

# A Universal Framework for Self-Replication

Bryant Adams<sup>1</sup> and Hod Lipson<sup>2</sup>

<sup>1</sup> Department of Mathematics,

<sup>2</sup> Department of Mechanical and Aerospace Engineering,  
Cornell University, Ithaca NY 14853, USA

badams@math.cornell.edu, hod.lipson@cornell.edu

**Abstract.** Self-replication is a fundamental property of many interesting physical, formal and biological systems, such as crystals, waves, automata, and especially forms of natural and artificial life. Despite its importance to many phenomena, self-replication has not been consistently defined or quantified in a rigorous, universal way. In this paper we propose a universal, continuously valued property of the interaction between a system and its environment. This property represents the effect of the presence of such a system upon the future presence of similar systems. We demonstrate both analytical and computational analysis of self-replicability factors for three distinct systems involving both discrete and continuous behaviors.

## 1 Overview and History

Self-replication is a fundamental property of many interesting physical, formal, and biological systems, such as crystals, waves, automata, and especially forms of natural and artificial life [1]. Despite its importance to many phenomena, self-replication has not been consistently defined or quantified in a rigorous, universal way. In this paper we propose a universal, continuous valued property of the interaction between a system and its environment. This property represents the effect of the presence of such a system upon the future presence of similar systems. Subsequently, we demonstrate both analytical and computational analysis of self-replicability factors for three distinct systems involving both discrete and continuous behaviors.

Two prominent issues arise in examining how self-replication has been handled when trying to extend the concept universally: how to deal with non-ideal systems and how to address so-called ‘trivial’ cases [2,3]. Moore [4] requires that in order for a configuration to be considered self-reproducing it must be capable of causing arbitrarily many offspring; this requirement extends poorly to finite environments. Lohn and Reggia [5] put forward several cellular-automata (CA) -specific definitions, and result in a binary criterion. A second issue that arose in the consideration of self-replicating automata was that some cases seemed too trivial for consideration, such as an ‘all-on’ CA, resulting in a requirement for Turing-universality [6].

The definition for self-replicability we propose here is motivated in part by (a) A desire to do more than look at self-replication as a binary property applicable only to

certain automata, and, (b) The goal of encapsulating a general concept in a means not reliant upon (but compatible with) ideal conditions.

We wish to do this by putting self-replication on a scale that is algorithmically calculable, quantifiable, and continuous. Such a scale would allow for comparisons, both between the same system in different environments, determining ideal environments for a system's replication, as well as between different systems in the same environment, if optimizing replicability in a given environment is desired.

Rather than viewing self-replicability as a property purely of the system in question, we view it as a property of the interaction between a system and its environment. Self-Replication, as we present it, is a property embedded and based upon information, rather than a specific material framework. We construct replicability as a property relative to two different environments, which indicates the degree to which one environment yields a higher presence of the system over time. *Self-replicability*, then, is a comparison between an environment lacking the system and an environment in which the system is present. We will first introduce a number of definitions, and then give examples of replicability of three types of systems.

## 2 Definitions

**Definition 1: *Environment.*** *By Environment we denote a single state of a (presumably closed) system.*

For example, with a closed system such as a 5×5 grid, each of whose cells may be either ‘on’ or ‘off’, one environment,  $E_1$ , would be ‘all 25 cells off’, while another,  $E_2$ , might be ‘every other cell on’.

**Definition 2: *Set of configurations.*** *Given an environment  $E$ , the set of configurations  $\bar{E}$  is the set of all possible states of the system that includes state  $E$ . We call elements of  $\bar{E}$  “ $E$ -configurations”.*

We do not assume an arrow of time, and thus if, for some environments  $E_1, E_2$ , we have  $E_1 \in \bar{E}_2$ , then  $E_2 \in \bar{E}_1$  and  $\bar{E}_1 = \bar{E}_2$ . In the 5×5 grid, for example, the set of configurations would be the collection of all  $2^{25}$  states the grid could be in.

**Definition 3: *Time development function.*** *A time development function is a map  $T: \bar{E} \times \mathfrak{R}^+ \rightarrow \bar{E}$ , (respectively  $T: \bar{E} \times \mathbb{Z}^+ \rightarrow \bar{E}$  for discrete time systems) constrained by  $T(E, 0) = E$  and  $T(T(E, y), x) = T(E, x+y)$ .*

With a time development function, we operate under the assumption that the progression between states as time passes is externally deterministic. We also assume there is no difference between a system that ‘ages’ five units and then ten units, and a system that first ages ten units, and then ages five units. When we write only  $T(E)$  rather than  $T(E, n)$ , it is assumed there is a natural increment of time and we are using  $T(E, 1)$ .

**Definition 4: Subsystem set.** We denote by  $X^*$  the collection of all subsystems of a given system  $X$ , and say  $X^*$  is the subsystem set of  $X$ .

Often, we are most interested in the subsystem set of an environment. For example, in the case where the system is a  $5 \times 5$  grid and  $E$  is a state with all points being off,  $E^*$  would include, among its  $2^{25}$  elements, four  $4 \times 4$  binary grids with all points being off, five  $1 \times 5$  binary grids with all points being off, and a number of L-shaped binary grids with all points being off. For further example, given a second system  $E_2$ , a  $5 \times 5$  grid in which the top row of points were on, the rest off,  $E_2^*$  would still have  $2^{25}$  elements, each the same shape as a corresponding element in  $E^*$ , but now some elements would include ‘on’ points.

**Definition 5: Possible Subsystems.** We define the possible subsystems  $\bar{E}^*$  as the union of all  $F^*$  such that  $F \in \bar{E}$ .

In the case of the binary grid, the elements of  $\bar{E}^*$  could be classified into  $2^{25}$  distinct shapes, and each shape-class would have  $2^n$  elements, with  $n$  being the number of cells in that shape.

**Definition 6: Dissimilarity pseudometric.** To quantify the ‘self’ portion of self-replication, we assume that a dissimilarity metric  $d : \bar{E}^* \times \bar{E}^* \rightarrow \mathfrak{R}^+$  is given. Recall a pseudometric  $d$  obeys  $d(x,y)+d(y,z) \geq d(x,z)$ ,  $d(x,y) \geq 0$ , and  $d(x,x)=0$ .

Note that  $d$  induces an equivalence relation on  $\bar{E}^*$ . That is, we say for  $S_1, S_2 \in \bar{E}^*$ , that  $S_1 \equiv S_2$  exactly when  $d(S_1, S_2)=0$ . Presumably, the dissimilarity metric would be chosen such that it induced a natural equivalence relation, but this choice is not assumed.

**Definition 7: Presence.** We define the presence  $P_\varepsilon(E, S)$  of a subsystem  $S$  in an environment  $E$  within tolerance  $\varepsilon$  to be the measure  $E^*$  normalized to 1. I.e, it is the probability that a randomly selected subsystem  $T \in E^*$  will satisfy  $d(T, S) \leq \varepsilon$ .

Essentially, the presence function measures ‘how much’  $S$  is found in  $E$ . As a probability,  $P$  takes values in the interval  $[0, 1]$ . In the case of discrete environments, the presence function essentially reduces to counting, and in a continuous case we (temporarily) increase the measure of the set by using a nonzero tolerance.

**Definition 8:  $\varepsilon$ -Present,  $\varepsilon$ -Possible.** When  $P_\varepsilon(E, S) \neq 0$ , we say that  $S$  is  $\varepsilon$ -present (in  $E$ ). Also, we say that  $S$  is  $\varepsilon$ -possible (in  $E$ ) when there is some time  $t \in \mathfrak{R}^+$  such that  $S$  is  $\varepsilon$ -present in  $T(E, t)$ .

**Definition 9: Replicability, Momentary.** Given a set of configurations  $\bar{E}$  and two  $E$ -configurations  $E_1, E_2$ , we define the momentary relative replicability of a system  $S$  in  $E_1$  relative to  $E_2$  with tolerance  $\varepsilon$  at time  $t$  as

$$R_M(S, E_1, E_2, \varepsilon, t) = \log \frac{P_\varepsilon(T(E_1, t), S)}{P_\varepsilon(T(E_2, t), S)} \quad (1)$$

The ratio in Eq. (1) serves to compare the probability at time  $t$  of finding  $S$  in the future of  $E_1$  to the probability at the same time of finding  $S$  in the future of  $E_2$ . There are a few cases where Eq. (1) is undefined. If  $P_\varepsilon(T(E_1, t), S) = P_\varepsilon(T(E_2, t), S)$  (including

zero) we define  $R_M$  as 0. When  $P_\varepsilon(T(E_1,t),S)=0$  but  $S$  is  $\varepsilon$ -present in  $T(E_2,t)$ , we define  $R_M$  as  $-\infty$ , and when  $P_\varepsilon(T(E_2,t),S)=0$  but  $S$  is  $\varepsilon$ -present in  $T(E_1,t)$ , we define  $R_M$  as  $\infty$ .

In the case where  $S$  is not  $\varepsilon$ -possible in either or both of  $E_1$  and  $E_2$ , we will not define replicability. In these cases, we would generate ratios comparing possibly non-zero quantities to zero quantities, which would fail to yield meaningful information. We explain the rationale for using the logarithm after all the definitions have been presented.

**Definition 10: Replicability, Over time.** In the case where  $S$  is  $\varepsilon$ -possible in both  $E_1$  and  $E_2$ , we also define the replicability over time  $\tau_0$  to  $\tau_1$  (in  $E_1$  relative to  $E_2$  with tolerance  $\varepsilon$ ) as:

$$R_T(S, E_1, E_2, \varepsilon, \tau_0, \tau_1) = \log \frac{\int_{t=\tau_0}^{\tau_1} P_\varepsilon(T(E_1, t), S) dt}{\int_{t=\tau_0}^{\tau_1} P_\varepsilon(T(E_2, t), S) dt} \quad (2)$$

Note that for discrete systems, integral reduces to a simple sum.

**Definition 11: Replicability, Overall.** We define the Overall Replicability as the limiting case:  $R_O(S, E_1, E_2, \varepsilon) = \lim_{t \rightarrow \infty} R_T(S, E_1, E_2, \varepsilon, 0, t)$

Note that, by using the logarithm of the fractions, we have an additive form for some basic relations. Letting  $R(S, E_1, E_2)$  stand in for any of the defined types of replicability, letting  $S$  be a fixed system, and letting  $A, B, C \in \overline{C}$ , we have:

$$\begin{aligned} R(S, A, B) + R(S, B, C) &= R(S, A, C) \\ R(S, A, B) &= -R(S, B, A) \\ R(S, A, A) &= 0 \end{aligned} \quad (3)$$

The last relation highlights that replicability is being taken as a relative, rather than absolute, concept. When a replicability value is zero, we have two environments in which a system fares equally well. In order to relate the replicabilities of systems in different environments, we will specifically consider cases where  $S$  is present in  $E_1$ , is not present in  $E_2$ , and  $d(E_1, E_2)$  is minimal. In these cases, we are making a comparison between a ‘blank’ environment and one that is minimally different, in order to reflect a minimally disturbing introduction of a system into the environment. For example, in modeling a crystal, we might consider a supersaturated solution as  $E_2$ , and a supersaturated solution with a tiny seed crystal as  $E_1$ , but we would not consider a fully crystallized configuration for  $E_1$ .

In particular, let an environment  $E$  and a system  $S$  be given, such that  $S$  is minimally  $\varepsilon$ -present in  $E$ . Let  $E' = E - S$  denote some  $E' \in \overline{E}$  such that  $d(E, E')$  is minimal, and  $S$  is not  $\varepsilon$ -present in  $E'$  but is  $\varepsilon$ -possible in  $E'$ . This leads to the particular case of self-replicability we wish to examine.

**Definition 12: Self-Replicability, Overall.** With  $E, S$ , and  $E - S$  given as above, we define the self-replicability of a system  $S$  in an environment  $E$  as  $R_S(S, E, \varepsilon) = R_O(S, E, E - S, \varepsilon)$ .

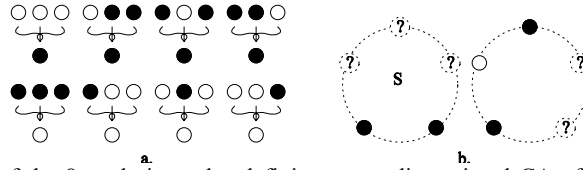
In essence,  $R_S$  looks at how present  $S$  becomes when it was at one point specifically in the environment, compared to how develops without prompting. If  $R_S$  is zero, then the inclusion of  $S$  neither adds nor detracts from the future presence of  $S$ . A

positive value of  $R_S$  indicates that including the system results in a higher presence of the system in the future, i.e. the system appears to be self-replicating. A negative value indicates that including the system has a detrimental effect, and that the system is essentially self-defeating. Infinite values, both positive and negative, arise when the magnitude of the influence of the system increases at a more than linear rate as time goes on. This can happen in particular when the lifespan of a system is finite in one of  $E$  or  $E-S$  and infinite in the other.

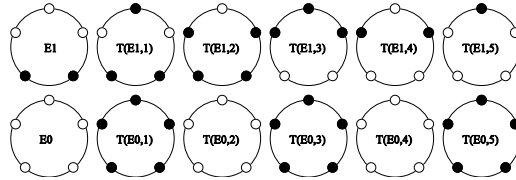
### 3 Examples

#### 3.1 Cellular Automaton

Our first system is a cellular automaton given by single-cycle graph with five nodes, each of which can take the state of ‘on’ or ‘off’, along with radius-1 evolution rules (Fig. 1a). Up to rotations and reflections of the graph, there are eight distinct states the system can take. There are two limit cycles under the evolution rule: First, the ‘all on’ state is sent to the ‘all off’ state, which is then sent to the ‘all on’ state again. Second, the remaining six states are transitive under the action of the evolution rule.



**Fig. 1.** (a) Table of the 8 evolution rules defining a one dimensional CA of radius 1 and (b) Two examples of subsystems. We will calculate the self-replicability of the one labeled  $S$



**Fig. 2.** Cellular Automaton system: Successive states of  $E_0$  and  $E_1$  under  $T$

Examples of two subsystems are shown in Fig. 1b. As this is a discrete system, we can move directly to the zero-tolerance case, and thus need to define only the preimage of zero under dissimilarity metric (i.e. establish the meaning of ‘identical’.) There are 32 configurations, of which only eight are distinct, and each configuration has 32 subsystems. We define a dissimilarity metric that is zero when the subsystems being compared have, up to rotation and flipping, the same shape and pointwise parity. We now compute the values  $T(E_0, n)$  and  $T(E_1, n)$  by applying the evolution rule. The results are seen in Fig. 2. Note that  $T(E_0, n) = T(E_0, n+2)$  and  $T(E_1, n) = T(E_1, n+6)$ . This cyclic nature of the environment under time development will allow us to find the (overall) self-replicability, rather than just the self-replicability over some constrained time. Next, we calculate the ‘presence’ of the subsystem  $S$  (see Fig. 1b) in

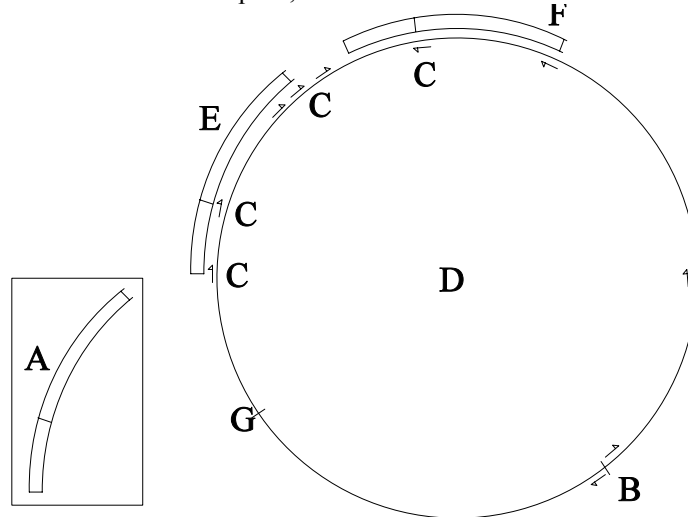
each of  $T(E_{0,l}, t)$ ,  $t \in \{0,1,2,3,4,5\}$ , by counting how many times it occurs in each time step. The results are presented in Table 1. We now take the log of the quotient of the sums, yielding  $\log(7/15) = -0.762$  for the self-replicability over  $t = 0 \dots 5$ .

**Table 1.** Cellular Automaton system: Time and Presence of  $S$  in  $E_0$  and in  $E_l$

<i>Time (t)</i>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Totals</b>
$P(T(E_0,t),S)$	0	5	0	5	0	5	<b>15</b>
$P(T(E_l,t),S)$	1	1	3	2	0	0	<b>8</b>

### 3.2 Ring system

The second system gives an example where the amount of information in the environment grows as the time development function is applied. For conceptualization purposes, the system can be seen as an ideal closed fiber-optic ring (Fig. 3D), with a number of irregularities (Fig. 3B, Fig 3G) that act as beam splitters (Fig. 3B), into which a pattern of moving light, packets (Fig 3C) can be injected. We assume that light packets travel at a constant speed, clockwise or counterclockwise.



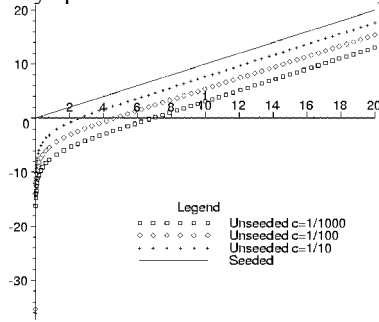
**Fig. 3.** Illustration of example Ring System. A: A system with potential replicability (a pattern to match) B: An irregularity with recently split packets C: optical packets D: center of the ring E: the system (A) acceptably matching a set of packets, F: the system (A) not acceptably matching a set of packets G: location of a second irregularity.

The problem becomes interesting when allowing for beam-splitting irregularities that divide the ring into irrational proportions. In this case, no steady state exists. While any initial distribution should become densely spread over the ring, the distribution is not necessarily uniform: thus, the presence of a given pattern is possibly nontrivial. For this example, a direct computer model was used, storing the initial conditions (initial light configuration and irregularity locations) and applying the time development rules to determine the location, direction, and intensity of all resulting light

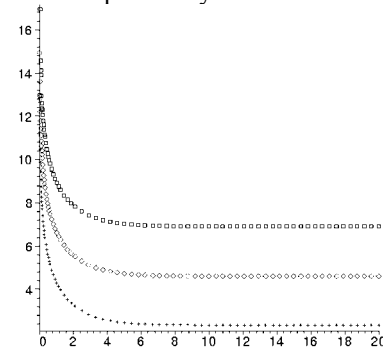
packets. The model was simulated and, in many cases, our hardware was not able to run enough time steps to suggest a long-term trend. However, all  $R_S$  values were no greater than 1 in absolute value. Those values that appeared to converge rapidly had a self-replicability factor between 0.5 and 0.9.

### 3.3 Crystal growth system

The final example is a model of a physical system that is analyzed analytically, rather than explicitly simulated. The goal is to measure the replicability of a crystal satisfying two properties: (a) once a seed crystal is established, the presence of the crystal in the environment is given as a function of time. For this example, we specify an exponential function of time, and (b) there is a fixed nonzero probability rate  $c \in (0, 1]$  at any given time of a seed crystal spontaneously forming. This guarantees that a crystal  $S$  is always possible in an environment  $E$ , and the replicability thus is defined.



**Fig. 4.** Plot of the log of the crystal model presence,  $\log(P_s(T(E_t, t), S))$  vs.  $t$  for different values of  $c$ , the probability of random crystal formation. We use the logarithm to make long-term behavior more visible.



**Fig. 5.** Plot of  $R_T(E_1, E_2, S, \epsilon, 0, t)$  vs.  $t$  for the crystal model, showing the same values of  $c$  as in Fig. 4. Note that  $R_T$  converges quickly (toward  $R_O$ ) as  $t$  becomes large.

In this model, we require that the presence of crystal in the environment increase exponentially from the time a seed crystal is inserted, i.e.  $P_s(T(E_0, t), S) = e^t$ . Thus, during the period while  $e^t < 1$  we compare an environment  $E_t$ , seeded with a crystal  $S$ , to an unseeded, homogeneous environment  $E_0$ . We specify that there is probability 'c' of a seed crystal spontaneously forming, and therefore a probability  $(1 - (1 - c)^t)$  that a seed has formed by time  $t$ . Working out the self-replicability (details omitted due to editorial constraints) we obtain:

$$R_O(S, E_0, E_1, \epsilon) = \lim_{x \rightarrow \infty} \log \left( \frac{e^x - 1}{\int_{t=0}^x \left( \int_{s=0}^t e^{t-s} (1 - (1 - c)^s) ds \right) dt} \right) \approx \log \left( 1 - \frac{1}{\log(1 - c)} \right) \quad (4)$$

So, we have a result that visibly converges in the long run (two examples see in Figs. 4,5), with the magnitude of the replicability being inversely proportional to the probability of seed formation. For example, if the probability of crystal formation  $c$  is at  $c=0.1$ , then  $R_S = 2.35$ , while when  $c$  is much smaller, such as  $c=0.001$ , we have  $R_S=6.88$ , and a high probability, like  $c=0.999$ , yields  $R_S=0.13$ .

## 4 Conclusions

We have proposed a means of measuring the fundamental property of self-replication, seeking a graded measure that can be universally applied to systems from cellular automata to living systems. We provided three examples that involved various mixtures of discrete and continuous variables and were handled both through direct simulation and analytical modeling. This property is not simply seen as intrinsic to a system, but is found in the information that arises from the interaction between a system and its environment.

We currently see two branches of further immediate investigation. On the theoretical side, while a few properties such as additivity in Eq (3) are known, other properties such as continuity of the replicability function warrant further investigation. On the applied side, it would be desirable to find practical ways to calculate replicability in more intricate systems, such as simple biological auto-catalyzing enzymes and cell models, and artificial systems ranging from molecules, to blocks [7] and to machines [8]. Ultimately, we seek to get a better idea of the scale into which interesting replicabilities fall and the conditions under which self-replication is maximized.

## Acknowledgment

This work was supported by the U.S. Department of Energy, grant DE-FG02-01ER45902.

## References

1. Sipper, M., Reggia, J.A., *Go forth and replicate*. In: Scientific American 285/265(2), August 2001, 35-43
2. Nehaniv C., Dautenhahn K., *Self-Replication and Reproduction: Considerations and Obstacles for Rigorous Definitions*. Abstracting and Synthesizing the Principles of Life, Verlag Harri Deutsch, pp. 283-290, 1998.
3. McMullin, B, *John von Neumann and the Evolutionary Growth of Complexity: Looking Backward, Looking Forward* In: Artificial Life 6 (2000) 347-361 Sanchez,
4. Moore, E. F., *Machine Models of Self-Reproduction* in Burks, A. W., *Essays on Cellular Automata* (1970) 187-203
5. Lohn J.D. Reggia J.A. (1997). *Automatic discovery of self-replicating structures in cellular automata*. IEEE Trans. Evolutionary Computation, 1(3):165-178
6. Von Neumann, J, completed and edited by Burks, A. W., *Von Neumann's Self-Reproducing Automata* In: Burks, A. W., *Essays on Cellular Automata* (1970) 4-65
7. Penrose, L.S., *Self-reproducing machines*. In: Scientific American, 200 (6) (1959) 105-114
8. Chirikjian, G.S., Zhou, Y., Suthakorn, J., *Self-replicating Robots for Lunar Development*, In: IEEE/ASME Trans. on Mechatronics 7. 4 (2002) 462-472