An Intergenerational Theory of Evolution

by Dr. Frederick B. Cohen ‡

‡ This research was funded by Management Analytics, PO Box 1480, Hudson, OH 44236, USA
and is extracted from the book "It's Alive" published by John Wiley and Sons of New York, NY USA

1 Background and Introduction

Limits on the rate of evolution, the likelihood of life given a particular environment, the survival of a species during environmental changes, and many other similar issues are open problems that students of the life sciences have been unable to solve. The answers to these questions would go directly to the core of our understanding of life, evolution, environmental impacts of people on the biosphere, the existence of life throughout the universe, and a host of other questions that have been on the minds of people throughout history.

In this paper, we examine Shannon's the well known syntactic information theory in a new light to show that the application of this theory to the communication of genetic information between generations may lead to the mathematial resolution of these historic questions.

We begin with some rudimentary issues regarding reproduction, and use these to lead into the application of Shannon's information theory [?] as a basis for the study of inter-generational communication of genetic information. We then show how this theory may be applied to resolving these well known problems.

In this work, space does not permit the final resolution of these historic issues, but the theoretical basis for their resolution is introduced, and their final resolution is left to other works.

2 Definitions

In this paper, live programs in an environment are defined as information-based entities that reproduce in that environment. This is a linguistic simplification of the formal definition [?] which defines live programs in terms of Turing machines [?], and encompasses the concepts of evolution and the interaction of life forms with their environment.

We describe several examples based on the game called Corewar creatures designed to operate in that environment as described in several issues of Scientific American. [?] We also discuss some results from Thomas Ray's work on Tierra. [?]

3 How Much Randomness?

Perhaps the most fundamental issue in the spontaneous generation of life is reflected in the issue of survival. Clearly life forms can survive indefinitely if they don't alter their environment and their environment doesn't alter them. All of the live programs we have seen [?] [?] are examples of this. In the other extreme, fire, which reproduces, seems to reign when randomness becomes too great. The question is how the level of randomness relates to the survivable life forms.

With no randomness, we see complete stability and essentially no evolution, and with high randomness, we see complete instability and massive evolution without any apparent survival selection other than location. With no randomness, we see minimal energy 'waste' for all energy is dedicated to life, while in high randomness, we see energy consumed at enormous rates, much of it producing electromagnetic emanations (i.e. heat, light, etc.), and with constant destruction of alternative life forms. In between high randomness and low randomness, we seem to find a very rich variety of life forms in which evolution plays a major role and competition forms a feedback system. Some energy is wasted, but much of it is consumed with the processes of life. It would seem there should be an equation relating randomness to life forms and their behavior, and more specifically, the nature of evolution.

4 An Intergenerational Communication Theory

In response to this, we introduce a new application of information theory relating the communication theory of Shannon to the intergenerational communication that is inherent in the reproductive process. Just as there are limits on the ability to communicate in a noisy channel, there are communications limits to the ability to reproduce in a noisy environment. Just as redundancy in communications can increase reliability in exchange for channel bandwidth, we can increase reliability of intergenerational communication by increasing the redundancy in the genetics of the creature, but at the expense of reducing the replication rate.

I don't yet have a mathematical theory to propose other than the one Shannon already proposed for communication. It seems to me that there is no real difference between the communication of genetic information between generations and the communication of other information between communicating parties. There are however some other issues to be considered in that some of the things we typically model in studying life are not commonly modeled in communications theory. For example, sexual reproduction introduces the mixing of two signals during communication. The resulting genetic structure is quite different from either parent, and yet there is clearly communication from both parents to the child. The introduction of parasites yields yet another model of communication wherein the 'transmitter' and/or 'receiver' are part of the communications channel rather than part of the living organism. These problems aside, I believe that researchers will soon show quite clearly that the same conditions of signal-to-noise ratio dictate the survivability of genes as the accurate transmission of signals.

Shannon's information theory considers communication between parties through a channel with noise characteristics. When trying to establish reliable communication, we introduce redundancy to fight off the effects of 'noise'. The amount of noise is related to the amount of 'signal', resulting in a 'signal to noise ratio'. Many results about the relationship between channel 'bandwidth' and signal to noise ratio have been developed, the most famous being the Shannon-Hartley law for channel capacity (C) given additive white Gaussian noise:

$$C = B\ log_2(1 + S/N)$$

where S/N is the signal-to-noise ratio and B is the bandwidth of the channel without noise present.

When we deal with properties of evolution, we might reasonably consider it in terms of a communications system with the communication taking place between generations. That is, when a creature reproduces, there is a communication of genetic information from parent to child. If the signal is not strong enough to overcome the random variations in the environment, the creature will not survive. A creature with enough redundancy should be able to survive any finite amount of noise, even if it reduces the effective bandwidth of intergenerational communication. We should therefore expect that life can exist in almost any environment capable of supporting general purpose computation, regardless of the noise.

Given that we believe this somewhat tenuous and as yet unproven conclusion, we may wish to address the issue of how such life might come to exist. One possibility to consider is that life springs from random noise. Given that there is random noise, there is a finite probability of any possible sequence of symbols occurring naturally. As soon as a live sequence of symbols appears with sufficient redundancy to overcome the signal to noise ratio required for reproduction, we should encounter the spontaneous generation of life.

Now this seems a bit strange at first. We need randomness in order to spontaneously create life, and yet that same noise makes it harder for that life to survive. Fortunately, it doesn't take much noise, given that we have ample time, to generate a living creature. Furthermore, the mere fact that life produces life means that there is a natural redundancy introduced by the very nature of life. Let's consider a simple example using a variation on Corewar.

In this variation, instead of giving two programs turns, we will randomly select a memory location to start operating from, and operate for some randomly selected number of instructions. Furthermore, every randomly selected number of instructions, we will randomly change one bit in the current instruction just prior to execution. If we come upon an illegal instruction, we will simply select the next random starting point.

Without even implementing this system, it should be quite clear that in a relatively short period of time, a version of IMP [**?**] will arise, it will rapidly take over memory, and even the randomness of the system will likely never come up with a more successful variant. One way to see this is to consider the probability of spontaneously generating the different variants and their relative success rate in the environment. We have already shown that with IMP and DWARF in memory, the likelihood is 60% that IMP will dominate if we grant only one step per program. [**?**] The likelihood of spontaneous generation of IMP (given that each corewar instruction contains 8 bits of information) is at best 1 in 2^8, while the likelihood of generating DWARF is at best 1 in 2^{40}. Therefore, we would expect IMP to dominate DWARF more than 99.999999995% of the time!

How often would we expect IMP to fail? It turns out that it's not very often. Even if we introduce a random bit change in 50% of the write operations, it means that a single copy of IMP will, on the average, turn into 2 copies of IMP on it's first execution. (50% chance of writing one copy, 25% chance of writing 2 copies, etc.) When the 1% chance of one of those IMP copies being run occurs, on the average, we will then have 2.5 copies of IMP. (The 1st copy has a 50% chance of keeping the second, a 25% change of creating a third, etc., while the 2nd copy has a 50% chance of creating a third, a 25% chance of creating a fourth, etc.) As we have more copies of IMP, the likelihood of running one of them increases. Clearly, the redundancy created by IMP is the only reason for it's survival in such a noisy environment.

There seems to be an obvious conclusion. Simplicity wins! The reason is that simplicity leads very rapidly to redundancy, which dominates in a small space. The lumbering giant falls every time to the swarm of little bees. But we must be careful here. We have made a great many implicit assumptions through the use of our simplistic model, and these assumptions may be more of the reason for the success of IMP than that simplicity wins. For example, this is a very small space. In a large space, we could easily write a live program with a large lead-in buffer that checks for IMP entering from behind every 10 instructions or so. If it sees it's lead-in area being overwritten, it knows IMP has entered in the last 10 cycles, and for the next so-many cycles does a binary search to locate the exact place IMP is currently executing. It then halts IMP, repairs it's lead-in area, and continues operation. We may even design it with so much redundancy that it can survive a 50% probability of a bit change on each write. We don't know for sure!

5 How Fast can Evolution Go?

Communications theory predicts that, lacking redundancy, and based on stochastic models of noise, etc. it will be impossible to communicate correctly when the noise level is the same as the signal level. Consider now that in Ray's experiments, there was an attempt to increase randomness. Eventually, the level of change was set so high that every replica had at least one alteration from it's parent. The result was disastrous. In effect, no life could be sustained over a substantial number of generations with the noise rate as high as 1 bit per reproduction. Although we could intentionally create living creatures with ample redundancy to survive single bit changes, no creature created without this condition could survive. This appears to confirm our signal to noise theory.

But what about fire? Surely there is so much randomness in fire that there can be no substantial intergenerational communication, and yet clearly fire reproduces essentially unchanged. I have two responses. One is that the conditions that allow fire to live are such that more fire can spontaneously arise when there is enough randomness to have fire in the first place. The second response is that the set of possible live conditions in a system with high randomness are so great that most of the random changes produce.

It would then appear that randomness works in conjunction with the brittleness of the system to produce the communications equations of evolution. In a very brittle system, almost no evolution survives, and the smallest amount of noise prevents the individual from surviving. In Tierra, a fairly small subset of evolutions of non-redundant programs survive (less than 1 in 100 produce grandchildren after a single bit change in producing their children). As the noise rate increases, fewer creatures survive, while lower noise rates tend to produce far slower rates of evolutionary change.

It gets even better! In Tierra, most users start with a fairly large creature that replicates. Over time, the system eventually generates a fairly small (22 symbol) replicating program, and many different forms come into being. But suppose we start with the 22 symbol version instead of the more robust 82 symbol being commonly used? We would expect that the lack of redundancy would yield a more brittle system, and indeed it does! In fact, in only a few generations, this program produces a very small number of variants, and eventually, they all die.

If only relatively redundant living creatures are able to survive some degree of randomness, that seems to imply that spontaneous generation of life with a rich heritage requires a far less likely event that the minimal generation of some living creature. So even though we can easily convince ourselves that some sort of spontaneous generation is within reason, we are still left with the issue of whether such generation will likely yield the rich sort of life forms we find on Earth. In other words, we still can't even say with any degree of certainty that it is reasonably likely that life on earth as we now know it arose from the random generation of a simple life form followed by reproduction, evolution, and eventually the generation of higher forms.

If the information theory of evolution in living systems is right, then we should be able to answer questions like this directly. The brittleness of the environment combined with the degree of randomness set a maximum rate of evolution. There is a tradeoff between evolution and survival. Once we exceed the survival threshold, we will produce a melting population, while backing away by reducing randomness reduces the evolution rate. As worlds cool, they pass through a continuous range of degrees of randomness, leaving many chances for niche life forms to appear.

Tierra is a fixed system, just as our physics is a fixed system, but the creatures in Tierra live in an environment where, except for the interactions between living creatures, there is no effect of the creatures on their environment. Now in some sense, this is true of all life. After all, we all live in a physical world, and the physics of our world dictates how things work, and except for physical objects, the evolution we speak of proceeds without any other interference from the physics.

But there are some differences. We live in a world with 4-dimensions, and Tierra has only two. We have three spatial and one temporal dimension, all potentially infinite, while Tierra has one spatial and one temporal dimension, one of which is finite and bounded. There are non-living things in our world that impact our environment and aren't created by living matter. We have rocks, for example, and water, and the activities of these physical entities impact our lives greatly. We compete not only at the genetic level, but at the body level, and the group level, and the regional level, and the national level, and the continental level. Our world also has a history of change over geologic periods.

The slowing pace and reduced gravity of the moon as it moves further from the Earth over time effects the tides and thus the weather. The temperature change since the Earth was formed has a direct impact on what lives and what dies, and the chemical reactions that take place, and how often they do so. The movement of the continents has impacts on the mixing of species between continents. The enormous asteroid that apparently hit Earth 70 million years ago is now supposed to be the reason for the end of the dinosaurs. The changes in electromagnetic field, volcanic activity, and unlisted other phenomena have dramatic effects, and are not reflected in simplistic models.

6 The Theory of Epidemics

Rather than try to analyze each of these circumstances one at a time, there is a theoretical basis for assessing the survival rates of creatures based only on the rates at which they are reproduced and destroyed. This is of course the theory of epidemics [**?**].

In the theory of epidemics, a statistical approach is used to analyze the movement of hosts between infected and uninfected states. By translating this into birth and death of the infection, we can produce a similar theory of the survival of creatures due to environmental effects. In essence, death occurs when reproduction is blocked by randomness, while birth occurs when reproduction succeeds or random creation occurs.

The theory essentially predicts genetic survival based on the birth rate of the creature (b), and the death rate (d) of the creature. If the birth rate exceeds the death rate, epidemic is reached with a probability of $b/(b+d)$ and the epidemic stabilizes at $b/(b+d)$ of the population infected. In the remaining $d/(b+d)$ cases, the genome will die out without ever reaching a substantial portion of the population. In numerical terms, if we have 3 births for every death, epidemic will occur in 3/4 of the incidents, and on the average, 3/4 of the population will be infected after the situation reaches equilibrium. In the remaining 1/4 of the cases, epidemic will never occur, and the infection will subside after only a small portion of the population is infected and cured.

The overall technique then consists of analyzing the signal rate based on redundancy in the genetic life form, the noise introduced by the environment, and the brittleness of the system containing the life form to changes in the life form. These then lead to the relative birth and death rates of creatures, which in tur leads to the likelihood of survival and relative population of the genome in the population. This analysis can be performed for specific situations to resolve the burning questions described in the introduction.

7 Conclusions and Further Work

This paper has probably introduced more questions than answers, but the one answer it has produced seems to be a vital one. The use of the intergenerational communication theory to mathematical questions of life formation, survival, and evolution may lead quickly to the solutions to many of the burning questions in this field. In the domain of artificial life, it is easy to test this theory because it predicts very clearly the likelihood and extent of success of genomes given the nature of the genome, the brittleness of the environment, and the noise characteristics of the environment.

Further work is required to verify this theory by experiment and to apply it to resolving many specific questions including those outlined in the introduction of this paper.

References

[1] F. Cohen, "It's Alive", John Wiley and Sons, New York.

[2] F. Cohen, "A Short Course on Computer Viruses", John Wiley and Sons, New York.

[3] C. Shannon, "A Mathematical Theory of Communications", Bell Systems Technical Journal, 1948.

[4] A. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem", London Math Soc Ser 2, 1936.

[5] A. Dewdney, "Computer Recreations", May 1984, Scientific American (pp 14-22) and in several following issues through 1984

[6] T. Ray, "", 3rd Conference on Artificial Life

[7] N. Baily, "The Mathematical Theory of Epidemics", Hafner Publishing Co., 1957, New York, NY.