms are not met in the chemistry where this limitation should be attributed to underlying reaction semantics of the chemistry as well as its design. We also discussed that probably the major aim of Algorithmic Chemistry is to define alternative pathways for the emergence of organizational structures (Level 1 and 2 organizations) and their abstract characterization, which though does not fit under the current scope of our framework. the results can be interpreted as clarifying the limit of the approach. The notion of evolution used in this thesis is very concrete and requires elements that are not present in Algorithmic Chemistry. This study highlights the fact that not all interesting dynamic processes are evolutionary in nature and therefore some of these non evolutionary processes are out of scope of the framework developed in this thesis.

Chapter 10

Case Study 3: The Reduced Instruction Set Artificial Chemistry

The Reduced Instruction Set Artificial Chemistry (RISAC) was designed and studied by J. Vedvyas, School of Computing, National University of Singapore as his honor's year project under the supervision of Dr. Martin Henz [Vedvyas05]. RISAC is an assembler automaton kind virtual machine model, similar to Tierra [Tierra-Webpage] and Avida [Avida-Webpage]. Their goal in the project was to find out the bare minimum instruction set required to demonstrate the behavior shown by Tierra and Avida. Reduction of instruction set implied smaller self replicating structures and so higher probability of the random emergence of such structures.

10.1 Design of the Chemistry

RISAC can be briefly described as follows: the virtual machine (VM) consists of an *alphabet of size* n of letters used in the individual strings (programs), total m memory registers holding instructions, a special *boolean register* b, the *register mapping* $r : [1, 2, \ldots, n] \Rightarrow$ [1,2,...,m] which maps each letter in the strings to registers and the *instruction mapping* $p : [1, 2, \ldots, n] \Rightarrow [set, test, write, jump]$ which maps each letter in the strings to an instruction. The environment containing the set of strings called

the *soup* (S). A transition from one state to another removes m strings from S and transforms them using the RISAC machine instruction semantics and finally puts them back to S resulting in a new state of the soup S'.

Each register in the VM contains a pair $r_i = (S_{r_i}, a_{r_i})$ where S_{r_i} points to a string and a_{r_i} to the position within that string in the soup. The content of the register refers to the location of a particular character in the soup. For example $(S_{r_4}, a_{r_4}) = (\text{“3210”}, 2)$ means that register 4 points to the string “3210” and 3^{rd} letter in that. The value of the Boolean register is either *true* or *false*, which is accessed using the letter “*b*”. Register r_1(P) is the initial instruction register.

Each instruction consists of 3 symbols which represent the instruction *Opcode* and two argument registers (*operands*) respectively. The machine design uses a flexible memory addressing mechanism using indirect addressing mode. The model also allows mutations of various kinds, which change the instruction strings like point mutations, insertions and deletions.

There are six different types of instructions in the chemistry: *Write* $[Wr_1r_2]$ instruction writes the character in the first register r_1 to the second register r_2. *Read* (Rr_1r_2) instruction sets the boolean register to the equality of the letters in first and second register. *Swap* (Ar_1r_2) instruction moves the location of the second register to the location of the first register. *Jump* (Pr_1r_2) instruction takes the instruction in the first register as pattern and searches for it in the second register. The search excludes the pattern from the search if both the registers are the same. The Jump instruction is excluded from the search if second register is the instruction register. *Beginning* (Mr_1r_2) instruction moves the second register to the beginning of the first register. *End* (Qr_1r_2) instruction moves the second register to the end of the first register.

The experiments were carried out with different parameter settings, usually involving at least one copier program present from the beginning. The experiments were carried out with or without spacial topological setup which had resulted in the way effects of various instructions are effective.

We will now proceed to instantiate our framework upon this RISAC machine with various experimental simulations defining various runs of the chemistry.

10.2 Instantiating the Framework

As mentioned above each state of the chemistry is defined by the soup S which in turn can be precisely defined as a set of registers in memory containing instructions. State transitions are the result of executions of the instructions, one at each step, which change the contents of the registers. Thus each sequence of states can be precisely defined as the chemistry is allowed to progress from some start state with simple instructions present in the memory.

Different Levels of Observation

Before we proceed to define the entity set observed by an observer, we should highlight the importance of the level at which these observations can be made. In case of RISAC, this aspect is of crucial importance because for the experiments involving topological constraints on memory an observer can observe the chemistry on two different levels - one the micro level where he has the access to the basic instruction registers in the chemistry and secondly when he works upon the high level graphical interface provided to him by the designer of the chemistry. And we will notice that observation made on these two different levels might bring out different amount of information and consequent inferences.

For these experiments with topology the interface for the observations on the chemistry is designed as two dimensional grid of cells, each cell representing one string (program) and color (state) of the cell distinguishes it from other programs occupying other cells.

Next I will informally define the various components of the framework with respect to RISAC experiments.

Entities

Entities are defined based upon the level of observation. For the non topological soup experiments there is only one level to define entities, where each entity can be regarded as an individual program occupying one or more registers in the memory. These programs can be given unique tags which can be used for defining function F, which will return the state of the entities based upon the tags given to them.

For experiments with topology, an observer can choose to define entities as individual cells or might choose to select spatially connected collections of cells (like Langton Loops) with same color as entities. Tags can be based upon the state of the chemistry as well as topological coordinates of these cells in both the cases.

For entities without topology, the nomological space Υ of characteristics of the entities can be defined as single dimensional space with all possible programs (in some of the experiments bounded by the maximum size) as values taken on that character. Thus each entity is completely characterized by the string of symbols defining it as a program.

For experiments with topology an observer might define color of the cell as its characteristics and in case of collection of cells as entities color of the cells as well as topological set up consisting of the exact subset of cells in the collection can be considered as a suitable definition of Υ.

More formally in case of experiments without topological interface, $\Upsilon = Char_1$, where $Char_1 = \{P : P$ is a syntactically valid program in of RISAC.$\}$. With topology, $\Upsilon = Char_1 = \{c : c$ is the color of a cell in the 2D grid. $\}$ for single cells as entities and $\Upsilon = Char_1 \times Char_2$, where $Char_1$ is the set of colors of cells in the grid, and $Char_2$ is the set of subsets of cells with same color and spatially connected in the grid.

The Distance Measure

Distance measure D can be defined without much difficulty for all three types of entities. For entities without topology, D should return the number of symbols where both entities (programs) differ. With topology - in case of single cells as entities, D can return 0 to indicate that both the cells have same color state and 1 otherwise, while with collection of entities D will return 0 when both entities contain same number of cells with same topological set up in the cells and when color of cells in both entities are also the same otherwise it will return the number of cells which have changes in the color.

Observable Limits on Mutational Changes

In case of experiments without topology, two entities are recognized under $\mathbf{R}_{\delta_{\mathbf{mut}}}$ as same in successive states using mutational limit δ_{mut} such that either of single symbol

based mutation or insertion or deletion based mutations are executed on first entity in previous state to get new entity in next state. Thus $\delta_{mut} = 1$, which denotes that two program entities can differ by only single symbol. We can similarly define the observable limits on the reproductive mutations. For non topological entities only single symbol mutations are allowed and therefore $\delta_{mut} = 1$.

For topological set up, situation is much difficult because of the interface provided by the chemistry. There an observer has no alternative but to recognize ($\mathbf{R}_{\delta_{\mathbf{mut}}}$) two entities in the successive states if they reside on the same coordinates in the 2 D grid with every possible change in the color. Thus for single cell entities $\delta_{mut} = 1$. For multi -cell entities an observer might not recognize cells with changed colors in successive states but might well recognize them under changes in the sets of cells under certain limit. Thus $\delta_{rep-mut} = [0, K]$, where K is the maximum number of cells where the two multi-cell entities may differ.

Evolutionary Relationships

Causal relationship between two entities in case of non topology based experiments can be defined when execution of one program affects the state of the other program. This can happen precisely only when first program executes the *Write* instruction. There if entity e executes (W) write instructions at any state of the chemistry then each of the entities which are written upon because of these executions will be causally connected to e, formally, $(e, e') \in C$ if and only if e executes a write instruction which contains an operand register pointing to e'.

In case of topological interface, there is no precise way to define the causal relation between cells because it is not known which cell was active and executed write instruction at any state of the chemistry. Same is true for the case of multi-cell entities. Therefore the observer might randomly choose to define any of the neighborhood cells as possible causally connected entity. In case of multi-cell entities, same can be generalized when the observer defines some fixed neighborhood radius between two multi-cell entities and then randomly selects one of the entities falling under the neighborhood radius as causally connected to the newly appeared entity.

Lemma 7. *Axioms* $1, 2, 3$ *for recognition relation* $\mathbf{R}_{\delta_{\mathbf{mut}}}$ *and the Axiom* 4 *for causal*

relation C can be established by our observer in a given Algorithmic Chemistry.

Proof. The axioms hold because of the way our observer has defined recognition as discussed above, though informally. In case of non topological set up causal relationship exists only when first entity executes write instruction to write upon the registers of second program. Therefore the written upon entity can not be recognized as mutant of any previous entity since only single symbol mutations are admissible. It must however be noted that in case of random multi-bit mutations, recognition and causal relations might be difficult to distinguish for an external observer unless he observes precise instructions which cause changes in the register contents. Similar is the case with topological experiments. □

10.3 Self Reproduction and Mutation in the Chemistry

For non topological set up - the observer defined above can observe that there are entities (programs) which point to their beginning and then write themselves to some other place in the memory. Therefore observer observes self replicators as self copiers. Since self copying might take several executions of several write instructions, therefore observer, in case of say such entity $e \in E$ can correctly associate new entities in intermediate stages as causally connected to e and finally at one state new entity $e' \in E$ will appear which will be exactly same as e and so e, e' can be associated with *Parent* relationship. In case of single symbol mutations as well, the observer can define the relationship correctly.

With experiments having topological set up - the plausible causal relationship defined by the observer which only restricts spatial neighborhood between entities can be used to define *Ancestor* as well as *Parent* entities.

Lemma 8. ***Axiom of Reproduction*** *and the* ***Axiom of Fecundity*** *can be established by our observer in given RISAC chemistry.*

Proof. These axioms go by demonstrating the examples of reproducing programs (entities). There are indeed two kinds of self reproducing entities in the chemistry - one which have copying instructions built into them, for example, [APR WRW MPP], while the second kind of self reproducers use copying instruction of other programs, so can be termed as symbiotic self reproducers. Example of such symbiotic self reproducer are

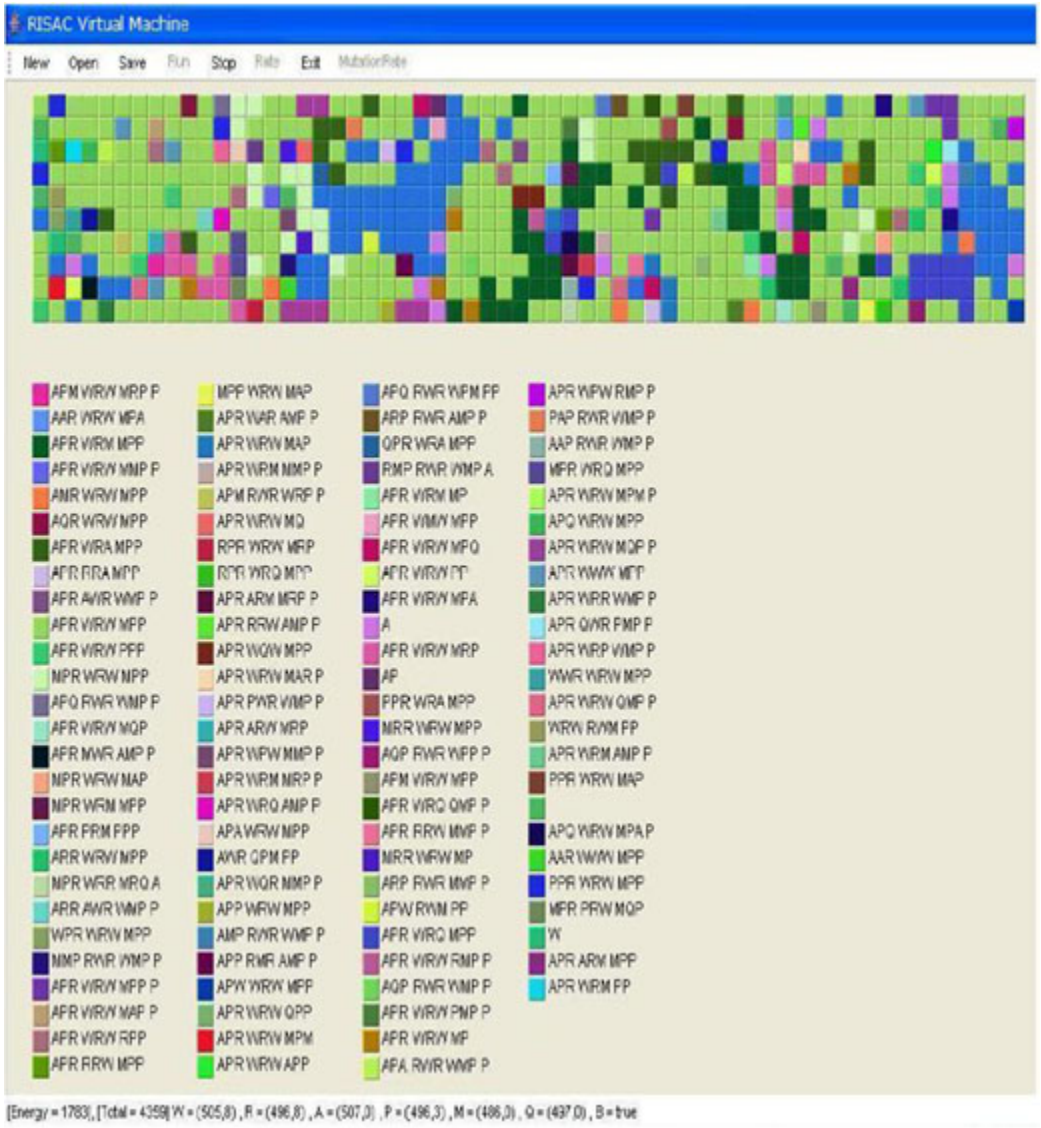

Figure 10.1: Simulation of Self Replicators subject to mutation. Each color in grid represents a particular type of program. A list of different program are shown below the grid.

given as [APR WRW MPP P].

For topological set up, again since population of self reproducers increases over time, our observer even with imprecise plausible associations of descendent relationships can infer self reproduction and fecundity.

Figure 10.1 is the screen shot of one such simulation with population of reproducers of various types under mutation.

□

Lemma 9. ***Axiom of Continuity of Reproduction under Mutation*** *can be established by our observer in given RISAC chemistry.*

Proof. Again this is established by observing explicit examples where mutations in self reproducers give rise to symbiotic self reproducers. The example is, [APR WRW MPP], which after insertion mutation gives rise to [APR WRW MPP P], which is a symbiotic self reproducer. The other kind of point mutations which change single bit in the registers, indeed give rise to whole set of self replicators for [APR WRW MPP], [APR WRA MPP], [APR WRQ MPP], [APR WRM MPP], [MPR WRM MPP], [MPR WRW MPP]. It must however be pointed out that further mutation on these mutants does not retain self reproducing capability. This is partly attributed to the design of the chemistry which has fixed semantics for the instruction, therefore not all instruction can keep copying machinery.

Some plausible way to demonstrate the same can be even found in case of topological set up. □

10.4 Natural Selection

Natural selection is the weakest aspect of the chemistry due to three factors: One, observations could not be made on the statistically large number of generations because most of the experiments terminate quickly. Secondly the variation under mutation could not be observed because as discussed before, only very limited number of mutants retain the self reproducing ability and thus expected variation is absent. Thirdly among those mutants that do reproduce, the reproduction rate tends to be exactly identical, giving rise to no selective reproductive variation.

10.5 Conclusions

Thus we have demonstrated that a precise observation process can be defined to work with new reduced instruction set based artificial chemistry, which can be used to discover the self replicating programs with limited reproductive mutations using two different kinds of interfaces for observation on the chemistry. These different interfaces demonstrate important informational aspect of observation on the chemistries and also demonstrate idea of multi level observations on the same model universe. We also noted that aspect of selection is not strong in the chemistry where this limitation can be attributed to underlying reaction semantics as well as the design of the programs, which do not allow large number of mutants of self reproducing programs to retain reproduction capability.

Chapter 11

Case Study 4: Artificial Graph Chemistry

The Artificial Graph Chemistry (AGC) was designed and studied in collaboration with Bee Peng, School of Computing, National University of Singapore as a part of his honor's year project under the supervision of Dr. Martin Henz [Peng05].

The artificial graph chemistry is designed with the motivation of capturing the basic "feel" of real chemistry. AGC captures the feel of real chemistry by applying abstractions on several fundamental levels. The abstractions are made on the level of basic molecular structure by the defining valid structural components of the chemistry as connected labeled graphs, by defining structure-guided reaction semantics as well as global environmental constraints such as preservation laws and global parameter like temperature.

11.1 Design of the Chemistry

The overall design of AGC can be described in the style of Dittrich [DZB01]- the syntactic structure of entities, reaction rules, and the reaction algorithm.

Syntactic Structure: AGC defines connected graphs as molecules. Every graph consists of finite number of nodes and edges between them. Each node is labeled by certain "node-type" identifier - type of a node restricts the number of edges it can be connected with as well as determines the relative affinity between nodes of different types. The relative affinities between different nodes types are represented as positive

integers making the edge labels. The edge labels are also referred to as *edge weights*.

The maximum number of possible edge connections for a particular node type is called its *saturation index*. In any graph, the number of edges a node is connected to is called its *saturation level*, and the difference between saturation index and saturation level is termed as *unsaturation level* of that node.

Reaction Semantics Chemistry defines the structure guided and globally constrained reaction semantics for reacting molecules. There are two basic principles governing reactions between molecules. *The principle of node preservation*- in every reaction, total number of nodes before and after the reaction remains the same. This means a reaction neither adds any new node not present in input graphs nor does eliminate any nodes. This implies that reactions are restricted to edge manipulations. *The principle of maximum stability* - reactions favor stable graphs. Node properties such as saturation level and edge weights affect the stability of graphs. Nodes of higher degree are more stable than nodes of lower degree in the same graph. Edges of greater weights are less likely to break. As stable graphs are favored, it increases the likelihood of complicated structure from arising and possible emerging behaviors such as self-replication as a result of the new structures. Apart from these reaction guiding principles, chemistry also defines global parameter like temperature which affect the way new molecules are generated after reactions take place.

Reaction Algorithm After setting up basic parameters like number of "node-types", saturation indices, and edge weights actual reaction proceeds to work upon the initial set of graphs in two phases at every step - the assembling phase and the disassembling phase described as follows.

In each reaction, algorithm selects randomly two input graphs. It defines global parameter for the reaction as temperature, which in turn determines the relative stability of nodes. In the assembly phase an intermediate graph is generated by rearranging the edges between reacting graphs to maximize stability. During disassembly phase, based upon value of the temperature, algorithm selects some random connection threshold level η, which determines which of the edges will remain intact in the graph generated from assembly phase. At high temperature value of η is low, while at low temperatures it will be high. Reaction removes all the node to node connections which fall below η.

At the end of this disassembling phase there might be several graphs with relatively high stability.

The new population is generated by removing input graphs from the population and placing the new set of graphs into the population.

Having described the design of the chemistry in brief, I will proceed to describe how our observer's framework can be effectively applied to it.

11.2 Instantiating The Framework

The *states* of chemistry can be defined as multisets representing population of graphs at any instance. As reaction algorithm is applied, chemistry progresses in time and gives rise to new multisets of graphs. Thus temporal progression of AGC can be defined as an ordered sequence of these multisets.

Entities

Individual graphs can be considered as observed entities, which can be given unique tags in each state. More complex multi-graph entities consisting of subsets of graphs can also be defined but with current experiments these multi-graph entities does not seem to have any evolutionary significance.

Because the observer is observing individual connected graphs as entities, the only relevant defining characteristics for these entities seems to be their graph definition, that is set of nodes and corresponding edges. This also gives simple characterization for distance measure D between entities as the number of nodes and number of edges where two entities differ, that is, for graphs $G_1 = (E_1, V_1), G_2 = (E_2, V_2) \in E$, $D(G_1, G_2) = [|V_1 \oplus V_2|, |E_1 \oplus E_2|]$, where $\oplus$ represents the symmetric difference between sets.

The mutational limits δ_{mut} and δ_{rep_mut} can be defined as to how much difference in nodes and edges between two graphs an observer can ignore while considering them similar. Thus the observer can define $\delta_{mut} = [n_v, n_e]$ to indicate that the observer will consider two graphs similar even if they have different nodes (size of the difference bounded by n_v) and different edges (size of the difference bounded by n_e).

Evolutionary Relationships

As per the design of the AGC, the entities which take part in a reaction at any state are removed from the chemistry and all the entities which otherwise do not take part in the reaction at any state remain intact. Therefore the observer can define recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$ based upon the reactions getting executed at any state of the chemistry. Thus to define recognition, the observer need to only look at the input participating entities and output entities. If any of the input entity graph is similar to any of the output graphs (difference is bounded by δ_{mut}), then output graph in next state can be recognized as input graph present in the previous state. It should be clear that this recognition relation satisfies all the axioms $(1, 2, 3)$ of the framework. Similarly if input and output entities for any reaction are not similar, that means, input entities cannot be recognized in next state (among output entities of the reaction), then they are related with causal relation C. It is clear that the causal relation satisfies the axiom 4 because causally connected entities are always present in consecutive states and output entities of the reaction were neither present in previous states nor do any of the entities in previous state can be recognized with them.

Reproduction and Selection

The underlying design of AGC with its generic reaction laws is such that there will always be some path consisting of one or more reaction steps in successive states of the chemistry which will yield reproduction of initially participating graph entities using recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$, causal relation C, and the limit on reproductive mutations bounded by δ_{rep_mut}. Though in actual experiments, all reaction sequences including the ones which yield the reproduction of certain entities are probable with certain probability and not always this probability attains high values. One such example appears in the Figure 11.1 (with statistical data in Figure 11.2), which though was not observed in explicit experiments but is possible to emerge if simulations run favoring the reaction sequence. Therefore major difficulty in exhibiting such reproduction for any arbitrary graph comes due to the stochastic nature of the reaction rules which allow for certain input graphs multiple probable sets of output graphs and at any state of the chemistry one of these possibilities is selected by the reaction algorithm in a probabilistic random

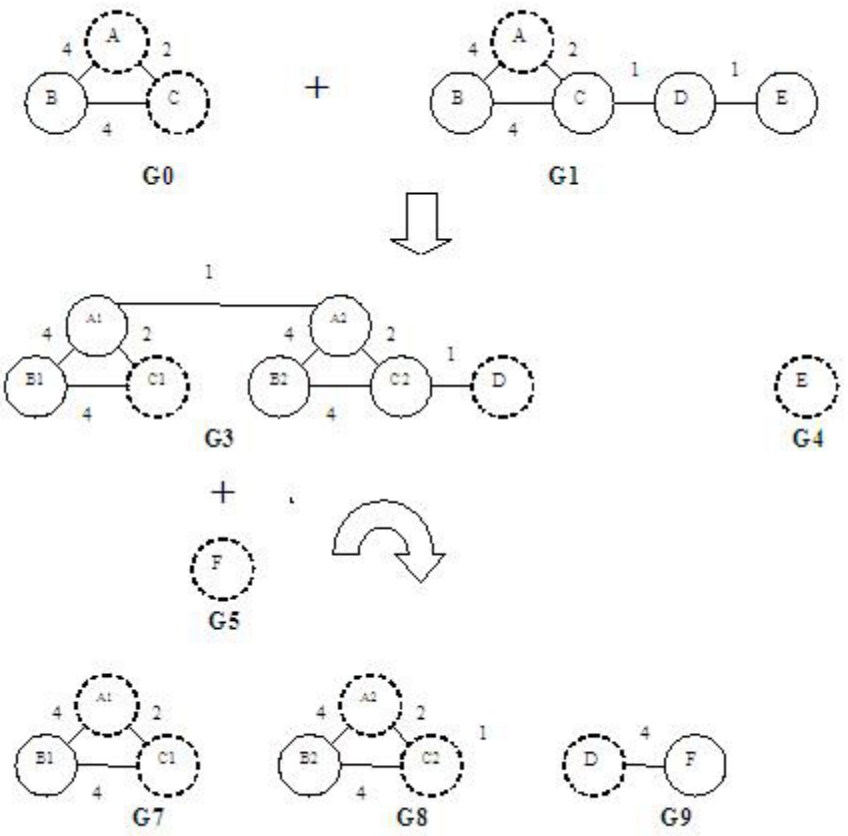

Figure 11.1: Self Replication of Graph $G0$ in two steps.

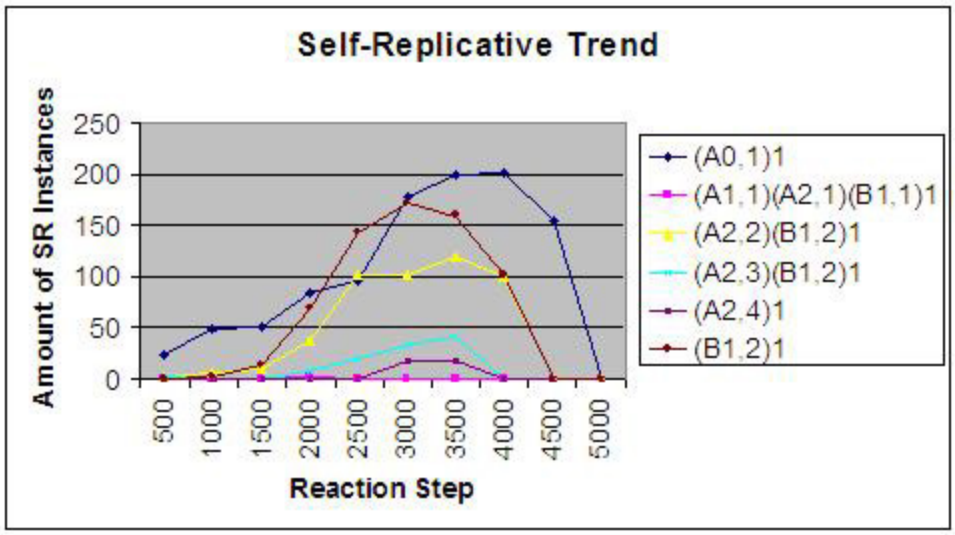

Figure 11.2: Statistical Data on the population of small reproductive graphs in ACG.

manner. This stochastic element in the chemistry demands that during experimental execution only strongly probable reaction will give rise to reproduction (if possible) of entities. Indeed with all the experiments which were carried out on the chemistry with different initial settings comprising of small seed graphs and simple weight matrix and saturation tables, only single node graphs were observed to be reproducing primarily because of high probability of corresponding reactions where single node graphs on reacting (colliding) with larger graphs would break off the larger graphs with multiple single node graphs as outputs.

Mutation as bounded by δ_{mut} and δ_{rep_mut} also permits possibility of mutant graphs to reproduce further using some related (with sequence of reaction by which original graph reproduces) or new chain (sequence) of reaction steps. Thus theoretically, chemistry allows observation of statistically significant number of reproducing mutants which will reproduce differently under the criterion of fitness defined as reproductive success per fix number of generations. And therefore , in theory, it is possible to observe selection effects in the chemistry. Nonetheless because of the stochastic nature of chemistry with sensitive dependence on initial setting in actual experiments were inconclusive on this aspect of mutation, heredity and selection because for single node reproducing graphs, it is not possible to define mutants in any reasonable manner.

11.3 Conclusions

Thus we have demonstrated that a precise observation process can be defined to work with new artificial graph chemistry, which can be used to discover the self replicating graphs (single node). Though, theoretically, every graph can reproduce, as well as its mutants, at different rates, the actual experiments were not conclusive to demonstrate upto this level.

Part IV

Related Work and Conclusion

Chapter 12

Related Work

In this chapter, we will present a brief discussion on related work on "observability" in real life and ALife studies. In biological literature there is little formal work on detecting life since there are sufficient biology-specific criteria (metabolism, morphological characters, bio-molecular structures etc) to distinguish life from non-life. There is, however some recent work in defining and developing methods to analyze genotype space structure based upon the observations on phenotype (mainly morphological and reactive characteristics) properties. This "observability problem" in real biology has been investigated using the concepts and tools from game theory to define necessary and sufficient conditions to uncover genotype space properties by observing only phenotype characteristics. In a static situation, in terms of the hereditary system, [GG98] presents a necessary and sufficient condition for the allele frequency-phenotype correspondence to be one-to-one. For the dynamic situation, in the strategic model of viability selection with symmetric genotype/phenotype correspondence, the observability problem was investigated in [GCKV03] providing sufficient conditions to recover the genetic process observing only phenotypic characteristics. In [LGC04], they extend these investigations to non-symmetric genotype/phenotype correspondences and also consider the observability problem in the classical replicator dynamics for the phenotypic evolution of an asexual population and in case of partial observation phenotype frequencies, providing sufficient conditions for observability.

To the authors' knowledge, there is not much work of similar nature focussing on observation processes for ALife studies reported in literature. In [BSP98] there is a

discussion on classification of long term adaptive evolutionary dynamics in natural and artificially evolving systems. To define their activity statistics, which quantifies the adaptive value of components (characteristics in our model) they employ similar mechanism of associating activity counters (tags) with all the components present in the system during simulation. Identification of new components is similar in nature to identifying new entities in our framework. Incrementing activity counters at every time step is similar to recognition of entities in the next states. Moreover their statistics can be directly integrated in our framework because appearance and continuation of any (new) component in the population is always associated with the appearance and continuation of entities which carry it.

Langton has defined in [Langton91] a quantitative matrix, called *lambda* parameter to detect life in any generic one dimensional cellular automata model based upon the characteristics of its transition rules (the fraction of rules which keep a cell in a particular state). His lambda parameter based analysis is based upon the assumption that any self organizing system can be treated as living and he does not consider evolutionary behavior as defining characteristics of life while defining his matrix. The lambda parameter tells when a particular CA model will undergo transition from an ordered state to chaotic state.

Self reproduction, which has long history of research starting from the late fifties [Sipper98] has evaded precise formal definition applicable to wide range of models. In some of the discussions related to self replication in cellular automata models [Samaya98b], [Morita98], there are formalizations of reproducing structures, but they do not attempt a general framework for observing reproduction or other components of evolutionary processes. Their formulation proceeds quite close to our definition of entities as discussed in the Section 8.2.

Chapter 13

Conclusion and Further Work

We will conclude the thesis by highlighting in brief the approach, the structure of the framework, and the main contributions of the work. This will be followed by discussion on the design suggestions for ALife researchers based upon the analysis of case studies. After that we will discuss major limitations of the framework and pointers for further work.

13.1 Conclusion

We have formalized an implicit underlying component of ALife studies, namely the observation process, by which entities are identified and their evolution is observed in a particular ALife simulation. Under the assumption that the essence of life-like phenomena is their evolutionary behavior, we developed a framework to formally capture basic components of evolutionary phenomena. This thesis, in essence, brings insights from evolutionary theory for real life into the realm of artificial life by defining a formal algebraic framework for observational processes, which are needed for the identification of life-like phenomena in ALife studies. We have argued that without such a formalism, claims pertaining to the evolutionary behavior in a chemistry will remain inconclusive.

In this thesis we focus upon the implicitly assumed notion of *observations to be carried out independent of the underlying chemistry structure* in ALife studies. We study observational processes independently by explicitly separating the chemistry from the abstractions used to describe it. We make it clear by placing observations into a distinct formal algebraic platform. The formalism to study observations is generic

enough to be applicable to a wide variety of models and specific differences among the models do not affect the applicability and analysis as illustrated in case studies on very different kind of ALife models, including two new experimental chemistries.

The discussion in Chapter 4 laid the conceptual foundation for the precise discussion of the evolutionary processes in ALife studies. The definitions presented in Chapter 5 form the basic formal foundation of the observation process and corresponding abstractions. The observation process is formally defined as a computable transformation from the simulations of the chemistry to various abstraction to be made upon it (Section 5.1). The Chemistry is observed as sequence of states and for each state entities are distinctly identified (Section 5.2). The framework demands that the entities to be observed and identified are represented as a vector of values for their measurable characteristics. The abstract space of entities is thus defined as nomological space in Section 5.3.

Having defined formally entities in terms of their measurable characteristics to be observed distinctly in successive states of the chemistry, the framework specifies that a suitable distance measure should be defined so that changes between the entities can be determined (Section 5.4). Moreover two entities can be considered alike only when the distance between them is below certain limits. Therefore framework specifies that limits on the observable changes in the entities to be defined precisely (Section 5.5). We introduced the notion of observed causality to determine precisely the parent-child relationship necessary for evolutionary phenomenon. The causal relationship specifies the entities in the chemistry which could have played any role in reproduction process of other entities. The observable limits when used in conjunction with the distance measure and the observed causality, gives precise definition of the ancestor-of relationship among entities. In order to avoid wrong inferences due to incomplete observation, we demand in the framework that relations to be defined for entities observed in successive states of the chemistry. The formalism developed so far in terms distance measure, observable limits, and causal relationship yielding ancestor-of relation gives us the precise formulation of sufficient reproduction as well as fecundity on entities as well as population levels as specified by the corresponding axioms (Section 6.2).

Reproduction is only one of the four essential components of evolution in neo-Darwinistic sense. The other components, variation, heritability, and selection, are

as well duly formalized in the framework in Chapter 6. The framework underlines two important kinds of changes causing variation in the characteristics of entities in the population. The first are the mutational changes in the entities as a result of their interaction with the environment (other entities) and second are the changes during reproduction. With mutational changes it becomes harder for the observation process to determine whether two entities in successive states are the same or different. Usually this requirement gets translated into determining as to whether two entities after the some reaction in the underlying chemistry should be treated alike. The second kind of reproductive changes make reproductive associations between entities difficult. The framework provides precise formulation of these issues and specifies basic axioms to guide such decision while working with actual ALife model.

Coupled with the basic formal structure of the observation process whereby entities are observed and characterized, the differences between entities are measured and bounded, entities are recognized across states of the chemistry and reproductive relationship between entities are established under changes, the heredity and the natural selection are then defined using axioms. We demand that in order to infer heredity and selection observation need to be made for statistically significant number of generations of entities in the chemistry. Furthermore the axiom of heredity is defined such that reproductive mutations should not undo the inherited changes occurred in ancestor generations to high proportion (Section 6.3). For natural selection the axioms of 'observations on evolutionary time scale', 'sorting', 'hereditary variation' and 'correlation' are defined which capture the requirements described in Section 1.3.1. In brief, starting with identification of entities, we defined the main ingredients of evolutionary processes algebraically and gave necessary conditions for evolution in the form of axioms.

The case studies on Langton loops (Chapter 8), Algorithmic Chemistry (Chapter 9), the Reduced Instruction Based Artificial Chemistry (Chapter 10), and the Artificial Graph Chemistry (Chapter 11) highlight the contributions that such an approach can make to the discussion of specific ALife experiments. An important property of such a study is to make explicit "multi-level observations", where entities and their relationship can be observed and defined on separate organizational levels. This was specifically discussed in case of example case study presented in Chapter 10 (Section 10.2). The

case studies also provide clues for ALife researchers for the design of their models as discussed next.

13.2 Design Suggestions for ALife Researchers

The formal framework defined in this thesis not only can be used to establish evolutionary behavior in a given ALife model, but can also be used to get some generic design suggestions which we believe should help ALife researcher while designing their models. Since the framework is based upon the neo-Darwinistic concepts of defining life in terms of evolutionary processes, the design suggestions we describe here are rather more suitable for those studies which aim to complement real life studies in an evolutionary framework.

Designed for Observation The chemistry should be designed and modeled for better observations. It should be clear from the case studies that the definition of entities in a chemistry which might be reproducing or are involved in evolutionary behavior is no trivial task. In some cases, like Langton's loops (Chapter 8), we can formalize the intuition regarding the nature of loops which reproduce but in general it is not clear how to determine on which level the entities in a chemistry might be evolving. Therefore it is imperative that chemistries which are designed such that entities on various possible levels can be easily defined and observed, are better candidates. In this respect RISAC (Chapter 10)) is a good candidate since we could easily design graphic interface which allows us to define entities on three different levels with or without spatial topology.

Maintenance of Variation and Heredity The chemistry should be designed such that variations in the characteristics of the entities which are generated by some means of mutational changes (environmental effects, reproductive mutations) are maintained and transferred to next generations with high probability. The requirement of strong heredity and maintenance of variation is critically needed to stop chemistries from converging rather quickly to a state consisting of only small number of different types of entities. This lack of maintenance of variation can be associated with the lack of strong selection in the simulations carried out in case

of artificial graph chemistry (Chapter 11), where reaction rules are such that most often entities change after every reaction before the changes could be transferred to next generation. Therefore entities cannot easily maintain their changes due to the nature of reaction rules.

Sufficient Reproduction with Variation The chemistry must be designed such that there exist potentially larger set of reproducing entities which should vary in their characteristics. Quite often this hinges upon the choice of reaction rules or the semantics of the chemistry and indeed it is a serious challenge for any chemistry designer to define the reaction semantics which permits potentially large set of reproducers. Another interesting aspect is that these reproducers must be relatively closely related to each other under the semantics. This means that sufficiently many variations of reproducers should also be reproducers in the chemistry otherwise the axiom of continuity of reproduction under mutation will not hold in the chemistry and most of the reproducers would have to appear *de novo* during simulations. We encounter this problem in most of the case studies discussed in Part III. In case of Langton loops, any kind of change in the loop structure would cause caseation of replication and thus Sayama work on designing Evoloops was fundamentally based upon the redefinition of the reaction semantics or transition rules. Similarly in the case of Algorithmic Chemistry, almost all of the single replicating λ terms arise *de novo* and their variations do not replicate under β reaction semantics. In the case of RISAC, we faced similar problem where small variations in the reproducing programs yield non replicating ones.

Measurable Rates of Reproduction The chemistry should be designed such that it is possible to impose some valid measure of determining the rates of reactions which in turn can be used to estimate differences in the rates of reproduction of different entities. This measurement of reproduction rates must be independent of the updation algorithm which uniformly selects entities to react. Therefore it can be argued that those chemistries, where all reactions take place in single step would be difficult to observe for natural selection, which works only when different entities reproduce at different rates. For example, it is not possible to infer differences in the rates of reproduction among different elementary hypercycles in

the Algorithmic Chemistry consisting of the same number of λ terms, which is because every reaction between any two λ terms occurs in single step. The other chemistries, like Evoloops can evolve natural selection precisely because different types of loops consisting of different number of cells reproduce at different rates based upon the number of state transitions.

13.3 Limitations

The decision to equate life with evolutionary processes also excludes some of the interesting complex phenomena that are not evolutionary in nature from the scope of this work. Indeed, we have shown in Chapter 9 that the framework cannot account for the dynamic non-evolutionary behavior of Level 1 and Level 2 organizations emerging in the Algorithmic Chemistry. We limit our attention to only those observations having evolutionary significance, though many other observations can be made upon the chemistry, which need to be addressed in further work. These include the phenomenon of metabolism [BF92, BFF92, Kitano94], the emergence of complexity, self organization and criticality under non linear dynamism, autonomous behavior [Kauffman89], which are not covered in our work. Our approach in the thesis has been limited to a population centric evolutionist approach of defining life as usually discussed in 'neo-Darwinistic' models for real life evolution [SS97, SS99]. This approach usually hinges upon the observations made only upon the evolutionarily active entities in the model and considers their reproductive relationship with each other.

We have not placed direct emphasis on certain concepts widely associated with ALife studies including the notion of "emergence". In our current setting the notion of "strong emergence" is only implicitly present and indeed "the element of surprise" [BE97] often associated with emergence is not immediate in the framework. Similarly "the element of autonomy" of emergent processes with respect to the underlying micro-level dynamics is not addressed in our framework. Indeed, the spirit of the high level of observations and corresponding abstractions upon which the framework rests, may preclude such inferences.

Nonetheless the idea of "weak emergence" [Bedau97], which lays emphasis on the simulations of the chemistry for the emergence of high level macro-states is fundamental

to our framework, where the observation process is by default based upon observations on the simulations of the chemistry and not on analytical derivations.

Another limitation of the framework in its current state is that it cannot be used effectively to make predictions regarding the possible observable evolutionary dynamics of the system during simulations. This limitation though carries forward from the nature of Darwinian theory which is too generic in its conceptualizations as well as is based upon random sources of change that make it very difficult to produce falsifiable claims or useful predictions.

13.4 Further work

The framework can be further extended in several directions including the following:

- To capture the essence of *strong emergence* by considering several observation processes at different organizational levels of the chemistry.
- To make explicit the genotype/phenotype dichotomy, which—we hope—will provide an adequate base for a formal definition of more complex evolutionary phenomena such as sexual reproduction.
- To capture the formal distinction between *Lamarckian* and *Darwinian* modes of evolution. While in case of Lamarckian evolution, entities change and these changes are inherited by the progenies in the next generations, in case of Darwinian evolution only changes during reproduction on genotype are inheritable.
- To offer definitions of life for systems not based upon evolution.
- To study overlapping evolutionary processes. Examples from real life include co-evolution, and sexual selection versus environmental selection.
- To study the tempo or rates of evolution in various systems. This will help to shed light on the controversies of punctuated equilibria versus gradualism (Section 1.3.5).
- The framework implies that the systems in which the axioms are met using some observational process possesses a qualitatively different kind of 'life' than those sys-

tems, in which axioms are not met. It will be interesting to explore this distinction in a formal way.

- As discussed in last section, the framework in its current state does not have enough structure to offer predictable analysis. It should be extended so that fruitful predictions for a given ALife model regarding the nature of evolutionary dynamics can be made.

Bibliography

[Adami98] Christoph Adami. Introduction to Artificial Life. Springer, New York 1998.

[Avida-Webpage] Introduction to Avida. `http://devolab.cse.msu.edu/software/avida/background.php`

[Bedau97] M. A. Bedau. Weak Emergence. In James Tomberlin, ed., *Philosophical Perspectives: Mind, Causation, and World*, vol. 11, pp. 375-399, Blackwell Publishers, 1997.

[BF92] Richard J. Bagley and J. Doyne Farmer. Spontaneous Emergence of Metabolism. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, eds., *Proceedings of the Workshop on Artificial Life, vol 5 of Santa Fe Institute Studies in the Sciences of Complexity*, pp. 93-140, 1992.

[BFF92] Richard J. Bagley, J. Doyne Farmer, and Walter Fontana. Evolution of a metabolism. n Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, eds., *Proceedings of the Workshop on Artificial Life, vol 5 of Santa Fe Institute Studies in the Sciences of Complexity*, pp. 93-140, 1992.

[Biology05] Science: Biology. `http://open-site.org/Science/Biology/`

[BSP98] M. A. Bedau, E. Snyder, N. H. Packard. A classification of long-term evolutionary dynamics. In C. Adami, R. Belew, H. Kitano, and C. Taylor, eds., *Artificial Life VI*, pp. 228-237, Cambridge: MIT Press, 1998.

[BE97] N. A. Baas, C. Emmeche. On Emergence and Explanation. *Intellectica* 1997/2, no.25, pp.67-83, 1997.

[BMRAGIKT00] Bedau, M., McCaskill, J. S., Packard, N., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. Open problems in artificial life. *Artificial Life 6*, pp. 363 - 376, 2000.

[Boden96] Boden, M. A. (ed.). The Philosophy of Artificial Life. Oxford University Press, New York, 1996.

[Burks70] A.W. Burks. Essays on Cellular Automata. University of Illinois Press. 1970.

[Capra97] Fritjof Capra. The Web of Life : A New Scientific Understanding of Living Systems. 1997.

[Cariani91] Cariani, P. Emergence and Artificial Life. In *Artificial Life II*, ed. C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, Addison-Wesley.

[Codd68] E.F. Codd. Cellular Automata. Academic Press 1968.

[DZB01] Peter Dittrich, Jens Ziegler, Wolfgang Banzhaf. Artificial Chemistries A Review (2001). *Artificial Life VII*, pp. 225 - 275, 2001.

[Darwin59] Charles Darwin. On the Origin of Species. 1859.

[Dawkins99] Richard Dawkins. The Extended Phenotype : the Long Reach of the Gene. Oxford University Press, 1999.

[ES79] M. Eigen, P. Schuster. The hypercycle: a principle of natural self-organization. In three parts: Part A, Emergence of the hypercycle. Part B, The abstract hypercycle. Part C, The realistic hypercycle. Springer-Verlag, Berlin, 1979.

[Evolution] Evolution - From Wikipedia. `http://en.wikipedia.org/wiki/Theory_of_evolution`

[FDB01] Pietro Speroni di Fenizio, Peter Dittrich, and Wolfgang Banzhaf. Spontaneous Formation of Proto-Cells in an Universal Artificial Chemistry on a Planar Graph. *Advances in Artificial Life*, LNCS 2159, Springer-Verlag, pp. 206 - 215, 2001.

[FJ94] S. Forrest, T. Jones. Modeling complex adaptive systems with Echo. In R.J. Stonier and X.H. Yu, eds. *Complex Systems: Mechanisms of Adaptation*, pp. 3-21, IOS Press, Amsterdam, 1994.

[FL94] W. Fontana and L. W. Buss. 'The Arrival of the Fittest': Toward a Theory of Biological Organization. *Math. Biol.*, vol 56, pp. 1 - 64, 1994.

[Fontana92] W. Fontana. Algorithmic chemistry. *Artificial Life V*, pp. 159 - 210, 1992.

[Crick70] Francis Crick. Central Dogma of Molecular Biology. Nature, vol 227, pp. 561-563, 1970.

[Futuyma98] Douglas J. Futuyma. Evolutionary biology. Sinauer Associates, c1998, 3rd ed.

[GCKV03] M. Gamez, R. Carreno, A. Kosa, Z. Varga. Observability in strategic models of viability selection. *Biosystems* vol 71 (3), pp. 249 255, 2003.

[GG98] J. Garay, M. B. Garay. Genetical reachability: when does a sexual population realize all phenotypic states. *J. Math. Biol.* vol 37, pp. 146 154, 1998.

[Gould81] Stephen Jay Gould. Evolution as Fact and Theory. In *Hen's Teeth and Horse's Toes: Further Reflections in Natural History*, pp. 253 - 262, 1994.

[GN77] Stephen Jay Gould and Eldredge, N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology, vol 3*, pp. 115-151, 1977.

[HJF97] P.T. Hraber, T. Jones and S. Forrest. The ecology of Echo. *Artificial Life III*, pp. 165-190, 1997.

[HS88] Josef Hofbauer, Karl Sigmund. The theory of evolution and dynamical systems : mathematical aspects of selection. Cambridge University Press, 1988.

[HMU00] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. Introduction to Automata Theory, Languages, and Computation. Second edition, Addition Wesley, 2000.

[Kauffman89] S. A. Kauffman. Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, Oxford, 1989.

[Kimura83] Motoo Kimura. The neutral theory of molecular evolution. Cambridge University Press, 1983.

[Kitano94] Hiroaki Kitano. Evolution of metabolism for morphogenesis. In Rodney A. In Brooks and Pattie Maes, eds., *Artificial Life IV*, pp. 49-58, 1994.

[Kuppers90] Bernd-Olaf Kuppers. Information And The Origin Of Life. MIT Press, 1990.

[Langton84] C. G. Langton. Self-reproduction in cellular automata. *Physica D, vol. 10*, pp. 135-144, 1984.

[Langton87] C. G. Langton. Artificial life. *Artificial Life: Proceedings of an interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, vol VI of SFI Studies in the Sciences of Complexity*, pp. 1-47, 1997.

[Langton89] C. G. Langton, ed. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, 1989.

[Langton91] C. G. Langton. Computation at the Edge of Chaos: Phase Transitions and Emergent Computation. In *Emergent Computation*, edited by Stephanie Forest, pp. 12-37, 1991.

[Langton95] C. G. Langton. Artificial Life: An Overview (Complex Adaptive Systems). MIT Press, Cambridge, MA, 1995.

[LGC04] I. Lopez., M. Gamez, R. Carreno. Observability in dynamic evolutionary models. *BioSystems vol 73*, pp. 99 109, 2004.

[McMullin00a] Barry McMullin. John von Neumann and the evolutionary growth of complexity: Looking backward, looking forward... *Artificial Life IV, vol 6*, pp. 347–361, 2000.

[McMullin00b] Barry McMullin. The Von Neumann Self-reproducing Architecture, Genetic Relativism and Evolvability. In *Evolvability Workshop at Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*, 2000.

[Margulis81] Lynn Margulis. Symbiosis in Cell Evolution: Life and its Environment on the Early Earth. Freeman, 1981.

[Margulis02] Lynn Margulis and Dorion Sagan. Acquiring Genomes: A Theory of the Origins of Species. Perseus Books Group, 2002.

[Mayr83] Ernst Mayr. The Growth Of Biological Thought. Harward University Press, 1983.

[MB97] M. Mahner, M. Bunge. Foundations of Biophilosophy. Springer, 1997.

[Morita98] Kenichi Morita. Cellular Automata and Artificial Life: Computation and Life in Reversible Cellular Automata. In E. Goles and S. Martinzez eds. *Complex Systems*, Kluwer Academic Publisher, pp. 151 - 200, 1998.

[MR98] Mayer B. and S. Rasmussen. The Lattice Molecular Automaton(LMA): A Simulation System for Constructive Molecular Dynamics. *International Journal of Modern Physics C, vol. 9(1)* , pp. 157-177, 1998.

[MSmith91] John Manyard Smith. Evolution: Natural and Artificial. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, eds., *Artificial Life II*, pp. 371 - 377, 1991.

[MSmith98] John Manyard Smith. Evolutionary Genetics, 1998.

[MV80] Humberto R. Maturana and Francisco J. Varela. Autopoiesis and Cognition: The Realization of the Living. Reidel, 1980.

[Neumann66] J. von Neumann. Theory of Self-Reproducing Automata. University of Illinois Press, 1966.

[OI01] N. Ono, T. Ikegami. Artificial Chemistry: Computational Studies on the Emergence of Self-Reproducing Units. *LINAI 2159* Springer, pp.186 - 195, 2001.

[Pattee95] Pattee H. Artificial life needs a real epistemology.*Advances in Artificial Life*, pp. 23 - 38, Springer-Verlag, 1995.

[Peng05] Bee Peng. An Experimental Plateform for Artificial Graph Chemistries. Honour's Year Thesis, School of Computing, National University of Singapore, 2005. Available upon request.

[Pinker02] Steven Pinker. The Blank Slate: the Mordern Denial of Human Nature. Penguin publication 2002.

[Popper63] K. R. Popper. Conjectures and Refutations: The Growth of Scientific Knowledge. London: Routledge and Kegan Paul, 1963.

[Ray91] Thomas 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*, pp. 371 - 408, 1991.

[Ridley96] Mark Ridley. Evolution. Blackwell Science, 2nd ed, 1996.

[Ridley97] Mark Ridley (Ed). Evolution. Oxford University Press, 1997.

[RKFH90] S. Rasmussen, C. Knudsen, R. Feldberg, and M. Hindsholm. The Coreworld: Emergence and Evolution of Cooperative Structures in a Computational Chemistry. *Physica D, vol. 42*, pp. 111 - 134, 1990.

[RKF91] S. Rasmussen, C. Knudsen, and R. Feldberg. Dynamics of Programmable Matter. *Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X*, pp. 211-254, 1991.

[Sayama98a] Hiroki Sayama. Introduction of Structural Dissolution into Langton's Self-Reproducing Loop. In Christoph Adami, Richard K. Belew, Hiroaki Kitano, and Charles E. Taylor eds., *Proceedings of the Sixth International Conference on Artificial Life*, 1998.

[Samaya98b] Hiroki Sayama. Constructing evolutionary systems on a simple deterministic cellular automata space, Doctoral Dissertation, Department of Information Science, Graduate School of Science, University of Tokyo, December 1998.

[Samaya99] Hiroki Sayama. A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life 5*, pp. 343-365, 1999.

[Sayama05] Hiroki Sayama. Java applet: Self-Replicating Loops in Cellular Space, available at `http://necsi.org/postdocs/sayama/sdsr/java/`, 2005.

[Schuster01] Peter Schuster. Mathematical Challenges from Molecular Evolution. In Bjrn Enquist and Wilfried Schmid, eds., *Mathematics Unlimited - 2001 and Beyond*, pp. 1019-1038, 2001.

[Sipper98] Moshe Sipper. Fifty Years of Research on Self-Replication: An Overview. *Artificial Life IV*, pp. 237–257, 1998.

[SO02] H. Suzuki, N. Ono. Several information space conditions for the evolution of complex forms of life. *Proceedings of the Seventh International Symposium on Artificial Life and Robotics (AROB 7th '02) , vol. 1*, pp. 175 - 178, 2002.

[SOY03] H. Suzuki, N. Ono, K. Yuta. Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life 9(2)*, pp. 537 - 558, 2003.

[SS97] John Maynard Smith and Eors Szathmary. The Major Transitions in Evolution. Oxford University Press; Reprint edition, 1997.

[SS99] John Maynard Smith and Eors Szathmary. The Origins of Life: From the birth of life to the origins of language. Oxford University Press, 1999.

[Stephen00] Stephen C. Stearns and Rolf F. Hoekstra, Evolution : An Introduction. Oxford University Press, 2000.

[Tierra-Webpage] Tierra home page `http://www.his.atr.jp/~ray/tierra/`.

[Volkenstein94] M.V. Volkenstein. Physical approaches to biological evolution. Springer-Verlag Berlin 1994.

[Vedvyas05] J. Vedvyas. The Reduced Instruction Set Artificial Chemistry (RISAC). Honour's Year Thesis, School of Computing, National University of Singapore, 2005. Available upon request.

[Werner05] Eric Werner. Genome Semantics, In Silico Multicellular Systems and the Central Dogma. FEBS Letters 579, pp. 1779-1782, 2005.