Towards a Framework for Observing Artificial Life Forms

Martin Henz

School of Computing

National University of Singapore

S16, Level 5, 3 Science Drive 2

Singapore 117543

Email: henz@comp.nus.edu.sg

Janardan Misra

Honeywell Technology Solutions Research Lab

151/1, Doraisanipalya

Bannerghatta Road

Bangalore 560076, India

Email: janardan.misra@honeywell.com

Abstract— **Evolutionary processes have emerged as the defining feature of "life" in Artificial Life (Alife). When studying the behavior of a particular Alife form, the question naturally arises, whether a particular run of an Alife experiment exhibits evolutionary behavior or not. This paper presents the Observer Framework, a formal framework for answering this question, based upon the notion of observations made in the Alife model at hand. Starting with defining entities and their relationships observed during the runs, the framework prescribes a series of definitions (decisions) that the observer of the Alife form needs to make, followed by axioms (conditions) that must be met in order to establish evolutionary behavior in particular runs. We use the example of Cellular Automata based Langton Loops to illustrate the Observer Framework, and suggest directions for further Alife research, based upon the framework design and the case study analysis.**

I. The Problem of Observing Artificial Life Forms

Researchers in Artificial Life (Alife) routinely propose computational environments and experiments that exhibit behavior of which life-like properties are claimed. Most of this work implicitly equates life-like behavior with evolutionary behavior (a standpoint that is not universally shared outside of Alife). Following this approach, the first question that arises when observing a "run" of an Alife experiment is what are the entities, or artificial "organisms" of which one may claim to observe evolutionary behavior. The problem of identifying time-varying entities has been recognized by Lehaniv and Dautenhahn [1], but is typically kept implicit in discussions of Alife experiments. Instead, researchers appeal to their reader's intuition to identify the entities of evolutionary behavior, similar to biologists who rarely worry about how to define the organisms that are the subject of study in a breeding experiment.[1] Problems arising from this informal approach include ambiguities in Alife research results, and the impossibility of automating the process of detecting artificial life in systematic experiments.

The first goal for a framework to precisely identify evolutionary behavior shall therefore be to formally distinguish the subjects of study, namely the objects of which life-like behavior is expected, from other phenomena in the experiment. Throughout the paper, we are emphasizing that such distinctions and deliberations are crucial decisions of the experimenter, which we call the "observer". Based on the precise notion of populations resulting from the identification process, the observer can attempt to identify the ingredients of evolutionary processes, including reproduction, heredity, variation due to mutational changes, and finally reproductive success based on natural selection [2], [3]. The Observer Framework describes each of these concepts as mathematical relationships between organisms, and thereby outlines a formal framework for identifying evolutionary behavior in Alife studies.

The paper is organized as follows: Section II formally defines the Observer Framework. Section III presents the central components of evolutionary processes cast as conditions on entities in the framework. Section IV applies the framework to a well-known Alife experiment—Langton Loops—as a case study. The remaining sections describe related work, limitations and conclusions, including design suggestions for Alife studies arising from the presented framework.

II. The Observer Framework

We shall see that the way in which the observer looks at the experiment has a decisive impact on his ability to identify evolutionary processes. The constructivist nature of our approach requires the observer to make these deliberations explicit. To distinguish them from general *Definitions*, we call them *Observer Decisions*. *Axioms* are used to specify conditions, which need to be established by the observer. For each fundamental component of evolution—self-reproduction, mutation, heredity, and natural selection—the Observer Framework specifies one or more Axioms specifying properties of the entities and relationships between entities (resulting from his decisions) that justify the claim of the respective component.

To illustrate the framework throughout this section, we shall use a simple example of a binary string based artificial chemistry, which we call **CBS** (Chemistry of Binary Strings).

---

[1]Sometimes, such a discussion is fruitful in biology, such as the view of mitochondria as organisms, replicating within higher life forms and through their maternal ancestor lines.

A. The Formal Structure of the Framework

We assume that the observer gains information on his Alife experiment by making snapshot-like observations, called *states*.

Definition 1: **[States]** Σ: set of observed states of an Alife model in a run.

The exact definition of a "state" varies from one Alife experiment to another owing to their irreducible design differences as well as the level of abstraction at which observations are being made. For example, in case of **CBS**, we can consider an observed state as a multiset of binary strings such as $\{00101, 00101, 10101, 010, 0100, 10100\}$.

Definition 2: **[Observed Run]** A sequence of states $T = [S_1, S_2, \ldots, S_n]$, ordered according to the temporal progression of the corresponding observations, represents one *observed run* of the experiment. The set of all observed runs of an experiment is denoted by $\mathcal{T}$.

The fundamental role of state sequences in the Observer Framework highlights the dynamic nature of evolution, reflecting the importance of *weak emergence* [4], a salient feature of most Alife studies. The model assumes that observations are made in form of snapshot-like states, which is suitable for many Alife experiments, but may often fail to account for the incremental nature of observations in organic life on earth.

B. Entities and their Characteristics

The first obligation of the observer is to define of what kinds of entities life-like behavior may be claimed.

Observer Decision 1: **[Entity Set]** The observer defines uniquely identifiable entities in every state. The set of all such entities is denoted by E.

The criterion to select the set of uniquely identifiable entities in a given state of the Alife model is entirely dependent on the observation process defined by the designer of the model. Thus for the same set of runs of an experiment, there may be different observed states as well as entities.

For example, in **CBS**, we might identify individual strings as entities. In order to make these entities uniquely identifiable, we would associate with every string an integer tag i. An entity corresponding to the binary string s can be represented as $[s]_i$. Thus a possible set of entities corresponding to the example state given above becomes $\{[00101]_1, [00101]_2, [10101]_3, [010]_4, [0100]_5, [10100]_6\}$. Alternately we may define an entity as a pair of strings with *three identical leftmost bits*, which would lead to $\{[\underline{001}01, \underline{001}01]_1, [\underline{101}01, \underline{101}00]_2, [\underline{010}, \underline{010}0]_3\}$ as the set of entities for the same state.

Observer Decision 2: **[State Function]** Let n be the number of states in an observed run. The function $F : E \mapsto \{1, \ldots, n\}$ returns for a given entity e the index i of the state S_i in which e is observed.

The next obligation of the observer is to identify characteristics of entities that are of interest. To this aim, we use a symbolic character space as follows:

Observer Decision 3: **[Character Space]** The observer defines a set of characteristics for entities in the model in a particular state or a sequence of states. Formally, define a multi-dimensional character space $\Upsilon = Char_1 \times \ldots \times Char_n$, where $Char_i$ denotes the set of values for i^{th} characteristic forming one dimension in Υ with zero element 0_{char_i}. Each entity $e \in E$ is thus a point in Υ, say $e = (v_1, v_2, \ldots v_n)$, where $v_i \in Char_i$.

For a vector $x = (a_1, a_2, \ldots, a_r)$, i^{th} element (a_i) will be denoted as $x[i]$. The absence of a characteristic $(Char_i)$ in an entity is represented by special zero element 0_{char_i}. Note that as a run of an experiment progresses, the dimensionality of Υ may vary because new entities with novel characteristic might emerge or entities with particular characteristic might disappear.

C. Distance Measures

The next component of the Observer Framework is the *dissimilarity measure* (D) to define the observable differences (**Diff**) between the characteristics of the entities in a population. The distance measure can be used by the observer to distribute entities into separate clusters such that entities in the same cluster are sufficiently similar to each other, compared to the entities outside the cluster.

Observer Decision 4: **[Distance Measure]** An observer defines a feasible distance measure $D : E \times E \rightarrow \mathbf{Diff}$, where **Diff** is the set of values to characterize the observable differences between entities in E.

The range of the distance measure D, **Diff** is a vector of values. Let $\mathbf{Diff} = (diff_0, diff_1, \ldots, diff_n)$, such that each element $diff_i$ represents the set of differences in the values for characteristics $Char_i$. The least element 0_{diff_i} is used when there is no observed difference between two entities for $Char_i$.

Examples include the Hamming distance between strings in **CBS**, the set of instructions where two programs may differ, and functional differences under reaction semantics of an artificial chemistry [5].

D. Observable Limits on Mutational Changes

Entities may change from state to state. The question arises, which entities are considered new in a given state, and which ones are recognizes as the "same" as entities observed in a previous state. In the Observer Framework, the observer needs to specify the bound vector δ_{mut}, under which he recognizes the persistence of an entity across states.

Reproduction also gives rise to changes in entities. In this case, an observer has to ascertain whether an entity is a descendent of another entity, or arises *de novo*. Thus, we introduce a second bound vector, δ_{rep}, for observable reproductive mutations. This bound is crucial when working with models where epigenetic development in the entities can be observed [6]. In such models—including typical examples from organic life on earth—the "child" entity and the "parent" entities initially do not resemble each other. The observer has to wait until the epigenetic developmental process unfolds, and then compare the entities for similarities in their characteristics. In less conventional experiments, the bound δ_{rep} allows us to distinguish between parent entities and other "input" entities

involved in the reproductive process, and between child entities and other "output" entities. Consider, for example, an artificial chemistry where entity A reproduces according to reaction $A + B \rightarrow 2A' + C$, where A' is a mutant child entity of A. Implicit in this notation is the observer decision that the pair A and A' enjoy sufficiently similarity with respect to their characteristics to be considered "parent" and "child", while the pairs B and A', and A and C do not. The bounds on observable differences are formally defined as follows:

Observer Decision 5: **[Mutation Bounds]** Based upon the choice of clustering distance measure D, an observer selects suitable vectors δ_{mut}, $\delta_{rep} \in$ **Diff**. Each component of δ_{mut} and δ_{rep} specifies a threshold on the recognizable mutational changes for the corresponding characteristic.

The choice of δ_{mut} and δ_{rep} critically affects further inferences. For example, large values for both bounds would result in the lack of identification of variability in the characteristics, which would result in difficulties to establish natural selection. Small values for δ_{mut} would lead to a failure to recognize persistence of an entity across states under changes, and small values for δ_{rep} make it hard to establish a reproductive relationship among entities, leading to a failure of establishing sufficient levels of fecundity.

E. Recognition and Causality

With these bounds in place, we can now formally define persistence of entities across states.

Definition 3: **[Recognition Relation]** An observer establishes persistence of entities across states of the model with (or without) mutations by defining the partial function **Rec**: $E \hookrightarrow E$.

The relation **Rec** must satisfy the following axioms:

Axiom 1: $\forall e, e' \in E$. $\mathbf{Rec}(e) = e' \Rightarrow F(e') = F(e) + 1$.

Informally, the axiom states that the entities to be recognized as same under mutational changes need to be observed in successive states. This is to avoid the cases where observed entities temporarily get out of the observations for certain states and then again reappear later, which might lead to unsound conclusions pertaining to persistence of entities. Note that in situations where the information available on each state is incomplete, such as typical observations on organic life on earth, this axiom would need to be weakened.

Axiom 2: $\forall e_1, e_2$ in S_i. $\nexists e'$ in $S_{i+1}.\mathbf{Rec}(e_1, e')$ and $\mathbf{Rec}(e_2, e')$.

This means that no two different entities in one state can be recognized as the same in the next state.

Axiom 3: $\forall e, e'$. if $\mathbf{Rec}(e, e')$ then $\mathbf{Diff}(e, e') < \delta_{mut}$.

Informally, if an entity in a state is recognised as being the same as another entity in the following state, then the difference between them must be below the mutation bound.

We aim to define reproduction from first principles, by reducing it to a causal relationship that satisfies a number of conditions.

Observer Decision 6: **[Observed Causality]** $C \subseteq E \times E$

If $C(e, e')$ holds, we say that e causes e' or e gives rise to e'. The relation C must satisfy the following axiom:

Axiom 4: **[Causality]** If $C(e, e')$ for some entities e, e', then $F(e') = F(e) + 1$ and $\nexists e'' \in F(e).\mathbf{Rec}(e'', e')$.

Informally, 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.

Notice that in order to establish a causal relation between entities, the observer does not need not refer to the underlying reaction semantics or "inner workings" of the experiment. Instead, the observer is free to claim causality, subject to the causality axiom above, as well as further conditions pertaining to the resulting notion of reproduction and are given in the next section. This may—in some cases—lead to problematic conclusions as further discussed in Section VI.

F. Observation Process

All observer decisions combined make up the *Observation Process* for a given Alife model $\Gamma = (\Sigma, \mathcal{T})$.

Definition 4: **[Observation Process]** An observation process Obs is defined as a computable transformation from the underlying universe of Alife model Γ to observer decisions $\Pi = (E, F, \Upsilon, D, \delta_{mut}, \delta_{rep}, C)$ and represented as $\Gamma \mapsto_{Obs} \Pi$.

The condition of computability ensures that the framework is feasible [7], which means the observation process only involves steps that can be algorithmically programmed by the designer of the model, and infeasible observations defined in terms of non-verifiable claims (e.g. claims based on 'meta-information') can be avoided.

III. COMPONENTS OF EVOLUTION

Having defined the observation process, we proceed now with observer decisions pertaining to the components of evolutionary processes.

A. Reproduction

Before we can stipulate the properties required for claiming the presence of reproduction in an Alife experiment, we need an auxiliary relation Δ to determine that the differences due to the reproductive mutations are bounded by δ_{rep}.

Definition 5: $\Delta \subseteq E \times E$ such that for (e, e') to be in Δ, their differences for each single characteristic $Char_i$ must be bounded by $\delta_{rep}[i]$, that is, $\forall Char_i \in \Upsilon$. $D(e, e')[i] \preceq_i \delta_{rep}[i]$

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

Definition 6: $\mathbf{AncestorOf} = ((C \cup \mathbf{Rec})^+ \cap \Delta)^+$

In this definition the (inner) transitive closure of $C \cup$ **Rec** captures the observed causality C across multiple states even in cases when parent entities might undergo mutational changes (**Rec**) before child entities complete their epigenetic development with possible reproductive mutations. Intersection with Δ ensures that causally related parent and child entities are not too different from each other, i.e. reproductive mutational changes remain below the bound δ_{rep}. The outer transitive closure ensures that all entities in an ancestor lineage

are considered ancestors themselves. When for two entities $e, e' \in E$, $(e, e') \in \textbf{AncestorOf}$, we say that e is *observed as an ancestor of e'*.
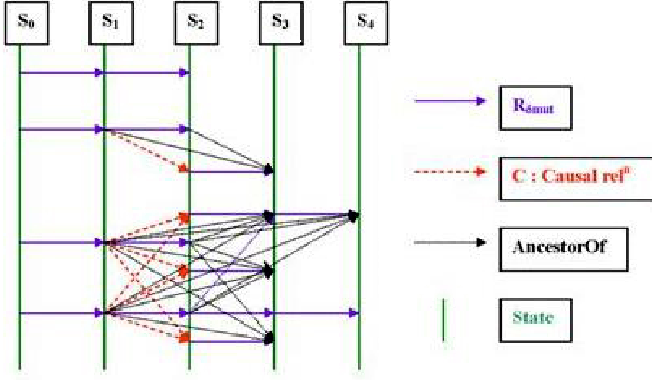

Fig. 1. Graphical view of the relationships between entities in successive states. Recognition relation $\mathbf{R}_{\delta_{\mathbf{mut}}}$, Causal relation C, and **AncestorOf**.

Figure 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.

This definition formally captures the recognition of reproductive relationships under parental mutations together with reproductive mutations in the child entities along with their epigenetic developments, which was believed to be difficult to formalize before [8]. The formalism also captures the case of multi-parent reproduction (without resorting to the concept of species) since a child entity can have several ancestors that are not ancestor of each other.

Using the **AncestorOf** relation, we now consider *reproduction*. For a given run of an experiment, the observer defines the following $\textbf{Parent}_\delta$ relation:

Definition 7: $\textbf{Parent}_\delta = \{ (p, c) \in \textbf{AncestorOf} \mid$ there is no intermediate ancestor of c *before* $p \}$.

The condition in defining $\textbf{Parent}_\delta$ is used to ensure that p is the immediate parent of c. Using the $\textbf{Parent}_\delta$ relation, in order for the observer to establish reproduction in the model, the following axiom must hold.

Axiom 5: **[Reproduction]** There exists an observed run $T \in \mathcal{T}$ of the model, where at least one instance of reproduction is observed, that is, $\textbf{Parent}_\delta \neq \emptyset$.

B. Fecundity

Though entity-level reproduction is essential, more significant for natural selection is the population-level collective reproductive behavior. The observer needs to establish that there is no perpetual decline in the size of the population, and thus there is a statistically sufficient number of generations that exhibit non-negative population growth. Formally we require the observer to establish fecundity by satisfying the following axiom:

Axiom 6: **[Fecundity]** There are statistically significantly many generations of entities $G_1, G_2, \ldots, G_n$ such that $(\forall G_i \subseteq E)(\exists G_{j>i} \subseteq E)$. $|G_j| \geq |G_i|$ where $G_j = \{c \in E \mid \exists a \in G_i$. $(a, c) \in \textbf{AncestorOf}\}$.
Here, the operator $|.|$ denotes the cardinality of a set.

We can now formulate another important axiom from evolutionary perspective, which asserts that reproduction in the model should not cease because of harmful mutations.

Axiom 7: **[Continuity of Reproduction under Mutations]** Some mutations preserve reproduction. In other words, there exists entities $e \in E$ that reproduce (with mutations) and one of the (mutant) children of e also reproduces.

C. Heredity

Heredity requires mechanisms to prevent the reversal of mutations in future generations by new mutations. To establish heredity in Alife models, sufficiently many generations of reproducing entities need to be observed to determine that the number of parent-child pairs, where certain characteristics were inherited by child entities without further mutations, is statistically significant. We can express this condition by the following axiom:

Axiom 8: **[Heredity]** Let Ω be a statistically large observed subsequence of a run T, and let $\textbf{Parent}_{\textbf{all}}$ be the set of all parent-child pairs observed in Ω. Then, there exists a characteristic i such that the set of entities in Ω, where the ith characteristics were inherited without (further) mutation is statistically significant.

The axiom of heredity together with the axiom of continuity of reproduction under mutation ensures that reproductive variation is maintained and propagated across generations.

D. Natural Selection

There are several notions of selection in the literature on evolutionary theory [9], [10], [6], [3]. In the spirit of the Observer Framework, we choose to define natural selection as a *statistical inference* of *average reproductive success*, which needs to be established over an evolutionary time scale i.e., over a statistically large number of states in a state sequence. Detailed notions of selection using fitness or adaptedness relative to the specific abstraction of "common environment" shared by entities and "the environment-entity interactions" are beyond the scope of this work. Here, we define the following (necessary) axioms for the natural selection:

Axiom 9: **[Significant Observations]** The observer must observe a statistically significant population of different reproducing entities, say Λ ($|\Lambda| \gg 1$), for a statistically large number of states in a state sequence $T \in \mathcal{T}$.

Axiom 10: **[Sorting]** The entities in Λ should be different with respect to their characteristics and there should exist differential rate of reproduction among these reproducing entities. The rate of reproduction for an entity is the number of child entities it reproduces before it undergoes mutations beyond δ_{mut}, or gets eliminated from the population.

The following axioms are aimed at establishing a distinction between natural selection and neutral selection [3].

Axiom 11: **[Heritable Variation]** There must be variation in heritable mutations in population of Λ. Formally, let $Child_{mut}$ be the set of child entities carrying reproductive mutations. Let $VarChild_{mut} \subseteq Child_{mut}$ be the subset of those child entities having mutations different from other entities. We require that $|VarChild_{mut}| \gg 1$.

This axiom implies that the number of child entities carrying different mutations is statistically significant.

Axiom 12: **[Correlation]** There must be a non-zero correlation between heritable variation and differential rate of reproduction. In other words as the value of characteristics inherited by the child entity changes, the rate of reproduction also consistently changes. Based upon the environmental pressures with respect to a particular characteristic, the rate of reproduction might either increase or decrease owing to the change in characteristic.

These two axioms state that a significant variation with respect to their characteristics is observed in a population of entities, which must be maintained for evolutionarily significant periods, that this variation is caused by differences in inherited characteristics, and that this variation directly affects the rate of reproduction.

IV. Case Study: Langton Loops

Having formalized the components of evolutionary processes to be observed in an experiment, we illustrate the Observer Framework the following section using Langton Loops as a case study.

A. General Considerations

In general, for a given Alife model, we suggest the following steps to instantiate the framework: The observation process works on runs of Alife experiments, which iteratively change the underlying states based upon the application of the updating rules. The observation process starts with the identification of states of the experiment (Σ) and state sequences ($\mathcal{T}$) during runs.

For every state in the state sequence, 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 measurable characteristics (Υ). The next important task is to define the limits on the observable mutational changes in individual characteristics of the entities (δ_{mut}, δ_{rep}), which will in turn define the recognition relation ($\mathbf{R}_{\delta_{\mathbf{mut}}}$) to determine entities persisting across states as well to determine whether two entities might be considered for descendent relationship.

Once the sets of entities in various successive states of the Alife model as well as their characteristics are known, we turn our attention to the evolutionary relationships between them. These relationships depend upon the intermediate causal relation (C) between the entities as observed under the mechanism 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 (**Parent$_{\boldsymbol{\Delta}}$**). These sets determine whether there are entities

which might be potentially reproducing in the model, even with observable differences between parent and child entities (Δ).

The next stage of the observation process is to ascertain the level of effectiveness of evolution in the model. Using the long term observations on the model for a statistically large number of generations, one can infer some statistical patterns for degrees of heredity and variation.

This process establishes the validity of all or some axioms of the framework for the given Alife model, which provides clues to the degree in which evolutionary processes might be at work for the Alife experiment at hand.

The case study on Langton Loops [11], based on Cellular Automata, will illustrate this process in detail. Alife models based on cellular automata offer a good example for an approach that emphasizes the observation process, since in such models replicating structures and their variations can be observed only with respect to a specific high level of abstraction.

B. Entities and Abstractions

We consider the case of two dimensional CA lattice based model. An observation is defined on the CA model 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. For a given cell $< (x, y), s > \in Cell$, we access its coordinates as follows: $co_x(< (x, y), s >) = x$, $co_y(< (x, y), s >) = y$, which can be extended to the set of cells: $\forall X \subseteq Cell$, $co_x^+(X) = \bigcup_{c \in X} co_x(c)$, $co_y^+(X) = \bigcup_{c \in X} co_y(c)$.

A CA-based model is initialized by setting a finite number of selected cells to non-quiescent states. At each step, the state of every cell of the model is changed synchronously as per the state transition rules. We define the *state* of the Langton model as the set of all non quiescent cells, Σ as the set of all possible states, and $\mathcal{T}$ as the set of state sequences obtained starting from some specific configuration. In the following discussion we will consider a fixed run given as $T \in \mathcal{T}$, starting with a specific initial state given in Figure 2 (Time 0).

a) Entities: We define entities as pairs consisting of two values: X, a connected set of non-quiescent cells, and an associated *pivot*. Two cells are connected only if there exists a consecutive sequence of neighboring non-quiescent cells joining them in the lattice. The *pivot* of such a set are coordinates of a cell uniquely associated with an entity in CA lattice in a particular state. To define a pivot function, an observer may choose the coordinates of the top left corner cell of an entity. Formally $pivot(X) = (min\{co_x^+(X)\}, max\{co_y^+(X)\})$This gives an obvious characterization for a two-dimensional character space $\Upsilon = Char_1 \times Char_2$ with $Char_1$ being the set of all non-quiescent connected sets of cells and $Char_2$ being the set of corresponding pivots, which can also be used to distinguish identical entities in the model.

b) Distance function: $D : E \times E \to \{0,1\} \times \{0,1\}$ is defined such that $\forall e, e' \in E . D(e,e') = (d_1, d_2)$, where d_1 and d_2 are defined as follows: $d_1 = 0$ if both entities have same number of cells arranged in same geometric arrangement, and $d_1 = 1$ otherwise; $d_2 = 0$ when the pivot for both the entities are the same and 1 otherwise.

c) Limits on Observable Mutations: The observer next selects $\delta_{mut} = [1, 0]$, which means that observer can recognize an entity in future states even with mutations (changes in the states, number, and the arrangement of cells comprising the entity) provided that the pivot remains the same. In contrast, the observer chooses $\delta_{rep} = [0, 1]$ which implies that for reproduction, the observer demands identical geometrical structure of the parent and child entities, although they may have different pivots—this is essential to capture exact replication of the loops.

d) Recognition relation: **Rec** $: E \to E$ is defined as follows: $\forall e, e' \in E, \mathbf{Rec}(e) = e' \Leftrightarrow [F(e') = F(e) + 1] \wedge [D(e,e') \leq \delta_{mut}]$ Informally this means two entities in consecutive states are recognized same only if they have the same pivots. This also means that the observer can recognize an entity even with changes in the number, state, and geometrical arrangement in the cells of an entity across states provided that entity does not shift in the CA lattice (which would result in the change of the pivot).

Lemma 1: **Rec** satisfies Axiom 1, Axiom 2, and Axiom 3.

Proof: Axiom 1 and Axiom 3 are satisfied by definition. Axiom 2, which states that **Rec** 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. ∎

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

$$(e, e') \in C \iff \begin{cases} 1. & co_x^+(X) \supset co_x^+(X') \\ 2. & co_y^+(X) \supset co_y^+(X') \\ 3. & pivot(X) \neq pivot(X') \\ 4. & F(e') = F(e) + 1 \end{cases}$$

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 2 at time step 127.

Lemma 2: The causal relation C defined above satisfies the Causality Axiom.

Proof: Condition $F(e') = F(e) + 1$ insures that e and e' are not observed in the same state. To establish that e' is not the result of mutations in some other entity e'' observed in past (i.e., $[F(e'') = F(e)] \wedge [\mathbf{Rec}(e'') = e']$) we note that because of the definition of **Rec**, e'' and e' would otherwise have the same pivots, which means pivot of e'' will be included in the set of cells in e (since $[co_x^+(Z_e) \supset co_x^+(Z_{e'})] \wedge [co_y^+(Z_e) \supset co_y^+(Z_{e'})]$), which is not possible because e and e'' being
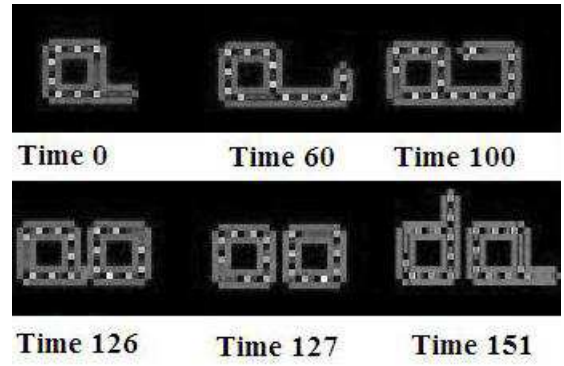

Fig. 2. Self-Reproduction in Langton loops (images generated using Bachmutsky's Java implementation [12])

different entities in the same state cannot have cells in common including pivot as argued above in the proof of previous lemma. ∎

C. Reproduction and Fecundity

Lemma 3: **Axiom of Reproduction** and the **Axiom of Fecundity** are satisfied by the entities and abstractions on Langton Loops described above.

Proof: These two axioms can be established by the observer in a specific state sequence as exemplified in Figure 2 and Figure 3 by repeatedly applying the recognition relation **Rec** when entities are changing in number and states of cells (retaining the pivots) and applying the causal relation when a parent entity splits (e.g. at Time=127). The relation Δ connects the initial parent entity and the child entity at Time=151.

With respect to Figure 2, the single parent entity is identified at Time=0 with associated pivot. Between time steps $[1 \dots 126]$ the parent entity changes in number and states of its cells but the pivot remains the same, hence according to the definition of **Rec** as explained above, the observer can recognize the entity in these successive states even while they change their structure. At Time=127, the parent entity is observed to be splitting into two identical copies. One of these is again recognized as the original parent entity because its pivot remains the same, and the second entity is causally related with the parent entity as per the definition of C. To see this, notice that the parent entity at Time=126 contains all the cells of the child entity appearing at Time=127, which satisfies the definition of C. Between time steps 128 and 151 both parent and child entities undergo changes in the number and states of their cells but their pivots remain fixed. Hence they can again be recognized. Finally at Time=151 the child entity becomes geometrically identical to the original parent entity, therefore the parent entity at Time=0 and the child entity at Time=151 are related using Δ. The transitive closure finally give us the final descendence relationship between the parent and the child entity. ∎
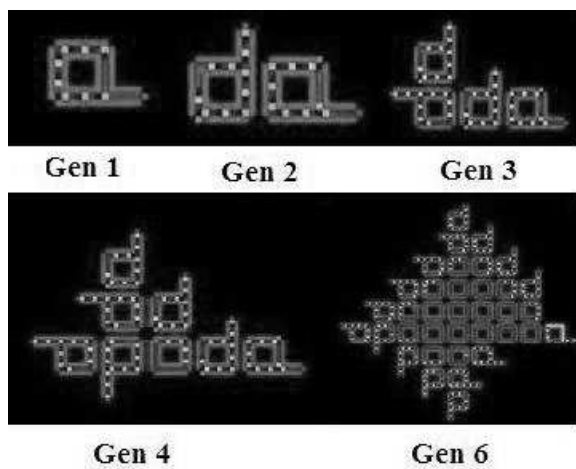
Fig. 3. Fecundity across generation in a population of Self Replicating Langton Loops (image generated using Bachmutsky's Java implementation [12])

D. Mutations, Inheritance, and Natural Selection

Langton loops, though self-replicating, do not exhibit behavior that can be interpreted as reproductive and inheritable mutations. This can be attributed to the choice of the underlying state transitions defined for the cells in the model. *Evoloops*, Samaya's extension of Langton loops represent an attempt of adding inheritable mutations to this style of Alife experiments [13]. In his model, the loops differ in the number and geometrical arrangement of cells. The population witnesses variations of different kinds such that different reproducing loops are scattered on the lattice, forming irregular colonies. Evoloops and their evolution can be analyzed in the Observer Framework by suitably modifying the definition of the distance measure D to measure the differences between the entities in the number and geometric arrangement of cells and by changing limit δ_{rep} such that the observer is able to establish a descendence relationship even when the parent and the child entities are not identical.

V. RELATED WORK

Not much work focussing on the observation process for Alife studies has been reported in literature. The framework presented here, however, can be seen in contrast to other proposals to define *numerical parameters* or *statistics* [7] to recognize life in a model. We are not sure whether there can be simple numerical definitions capturing the essence of life in arbitrary models and even if so does not seem to be the case with the current proposals. The difficulty arises out of intricate nature of selection inevitably involving non trivial identification of the population of evolving entities.

Langton defined a quantitative metric, called *lambda* parameter, to detect life in any generic one-dimensional cellular automata model based on transition rules [14]. Bedau et al. discuss a classification of long-term adaptive evolutionary dynamics in natural and artificially evolving systems by defining activity statistics for the components, which quantifies

their adaptive value [15]. These approaches implement the identification process of entities emphasized in the Observer Framework.

Self-reproduction, which has a long history of research starting from the late 1950s [16] has evaded precise formal definition applicable to a wide range of models [1]. In some of the discussions related to self-replication in cellular automata models [13], [17], formalizations of reproducing structures are presented, but they do not attempt to provide a general framework for observing reproduction or other components of evolutionary processes. The existing attempts at formalizing reproduction are reminiscent of our definition of entities (loops) as discussed in Section IV.

VI. LIMITATIONS

The decision to equate life with evolutionary processes excludes for the scope of this work interesting processes that in our view lie outside the set of essential ingredients of evolutionary processes, including metabolism [18], self-organization [19], and autonomous and autopoitic processes [20].

The framework does not place direct emphasis on the notion of *emergence*. In our current setting, the notion of *strong emergence* is only implicitly present and indeed the *element of surprise* [21] often associated with emergence is not represented in the framework. Similarly the *the element of autonomy* of emergent processes with respect to the underlying micro-level dynamics is not addressed in the framework. Nonetheless the idea of *weak emergence* [22], which emphasizes the importance of experimental simulation for the emergence of high level macro-states, is fundamental to the framework.

Like any other generic specification framework, the Observer Framework also suffers from the weakness of administering false positives. *False positive* refers to a situation where observations and consequent inferences on a model result in a claim of the presence of a certain property in the model which actually does not exist. The problem of false positives is due to the necessarily domain-independent definition of causality, which cannot account for actual causal relationships within the underlying micro-level dynamics of the experiment. The generic nature of causality might give rise to false claims on the presence of evolution in the model. For example, an observer might decide to "ignore" entities in some states in the beginning and then choose later on to observe them in some other states so that to use them for establishing (false) evolutionary relationships, which would not have been possible had he not preferred to ignore them earlier. The problem of selectively observing entities in various states requires additional constraints in the framework.

VII. CONCLUSION

A. General Remarks

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 par-

ticular runs. 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. The observation process as specified in the framework may be carried out manually or can be automated and integrated within the model.

The Observer Framework defines aspects of life including recognition of reproductive relationships under parental mutations as well as reproductive mutations in children along with their epigenetic developments, which were previously believed to be difficult to formalize [1], [8]. The framework also captures the case of multi-parent reproduction (without resorting to the concept of species), and the case of reproduction without overall growth of the population [1].

The framework design and the case study analysis provides the following clues for Alife experiment design to improve the ability to witness evolutionary phenomena in runs.

B. Design Suggestions for Alife Research

1) Sufficient Reproduction with Variation: Alife experiments should be designed such that there exist a potentially large set of reproducing entities, which are semantically related and have significant variation in their characteristics. Semantic relatedness means that sufficiently many variations of reproducing entities should be reproducing themselves, since otherwise most of the reproducing entities would have to appear *de novo* during experiment runs, which will make it difficult to meet the axiom of Heritable Variation (Axiom 11).

2) Measurable Rates of Reproduction: Alife experiments should be designed such that it is possible to impose measures for determining the rates of reactions which can be used to estimate variation in the rates of reproduction in a population. This measurement of reproduction rates should be independent of the algorithm that selects entities for reaction. It can be argued that in Alife experiments, where all (reproductive) reactions take place in a single step, natural selection—which can be observed only when different entities reproduce at different rates—may be difficult to observe.

C. Further Work

The Observer Framework can be extended in several interesting directions, including the following:

- The essence of *strong emergence* could be captured by considering several observation processes at different organizational levels.
- Conditions for overlapping evolutionary processes—examples from real life include co-evolution, and sexual selection versus environmental selection—could be formulated within the framework.
- Stricter axioms may be able to partially overcome the problem of false positives, such that false claims of causality are bound to give rise to insurmountable difficulties in meeting other aspects of the framework.
- Additional concepts and axioms may lead to distinction between genotype and phenotype, and a definition of Darwinian and Lamarckian evolution within the Observer

Framework [3]. This distinction may then lead to a precise definition of sexual reproduction [23].
- Further insights shall be gained by applying the framework to novel classes of Alife experiments to refine the framework.

REFERENCES

[1] C. L. Nehaniv and K. Dautenhahn, "Self-replication and reproduction: Considerations and obstacles for rigorous definitions," *Third German Workshop on Artificial Life: Abstracting and Synthesizing the Principles of Life*, pp. 283–290, 1998.

[2] R. Dawkins, "Universal darwinism," in *Evolution from Molecules to Men*, D.S.Bendall, Ed. (Cambridge: Cambridge University Press, 1982, pp. 403–25.

[3] S. C. Stearns and R. F. Hoekstra, *Evolution–An Introduction*. Oxford University Press, 2000.

[4] M. A. Bedau, J. S. McCaskill, N. H. Packard, S. Rasmussen, C. Adami, D. G. Green, T. Ikegami, K. Kaneko, and T. S. Ray, "Open problems in artificial life," *Artificial Life*, vol. 6, no. 4, pp. 363–376, 2000.

[5] P. Dittrich, J. Ziegler, and W. Banzhaf, "Artificial chemistries—a review," *Artificial Life*, vol. 7, no. 3, 2001.

[6] M. Mahner and M. Bunge, *Foundations of Biophilosophy*. Springer, 1997.

[7] M. A. Bedau, "Can unrealistic computer models illuminate theoretical biology?" in *Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program*, Orlando, Florida, 1999, pp. 20–23.

[8] C. L. Nehaniv, "Self-replication, evolvability, and asynchronicity in stochastic worlds," *Proc. 3rd Symposium on Stochastic Algorithms, Foundations and Applications*, vol. 3777, p. (in press), 2005.

[9] M. Kimura, *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.

[10] D. J. Futuyma, *Evolutionary Biology*, 3rd ed. Sinauer Associates, 1998.

[11] C. G. Langton, "Self-reproduction in cellular automata," *Physica D*, vol. 10D, no. 1-2, pp. 135–44, 1984.

[12] E. Bachmutsky, *Java applet: Self-Replicating Loops in Cellular Space*. Available at http://necsi.org/postdocs/sayama/sdsr/java/, 2005.

[13] H. Sayama, "Constructing evolutionary systems on a simple deterministic cellular automata space," Ph.D. dissertation, Department of Information Science, Graduate School of Science, University of Tokyo, December 1998.

[14] C. G. Langton, "Computation at the edge of chaos: Phase transitions and emergent computation," *Physica*, vol. D 42, pp. 12–37, 1990.

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

[16] M. Sipper, "Fifty years of research on self-replication: An overview," in *Artificial Life IV*, 1998, pp. 237–257.

[17] K. Morita, "Cellular automata and artificial life: Computation and life in reversible cellular automata," in *Complex Systems*, E. Goles and S. Martinzez, Eds. Kluwer Academic Publisher, 1998, pp. 151–200.

[18] R. J. Bagley, J. D. Farmer, and W. Fontana, "Evolution of a metabolism," in *Artificial Life II*, C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, Eds. Redwood City, CA: Addison-Wesley, 1992, pp. 141–158.

[19] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press, 1993.

[20] M. Zeleny, Ed., *Autopoiesis: A Theory of Living Organization*. New York: North Holland, 1981.

[21] N. A. Baas and C. Emmeche, "On emergence and explanation," *Intellectica*, vol. 1997/2, no. 25, pp. 67–83, 1997.

[22] M. A. Bedau, "Weak emergence," in *Philosophical Perspectives: Mind, Causation, and World*, J. Tomberlin, Ed. Blackwell Publishers, 1997, vol. 11, pp. 375–399.

[23] D. Misevic, R. Lenski, and C. Ofria, "Sexual reproduction and Muller's ratchet in digital organisms," in *ALIFE 9: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems*, R. Watson, Ed., 2004, pp. 340–346.