Bio-inspired computation

Biologically inspired computing (also bio-inspired computing) is a field of study that loosely knits together subfields related to the topics of connectionism, social behavior and emergence. It is often closely related to the field of artificial intelligence, as many of its pursuits can be linked to machine learning. It relies heavily on the fields of biology, computer science and mathematics. Biologically inspired computing is a major subset of natural computation. The field of biocomputation has a twofold definition: the use of biology or biological processes as metaphor, inspiration, or enabler in developing new computing technologies and new areas of computer science; and conversely, the use of information science concepts and tools to explore biology from a different theoretical perspective. In addition to its potential applications, such as DNA computation, nanofabrication, storage devices, sensing, and health care, biocomputation also has implications for basic scientific research. It can provide biologists, for example, with an IT-oriented paradigm for looking at how cells compute or process information, or help computer scientists construct algorithms based on natural systems, such as evolutionary and genetic algorithms. Biocomputing has the potential to be a very powerful tool.

The domain of bio-inspired computing is gradually getting prominence in the current times. As organizations and societies are gearing towards a digital era, there has been an explosion of data. This explosion of data is making it more and more challenging to extract meaningful information and gather knowledge by using standard algorithms, due to the increasing complexity of analysis. Finding the best solution increasingly becomes very difficult to identify, if not impossible, due to the very large and dynamic scope of solutions and complexity of computations. Often, the optimal solution for such a NP hard problem is a point in the n-dimensional hyperspace and identifying the solution is computationally very expensive or even not feasible in limited time. Therefore intelligent approaches are needed to identify suitable working solutions.

In this context, intelligent meta-heuristics algorithms can learn and provide a suitable working solution to very complex problems. Within meta-heuristics, bio-inspired computing is gradually gaining prominence since these algorithms are intelligent, can learn and adapt like biological organisms. These algorithms are drawing attention from the scientific community due to the increasing complexity of the problems, increasing range of potential solutions in multi-dimensional hyper-planes, dynamic nature of the problems and constraints, and challenges of incomplete, probabilistic and imperfect information for decision making. However, the fast developments in this domain is increasingly getting difficult to track, due to different algorithms which are being introduced very frequently. However, no study has attempted to identify these algorithms exhaustively, explore and compare their potential scope across different problem contexts.

In fact very few researchers are often familiar with the developments in the domain, where more and more new algorithms are gaining acceptance and prominence. Therefore, with limited visibility across algorithms, new researchers working in this domain tend to focus on very limited and popular approaches, and therefore often “force-fit” algorithms rather than exploring the most suitable one, based on the problem statement, due to limited awareness. To address this gap, we review some of the popularly used bio-inspired algorithms as well as introduce the newly developed algorithms which have a huge potential for applications. Further to that, we also explore the potential scope of applications of the algorithms in specific domains, based on published scientific literature. While twelve of the slightly popular algorithms have been discussed, the scope of future research in other bioinspired algorithms has been discussed. However, in depth discussion about the implementation (e.g. pseudocode, etc) and enhancements in each algorithm is beyond the scope of the current article. Further, specific detailed citations of each application could not be provided, but we attempt to generalize whenever possible based on other focused reviews.

I-485/H-400 ***Biologically-inspired computing*****,** Lecture Notes
Spring 2015, Indiana University, By Luis M. Rocha

1. What is Life?

> "*What was life? No one knew. It was undoubtedly aware of itself, so soon as it was life; but it did not know what it was*". Thomas Mann [1924]

Threshold of Complexity

> "*Seeking a connecting link, they had condescended to the preposterous assumption of structureless living matter, unorganized organisms, which darted together of themselves in the albumen solution, like crystals in their mother-liquor; yet organic differentiation still remained at once condition and expression of all life. One could point to no form of life that did not owe its existence to procreation by parents*". Thomas Mann [1924].
>
> "*Nothing in biology makes sense without evolution*". Theodosius Dobzhansky [1973]

Biologically-inspired computing is an interdisciplinary field that formalizes processes observed in living systems to design computational methods for solving complex problems, or simply to endow artificial systems with more natural traits. But to draw more than superficial inspiration from Biology we need to understand and discuss the concept of *life*. It should be noted that for the most part of the history of humanity, the question of what *life* is was not an important issue. Before the study of mechanics became important, everything was thought to be alive: the stars, the skies, the rivers and mountains, etc. There was no non-life, so the concept was of no importance. It was only when people started to see the World as determined by the laws of mechanics that the question arose. If all matter follows simple physical laws, then what is indeed the difference between life and non-life, between biology and physics? Let us then start with a current dictionary definition:

> "life adj.— n.1. the general condition that distinguishes organisms from inorganic objects and dead organisms, being manifested by growth through metabolism, a means of reproduction, and internal regulation in response to the environment. 2. the animate existence or period of animate existence of an individual. 3. a corresponding state, existence, or principle of existence conceived of as belonging to the soul. 4. the general or universal condition of human existence. 5. any specified period of animate existence. 6. the period of existence, activity, or effectiveness of something inanimate, as a machine, lease, or play. 7. animation; liveliness; spirit: The party was full of life. 8. the force that makes or keeps something alive; the vivifying or quickening principle." [Random House Webster's Dictionary]

The definitions above fall into three main categories: (1) life as an organization distinct from inorganic matter (with an associated list of properties), (2) life as a certain kind of *animated behavior*, and (3) life as a special, incommensurable, quality—*vitalism*. Throughout this course we will see that all principles, and indeed all controversies, associated with the study of life fall into one of these categories or the differences among them. The third category has been discarded as a viable scientific explanation, because for science nothing is in principle incommensurable. The question of whether life is organized according to a special design, intelligent or mysterious, pertains to metaphysics. If the agent of design cannot be observed with physical means, then it is by definition beyond the scope of science as it cannot be measured, and any theories derived from such a concept cannot tested.

While metaphysical dispositions do not pertain to science, many scientists have observed that a naive mechanistic decomposition of life may also fail to explain it. The traditional scientific approach has lead the study of living systems into a reductionist search for answers in the nitty-gritty of the biochemistry of living organisms. This alternative sees life as nothing more than the complicated physics of a collection of moving bodies. However, the question remains unanswered since there are many ways to obtain some complicated dynamics, but of all of these, which ones can be classified as *alive*? What *kind of complexity* are we looking for? No one disputes that life is some sort of complex material arrangement, but when do we reach a necessary *threshold of complexity* after which matter is said to be living? Is it a discrete step, or is life a fuzzy concept? To understand it without meaningless reduction, must we synthesize organizations with the same threshold of complexity (first category above), or is it enough to simulate its animated behavior (second category above)?

Information Organizes and Breeds Life

> "*Life is a dynamic state of matter organized by information*". Manfred Eigen [1992]

> "*Life is a complex system for information storage and processing*". Minoru Kanehisa [2000]

Traditionally life has been identified with material organizations which observe certain lists of properties, e.g. *metabolism*, *adaptability*, *self-maintenance* (autonomy), *self-repair*, *growth*, *replication*, *evolution*, etc. Most living organisms follow these lists, however, there are other material systems which obey only a subset of these rules, e.g. viruses, candle flames, the Earth, certain robots, etc. This often leads to the view that life is at best a fuzzy concept and at worst something we are, subjectively, trained to recognize—life is what we can eat—and is thus not an objective distinction. The modern-day molecular biology view of life, on the other hand, tends to see life as a material organization that if not completely defined by *genomic information*, is at least fully controlled by it. Thus, when Craig Venter's team [Gibson et al, 2010] recently produced a bacteria with a "*prosthetic genome*" [a termed coined by Mark Bedau, see *Nature* | Opinion, 2010] copied from another bacteria but synthesized in the lab, the momentous synthetic biology feat was announced as the creation of the first synthetic or artificial life form.

The artificial life field, whose members tend to follow the fuzzy list of properties conception of life, does not typically recognize Venter's bacteria with a prosthetic genome as a bona fide synthesis of artificial life, since it relies on the pre-existence of a working, naturally-obtained cell to implant a prosthetic genome into. Even most molecular biologists will agree that we are nowhere near understanding, let alone synthesizing an artificial cell from scratch [e.g. George Church, see *Nature* | Opinion, 2010]. Nonetheless, Venter's achievement begs at least the question of what is it about life's *design principle* that makes it easier to synthesize a working prosthetic genome than a working "prosthetic proteome or metabolome"? It also makes us think about what does "understanding life" mean for biology, biomedical technology, artificial life, and informatics? Why is *genetic information* so important and how does it relate to *information technology*?

Life requires the ability to both categorize and control events in its environment in order to survive. In other words, organisms pursue (or even decide upon) different actions according to *information* they perceive in an environment. Furthermore, living organisms reproduce and develop from genetic *information*. More specifically, genetic information is ***transmitted*** "vertically" (inherited) in phylogeny and cell reproduction, and ***expressed*** "horizontally" within a cell in ontogeny for the functioning of living organisms as they interact and react with their environments—we are now sure that genetic information can also be transmitted horizontally between organisms and play an important role in evolution [Goldenfeld & Woese 2007; Riley, 2013]. Indeed, the difference between living and non-living organizations seems to stand on the ability of the former to use *relevant* information for their own functioning. It is this "relevant" which gives life an extra attribute to simple mechanistic interactions. When an organization is able to recognize and act on aspects of its environment which are important *to its own* survival, we say that the mechanisms by which the organization recognizes and acts are *functional* in reference to the organization itself (*self-reference*). Physics is not concerned with function. A physical or chemical description of DNA is certainly possible, but will tell us nothing as to the function of a DNA molecule as a gene containing relevant information for a particular organism. Only in *reference* to an organism does a piece of DNA *function* as a gene (e.g. an enzyme with some effect in an environment).

Thus it is remarkable that in Venter's experiment, a cell with a synthesized prosthetic genome from a similar but distinct organism, was able to reproduce over and over resulting in a cell with a different phenotype from the original, implanted cell—in effect, a cell *re-programmed* by a synthesized genome. Is life then a type of computer that can be reprogrammed? This also leads us to question how general-purpose can such genomic re-programming be? Will it be restricted to very narrow classes of similar organisms, or will it ever be possible to re-program any prokaryotic or eukaryotic cell ?

Emergence and Explanation

> "*First, nothing in biology contradicts the laws of physics and chemistry; any adequate biology must be consonant with the 'basic' sciences. Second, the principles of physics and chemistry are not sufficient to explain complex biological objects because new properties emerge as a result of organization and interaction. These properties can only be understood by the direct study of the whole, living systems in their normal state. Third, the insufficiency of physics and chemistry to encompass life records no mystical addition, no contradiction to the basic sciences, but only reflects the hierarchy of natural objects and the principle of emergent properties at higher levels of organization*". Stephen Jay Gould [1984].

This issue could be rephrased in terms of the notion of *emergence*. Whatever (macro-level) organization exists after the complexity threshold for life is passed, we may say that it is emergent because its attributes cannot be completely explained by the (micro-) physical level. In particular, function, control, and categorization cannot be explained by the mechanics and dynamics of the components of life alone. Notice, however, that emergence does not imply vitalism or dualism. When we say that certain characteristics of life cannot be explained by physics alone, we mean that they must be explained by different, additional models—namely, *informational*, *historical* and *functional* descriptions. In other words, though biological function, control, and categorization cannot be explained by physics alone, organisms, like anything else, must nonetheless follow physical laws. But information is contextual, and therefore requires more than universal models: it requires contingent, context-specific descriptions. In particular, the *origin of life*, is a problem of emergence of information from a physical milieu under specific constraints [Eigen, 1992]. This is the crux of complex systems: the interplay between micro- and macro-level descriptions determines their behavior, and both levels (emergence) are required to understand complexity.

The definition of emergence as an epistemological, explanatory requirement, is related to the notion of *emergence-relative-to-a-model* [Rosen, 1985; Cariani, 1989] or *intensional emergence* [Salthe 1991]. It refers to the impossibility of epistemological reduction of the properties of a system to its components [Clark, 1996]. As an example, we can think of phase transitions such as that of water in its transition from liquid to gas. Water and its properties cannot be rephrased it terms of the properties of hydrogen and oxygen, it needs a qualitatively different model. Another example of complementary models of the same material systems is the wave-particle duality of light.

Physicists *understand* the laws of nature (as best they can), but it takes engineers to *control* nature. The very best physicists are the very best engineers, but those are exceedingly rare (e.g. Von Neumann). The goal of complex systems is to understand *organized complexity* (life, society, cognition) in the same way physicists understand nature [Weaver, 1948]. Biology, as a discipline, has not entirely "made up its mind" if it wants to understand life as a physicist or control it as an engineer. Due to its focus on the micro-level of life, its biochemistry, molecular biology follows essentially a (reverse-) engineering, black-box methodology (knockouts, controls, etc.). This leads to a bit of a schizophrenic agenda: focusing exclusively on micro-level experiments in order to *suggest* macro-level understandings. If the goal is control of biology, say for biomedical advances, then we really need to focus on biotechnology engineering. If the goal is understanding, then we need to focus more on macro-level organized complexity. Ideally, a healthy life sciences program would tie the need to understand with the need to control better—like physicists and engineers do.

This is where complex systems, artificial life, and bio-inspired computing can contribute to a wider arena of the life sciences; they can be used as laboratories for experimenting with theories of organized complexity, and thus enrich our understanding of life. Artificial life concerns both the simulation and realization of life in some artificial environment, usually the computer. At least regarding the second of its goals, artificial life aims to understand the fundamental micro/macro-level interaction that leads to organized complexity. Bio-inspired computing, as a more pragmatic endeavor, does not need to concern itself with synthesizing actual life, but only with drawing analogies from life (real and artificial). Nonetheless, if the main motivation of bio-inspired computing is that life with its designs has already solved versions of many complex engineering problems we are interested in, then a thorough and accurate understanding of the essential characteristics of life is inescapable. Moreover, by abstracting context-specific principles of life to make them relevant in other settings, provides a useful laboratory to experiment with theoretical biology.

Further Readings and References:

Cariani, Peter [1989].*On the Design of Devices with Emergent Semantic Functions*. PhD.Dissertation. SUNY Binghamton.

Clark, Andy [1996]. “Happy couplings: emergence and explanatory interlock.” In: *The Philosophy of Artificial Life*. M. Boden (ed.). Oxford University Press, pp. 262-281.

Dobzhansky, T. [1973]. “Nothing in Biology Makes Sense Except in the Light of Evolution”. *The American Biology Teacher*, March 1973 (35:125-129)

Eigen, M. [1992]. *Steps Towards Life*. Oxford University Press.

D. G. Gibson et al [2010]. “Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome”. *Science*. **329** (5987): 52-56

Goldenfeld, Nigel, and Carl Woese [2007]. “Biology’s Next Revolution.” *Nature* **445** (7126): 369.

Gould, Stephen Jay [1984]. *Natural History*; Jan84, Vol. 93 Issue 1, p24.

Mann, T. [1924]. *The Magic Mountain*. As quoted by Eigen [1990].

Nature | Opinion [2010] “Life after the synthetic cell”. *Nature* **465**: 422–424

Pattee, Howard H. [1978]."The complementarity principle in biological and social structures." In: *Journal of Social and Biological Structures* Vol. 1, pp. 191-200.

Polanyi, M. [1968]. “Life’s irreducible structure”. *Science*, **160** (3834), 1308-1312.

Riley, D. R., K.B. Sieber, K. M. Robinson, J. R. White, A. Ganesan, S. Nourbakhsh, and J. C. Dunning Hotopp [2013]. “Bacteria-Human Somatic Cell Lateral Gene Transfer Is Enriched in Cancer Samples.” *PLoS Computational Biology* **9** (6): e1003107.

Salthe, Stanley N. [1991], “Varieties of Emergence”. *World Futures* Vol. 32, pp.69-83

Schrödinger, Erwin [1944]. *What is Life?.* Cambridge University Press.

Weaver, W. [1948]. "Science and Complexity". *American Scientist*, **36**(4): 536-44.

For next lectures read:

Dennet, D.C. [2005]. "Show me the Science". *New York Times*, August 28, 2005.

Gleick, J. [2011]. *The Information: A History, a Theory, a Flood*. Random House. Chapter 8.

Kanehisa, M. [200]. *Post-genome Informatics*. Oxford University Press. **Chapter 1, Blueprint of life**, pp. 1-23.

Langton, C. [1989], “Artificial Life” In *Artificial Life*. C. Langton (Ed.). Addison-Wesley. pp. 1-47.

Nunes de Castro, Leandro [2006]. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. Chapman & Hall. **Chapter 1**, pp. 1-23.

Polt, R. [2012]. "Anything but Human". *New York Times*, August 5, 2012

2. The logical Mechanisms of Life

> "The designs found in nature are nothing short of brilliant, but the process of design that generates them is utterly lacking in intelligence of its own". Daniel Dennett, NY Times 2005

Life-As-It-Could-Be: but, what is *non*-life-as-it-could-be?

> "Artificial Life [AL] is the study of man-made systems that exhibit behaviors characteristic of natural living systems. It complements the traditional biological sciences concerned with the *analysis* of living organisms by attempting to *synthesize* life-like behaviors within computers and other artificial media. By extending the empirical foundation upon which biology is based *beyond* the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be.* [...] [AL] views life as a property of the *organization* of matter, rather than a property of the matter which is so organized. Whereas biology has largely concerned itself with the material basis of life, Artificial Life is concerned with the formal basis of life. [... It] starts at the bottom, viewing an organism as a large population of *simple* machines, and works upwards *synthetically* from there — constructing large aggregates of simple, rule-governed objects which interact with one another nonlinearly in the support of life-like, global dynamics. The 'key' concept in AL is *emergent behavior*." [Langton, 1989, pp 1-2]

> "Artificial Life is concerned with tuning the behaviors of such low-level machines that the behavior that emerges at the global level is essentially the same as some behavior exhibited by a natural living system. [...] Artificial Life is concerned with generating lifelike behavior." [Langton, 1989, pp 4 and 5]

The previous quotes indicate the goals of Artificial Life according to Chris Langton: the search for complex, artificial, systems which instantiate some kind of lifelike *organization*. The field is interested in both synthesizing an actual artificial living organization, as well as simulating lifelike behavior. The first goal is more ambitious and related to the first definition of life introduced in lecture one, while the second goal is related to the second definition. The methodology to reach either of these goals is also in line with the notion of emergence mentioned in lecture one: from the non-linear interaction of simple, mechanistic, components, we wish to observe the emergence of complicated, life-like, unpredictable, behavior. Natural living organisms are likewise composed of non-living components. As pointed out in lecture one, the origin problem in biology is precisely the emergence of life from non-living components. The material components follow, and are completely described, by physical laws, however, a mechanistic explanation of the overall living system is incomplete. Similarly, in Artificial Life, we have formal components obeying a particular set of axioms, and from their interaction, global behavior emerges which is not completely explained by the local formal rules. Clearly, the formal rules play the role of an artificial matter and the global behavior, <u>if recognized as life-like</u>, plays the role of an artificial biology.

> "Of course, the principle assumption made in Artificial Life is that the 'logical form' of an organism can be separated from its material basis of construction, and that 'aliveness' will be found to be a property of the former, not of the latter." [Langton, 1989, page 11]

The idea is that if we are able to find the basic design principles of living organization, then the material substrate used to realize life is irrelevant. By investigating these basic principles we start studying not only biological, carbon-based, life — life-as-we-know-it — but really the universal rules of life, or life-as-it-could-be. Moreover, from a better understanding of the design principles of life, we can use them to solve engineering problems similar to those that living organisms face [Segel and Cohen, 2001; DeCastro and Von

Zuben, 2005]. Several problems have been raised regarding this separation of matter from form, or the search for a universality without matter [Cariani, 1992; Moreno et al, 1994], which will not be discussed here. What needs to be made more explicit is the relationship between the two distinct goals of AL.

Looking at emergent behavior, obtained from formal complex systems, in search of *interesting* behavior leads to a certain circularity. If AL is concerned with finding life-like behavior in artificial, universal, systems, we are ultimately binding life-as-could-be to the behavior of life-as-we-know-it by virtue of some subjective resemblance. This can hardly be accepted as the search for universal principles.

> "They say, 'Look, isn't this reminiscent of a biological or a physical phenomenon!' They jump in right away as if it's a decent model for the phenomenon, and usually of course it's just got some accidental features that make it look like something." [Jack Cowan as quoted in *Scientific American*, June 1995 issue, "From Complexity to Perplexity", by J. Horgan, page 104]
>
> "Artificial Life — and the entire field of complexity—seems to be based on a seductive syllogism: There are simple sets of mathematical rules that when followed by a computer give rise to extremely complicated patterns. The world also contains many extremely complicated patterns. Conclusion: Simple rules underlie many extremely complicated phenomena in the world. With the help of powerful computers, scientists can root those rules out." [J. Horgan, *Scientific American*, June 1995 issue, "From Complexity to Perplexity", page 107]
>
> "Artificial Life is basically a fact-free science". [John Maynard Smith as quoted in *Scientific American*, June 1995 issue, "From Complexity to Perplexity", by J. Horgan, page 107]

The problem is that Artificial Life must be compared to something, otherwise it becomes a factless manipulation of computer rules with subjective resemblances to real life. Again, we are faced with many possible types of emergent complex behaviors, this time formal, but what kinds of behaviors can be classified as "life-as-could-be"? What is the formal *threshold of complexity* needed? In the natural world we are able to distinguish life from non-life, biology from physics due to the known signatures of bio-chemistry. In the logical realm, we likewise need a formal criteria to distinguish logical life from logical non-life, artificial life from artificial physics.

> "Artificial Life must be compared with a real or an artificial nonliving world. Life in an artificial world requires exploring what we mean by an alternative physical or mathematical reality." [Pattee, 1995]

The two goals of AL are usually described as *hard* and *soft AL* respectively. The first concerns the synthesis of artificial life from computational or material (e.g. embodied robotics) components. The second is interested in producing life-like behavior and is essentially metaphorical. To be accepted as a scientific field, Alife cannot settle for subjective rules of what constitutes living behavior. Indeed, whether we want to synthesize life or merely simulate a particular behavior of living organisms, we need investigate the rules that allow us to distinguish life from non-life . Only by establishing an artificial physics, from which an artificial biology can emerge, and a *theory*, or set of rules, distinguishing the two, can we aim at a proper science based on fact. In other words, the methodology of Artificial Life requires existing theories of life to be compared against; it can also contribute to the meta-methodology of Biology by allowing us to test and improve its theories beyond the unavoidable material constraints, such as the incomplete fossil record or measurement of cellular activity. Naturally, the requirements for hard AL are much stricter, as we are not merely interested in behaviors that can be compared to real biological systems with looser or stricter rules, but the actual realization of an artificial organization that must be agreed to be living against some theory. Soft AL, may restrict itself to particular behavioral traits which need only to be simulated to a satisfactory degree.

Simulations, Realizations, Systemhood, Thinghood, and Theories of Life

> “Boids are *not* birds; they are not even remotely like birds; they have no cohesive physical structure, but rather exist as information structures — processes — within a computer. But — and this is the critical ‘but’— at the level of behaviors, *flocking Boids and flocking birds are two instances of the same phenomenon*: flocking. [...] The ‘artificial’ in Artificial Life refers to the component parts, not the emergent processes. If the component parts are implemented correctly, the processes they support are genuine — every bit as genuine as the natural processes they imitate. [...] Artificial Life will therefore be genuine life —it will simply be made of different stuff than the life that has evolved on Earth.” [Langton, 1989, pp. 32-33]

> “Simulations and realizations belong to different categories of modeling. Simulations are metaphorical models that symbolically ‘stand for’ something else. Realizations are literal, material models that implement functions. Therefore, accuracy in a simulation need have no relation to quality of function in a realization. Secondly, the criteria for good simulations and realizations of a system depend on our theory of the system. The criteria for good theories depend on more than mimicry, e.g., Turing Tests.” [Pattee, 1989, page 63]

As Pattee points out, the bottom line is that a simulation, no matter how good it is, is not a realization. Nonetheless, it may still be possible to obtain artificial living organisms (realizations) if, from an artificial environment, we are able to generate, in a bottom-up manner, organizations which conform to some theory of life we wish to test. Howard Pattee [1989] has proposed that if emergent artificial organisms are able to perform measurements, or in other words, categorize their (artificial) environment, then they may be considered realizations. Some claim that computational environments do not allow for this creative form of emergence [see Cariani, 1992; Moreno, et all, 1994]. In any case, whatever artificial environment we may use, computational or material, we need a theory allowing us to distinguish life from non-life.

Related to this issue, and in the context of complex systems science, is the search of those properties of the world which can be abstracted from their specific material substrate: systemhood from thinghood. Systems science is concerned with the study of *systemhood* properties, but there may be systems from which systemhood cannot be completely abstracted from *thinghood*. Life is sometimes proposed as one of those systems [see Rosen, 1986, 1991; Moreno et al, 1994; Pattee, 1995]. The difficulty for systems science, or complexity theory, lies precisely in the choice of the appropriate level of abstraction. If we abstract enough, most things will look alike, leading to a theory of factless, reminiscent analogies, exposed by Cowan and Maynard-Smith above. If, on the other hand, we abstract too little, all fields of inquiry tend to fall into increasingly specific niches, accumulating much data and knowledge about (context-specific) components without much *understanding* of, or ability to *control*, the (general) macro-level organization. In the context of life, we do not want to be tied uniquely to carbon-based life, or *life-as-we-know-it*, but we also do not want *life-as-could-be* to be anything at all. The challenge lies precisely on finding the right amounts of systemhood and thinghood, as well as the interactions between the two, necessary for a good theory of life, real or artificial.

Further Readings and References

Cariani, P. [1992], “Emergence and Artificial Life” In *Artificial Life II*. C. Langton (Ed.). Addison-Wesley. pp. 775-797.

de Castro, L. N. & Von Zuben, F. J. (eds.) [2005]. *Recent Developments in Biologically Inspired Computing*. Idea Group Publishing.

Dennet, D.C. [2005]. "Show me the Science". *New York Times*, August 28, 2005

Langton, C. [1989], “Artificial Life” In *Artificial Life*. C. Langton (Ed.). Addison-Wesley. pp. 1-47.

Klir, G. [1991], *Facets of Systems Science*. Plenum Press. (On Constructivism pp. 12-13)

Moreno, A., A. Etxeberria, and J. Umerez [1994], “Universality Without Matter?”. In *Artificial Life IV*, R. Brooks and P. Maes (Eds). MIT Press. pp 406-410

Pattee, H. [1989], "Simulations, Realizations, and Theories of Life". In *Artificial Life*. C. Langton (Ed.). Addison-Wesley. pp. 63-77.

Pattee, H. [1995], "Artificial Life needs a real Epistemology". In Advances in Artificial Life. F. Moran, A Moreno, J.J. Merelo, P. Chacon (Eds.). Springer-Verlag.

Rosen, R. [1986], "Some Comments on Systems and System Theory". In *Int. Journal of General Systems*. Vol. 13, No.1.

Rosen, R. [1991]. *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life.* Columbia University Press.

Segel, L.A. and I.C. Cohen [2001]. *Design Principles for the Immune System and Other Distributed Autonomous Systems*. Santa Fe Institute Studies in the Sciences of Complexity. Oxford University Press.

For next lecture read:

Nunes de Castro, Leandro [2006]. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. Chapman & Hall. **Chapter 7, sections 7.1, 7.2 and 7.4**.

Optional: Part I of Flake's [1998], *The Computational Beauty of Life*. MIT Press.

3. Formalizing and Modeling the World

> "*When you can measure what you are speaking of and express it in numbers you know that on which you are discoursing. But if you cannot measure it and express it in numbers. your knowledge is of a very meagre and unsatisfactory kind.*". (Lord Kelvin)

The Nature of Information and Information Processes in Nature[1]

The word **information** derives from the Latin *informare* (*in* + *formare*), meaning to give form, shape, or character to. Etymologically, it is therefore understood to be the formative principle of something, or to imbue with some specific character or quality. However, for hundreds of years, the word information is used to signify knowledge and aspects of cognition such as meaning, instruction, communication, representation, signs, symbols, etc. This can be clearly appreciated in the *Oxford English Dictionary*, which defines information as "the action of informing; formation or molding of the mind or character, training, instruction, teaching; communication of instructive knowledge".

When we look at the world and study reality, we see order and structure everywhere. There is nothing that escapes description or explanation, even in the natural sciences where phenomena appear sometimes catastrophic, chaotic and stochastic. A good example of order and information are our roads. Information can be delivered by signs. Drivers know that signs are not distant things, but they are *about* distant things in the road. What signs deliver are not things but a sense or knowledge of things – a message. For information to work that way, there have to be signs. These are special objects whose function is to be *about* other objects. The function of signs is reference rather than presence. Thus a system of signs is crucial for information to exist and be useful in a world, particularly for the world of drivers!

The central structure of information is therefore a relation among signs, objects or things, and agents capable of understanding (or decoding) the signs. An AGENT is *informed* by a SIGN about some THING. There are many names for the three parts of this relation. The AGENT can be thought of as the recipient of information, the listener, reader, interpretant, spectator, investigator, computer, cell, etc. The SIGN has been called the signal, symbol, vehicle, or messenger. And the about-some-THING is the message, the meaning, the content, the news, the intelligence, or the information.

The SIGN-THING-AGENT relation is often understood as a sign-system, and the discipline that studies sign systems is known as *Semiotics*. In addition to the triad of a sign-system, a complete understanding of information requires the definition of the relevant context: an AGENT is informed by a SIGN about some THING in a certain CONTEXT. Indeed, (Peircean) semiotics emphasizes the *pragmatics* of sign-systems, in addition to the more well-known dimensions of *syntax* and *semantics*. Therefore, a complete (semiotic) understanding of information studies these three dimensions of sign-systems:

1. ***Semantics***: the content or meaning of the SIGN of a THING for an AGENT; it studies all aspects of the relation between signs and objects for an agent, in other words, the study of meaning.
2. ***Syntax***: the characteristics of signs and symbols devoid of meaning; it studies all aspects of the relation among signs such as their rules of operation, production, storage, and manipulation.
3. ***Pragmatics***: the context of signs and repercussions of sign-systems in an environment; it studies

---

[1] This subsection is an excerpt of [Rocha and Schnell, 2005]

how context influences the interpretation of signs and how well a signs-system represents some aspect of the environment.

Signs carry information content to be delivered to agents. However, it is also useful to understand that some signs are more easily used as referents than others. In the beginning of the 20th century, Charles Sanders Peirce defined a typology of signs:

1. ***Icons*** are direct representations of objects. They are similar to the thing they represent. Examples are pictorial road signs, scale models, and of course the icons on your computer. A footprint on the sand is an icon of a foot.
2. ***Indice***s are indirect representations of objects, but necessarily related. Smoke is an index of fire, the bell is an index of the tolling stroke, and a footprint is an index of a person.
3. ***Symbols*** are *arbitrary* representations of objects, which require exclusively a social convention to be understood. A road sign with a red circle and a white background denotes something which is illegal because we have agreed on its arbitrary meaning. To emphasize the conventional aspect of the semantics of symbols, consider the example of variations in road signs: in the US yellow diamond signs denote cautionary warnings, whereas in Europe a red triangle over a white background is used for the same purpose. We can see that convention establishes a code, agreed by a group of agents, for understanding (decoding) the information contained in symbols. For instance, smoke is an index of fire, but if we agree on an appropriate code (e.g. Morse code) we can use smoke signals to communicate symbolically.

Clearly, signs may have iconic, symbolic and indexical elements. Our alphabet is completely symbolic, as the sound assigned to each letter is purely conventional. But other writing systems such as Egyptian or Mayan hieroglyphs, and some Chinese characters use a combination of phonetic symbols with icons and indices. Our road signs are also a good example of signs with symbolic (numbers, letters and conventional shapes), iconic (representations of people and animals) and indexical (crossing out bars) elements.

Finally, it is important to note that due to the arbitrary nature of convention, symbols can be manipulated without reference to content (syntactically). This feature of symbols is what enables computers to operate. As an example of symbol manipulation without recourse to content, let us re-arrange the letters of a word, say “deal”: dale, adel, dela, lead, adle, etc. We can produce all possible permutations ($4! = 4\times3\times2\times1 = 24$) of the word whether they have meaning or not. After manipulation, we can choose which ones have meaning (in some language), but that process is now a semantic one, whereas symbol manipulation is purely syntactic. All signs rely on a certain amount of convention, as all signs have a pragmatic (social) dimension, but symbols are the only signs which require exclusively a social convention, or code, to be understood.

We are used to think of information as pertaining purely to the human realm. In particular, the use of symbolic information, as in our writing system, is thought of as technology used exclusively by humans. Symbols, we have learned, rely on a code, or convention, between symbols and meanings. Such a conventional relation usually specifies rules created by a human community. But it can have a more general definition:

> “A code can be defined as a set of rules that establish *a correspondence between two independent worlds*”. The Morse code, for example, connects certain combinations of dots and dashes with the letters of the alphabet. The Highway Code is a liaison between illustrated signals and driving behaviours. A language makes words stand for real objects of the physical World.” [Barbieri, 2003, page 94]

We can thus think of a code as a process that implements correspondence rules between two independent worlds (or classes of objects), by ascribing meaning to arbitrary symbols. Therefore, meaning is not a characteristic of the individual symbols but a convention of the collection of producers and recipients of the encoded information.

Interestingly, we can see such processes in Nature, where the producers and recipients are not human. The prime example is the genetic code, which establishes a correspondence between DNA (the symbolic genes which store information) and proteins, the stuff life on Earth is built of. With very small variations, the genetic code is the same for all life forms. In this sense, we can think of the genetic system and cellular reproduction as a symbolic code whose convention is "accepted" by the collection of all life forms.

Other codes exist in Nature, such as signal transduction from the surface of cells to the genetic system, neural information processing, antigen recognition by antibodies in the immune system, etc. We can also think of animal communication mechanisms, such as the ant pheromone trails, bird signals, etc. Unlike the genetic system, however, most information processes in nature are of an analog rather than digital nature. Throughout this course we will discuss several of these natural codes.

Formalizing Knowledge: Uncovering the Design Principles of Nature[2]

Once we create symbols, we can also hypothesize relationships among the symbols which we can later check for consistency with what we really observe in the World. By creating relationships among the symbols of things we observe in the World, we are in effect formalizing our knowledge of the World. By formalizing we mean the creation of *rules*, such as verbal arguments and mathematical equations, which define how our symbols relate to one another. In a formalism, the rules that manipulate symbols are independent of their meaning in the sense that they can be calculated mechanically without worrying what symbols stand for.

It is interesting to note that the ability to abstract characteristics of the world from the world itself took thousands of years to be fully established. Even the concept of number, at first was not dissociated from the items being counted. Indeed, several languages (e.g. Japanese) retain vestiges of this process, as different objects are counted with different variations of names for numbers. Physics was the first science to construct precise formal rules of the things in the world. Aristotle (484-322 BC) was the first to relate symbols more explicitly to the external world and to successively clarify the nature of the symbol-world (symbol-matter) relation. "In his Physics he proposed that the two main factors which determine an object's speed are its weight and the density of the medium through which it travels. More importantly, he recognized that there could be mathematical rules which could describe the relation between an object's weight, the medium's density and the consequent rate of fall." [Cariani, 1989, page 52] The rules he proposed to describe this relations were:

1. *For freely falling or freely rising bodies: speed is proportional to the density of the medium.*
2. *In forced motion: speed is directly proportional to the force applied and inversely proportional to the mass of the body moved*

This was the first time that the relationships between observable quantities were hypothesized and used in calculations. Such a formalization of rules as a hypothesis to be tested is what a model is all about. Knowledge is built upon models such as this that sustain our observations of the World.

---

[2] This subsection is an excerpt of [Rocha and Schnell, 2005b]

> “While these quantities were expressed in terms of numbers, they were still generally regarded as inherent properties of the objects themselves. It was not until Galileo took the interrelationships of the signs themselves as the objects of study that we even see the beginnings of what was to be progressive dissociation of the symbols from the objects represented. Galileo's insight was that the symbols themselves and their interrelations could be studied mathematically quite apart from the relations in the objects that they represented. This process of abstraction was further extended by Newton, who saw that symbols arising from observation […] are distinct from those involved in representing the physical laws which govern the subsequent motion”. [Cariani, 1989, page 52]

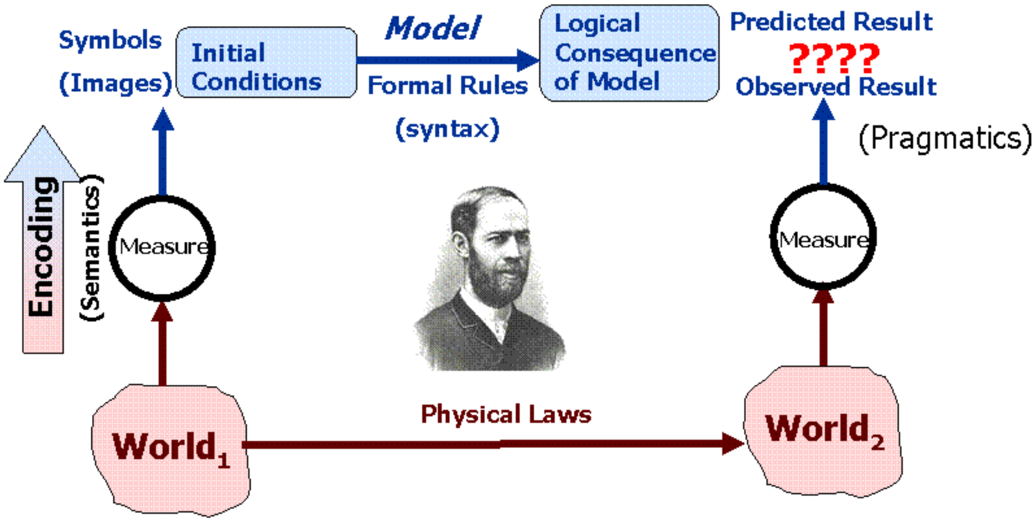


Figure 1: The Modeling Relation between knowledge and reality according to Hertz (adapted from Cariani, 1989)

“In 1894 Heinrich Hertz published *his Principles of Mechanics* which attempted […] to purge mechanics of metaphysical, mystical, undefined, unmeasured entities such as force and to base the theory explicitly on measurable quantities. Hertz wanted to be as clear, rigorous, and concise as possible, so that implicit, and perhaps unnecessary, concepts could be eliminated from physical theories, [which he thought should be based solely on measurable quantities].” [Cariani, 1989, page 54]. Since the results of measurements are symbols, physical theory should be about building relationships among observationally-derived symbols, that is, it should be about building formal *models*, which Hertz called "images”:

> “The most direct and in a sense the most important, problem which our conscious knowledge of nature should enable us to solve is the anticipation of future events, so that we may arrange our present affairs in accordance with such anticipation. As a basis for the solution of this problem we always make use of our knowledge of events which have already occurred, obtained by chance observation or by prearranged experiment. In endeavoring thus to draw inferences as to the future from the past, we always adopt the following process. We form for ourselves images or symbols of external objects; and the form which we give them is such that the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured. In order that this requirement may be satisfied, there must be a certain conformity between nature and our thought. Experience teaches us that the requirement can be satisfied, and hence that such a conformity does in fact exist. When from our accumulated experiences we have succeeded in deducing images of the desired nature, we can then in a short time develop by means of them, as by

> means of models, the consequences in the external world which only arise in a comparatively long time, or as a result of our own interposition. We are thus enabled to be in advance of the facts, and to decide as to present affairs in accordance with the insight so obtained. The images which we here speak are of our conceptions of things. With the things themselves they are in conformity in *one* important respect, namely, in satisfying the above mentioned requirement. For our purpose it is not necessary that they should be in conformity with the things in any other respect whatever. As a matter of fact, we do not know, nor do we have any means of knowing, whether our conceptions of things are in conformity with them in any other than the *one* fundamental respect. [Hertz, 1894 pp. 1-2]"

A model is any complete and consistent set of verbal arguments, mathematical equations or computational rules which is thought to correspond to some other entity, its prototype. The prototype can be a physical, biological, social psychological or other conceptual entity.

The etymological roots of the word model lead us to the Latin word "modulus", which refers to the act of molding, and the Latin word "modus" (a measure) which implies a change of scale in the representation of an entity. The idea of a change of scale, can be interpreted in different ways. As the prototype of a physical, social or natural object, a model represents a change on the scale of abstraction: certain particularities have been removed and simplifications are made to derive a model.

In the natural sciences, models are used as tools for dealing with reality. They are caricatures of the real system specifically build to answer questions about it. By capturing a small number of key elements and leaving out numerous details, models help us to gain a better understanding of reality and the design principles it entails.

Computational Models[3]

> "Insofar as the propositions of mathematics are certain they do not refer to reality; and insofar as they refer to reality, they are not certain". Albert Einstein

Computation is the ultimate abstraction of a formal mathematical system, or an axiomatic system. It is defined by the purely syntactic process of mapping symbols to symbols. Such mapping is the basis of the concept of mathematical function, and it is all that computers do. This abstraction requires that all the procedures to manipulate symbols are defined by unambiguous rules that do not depend on physical implementation, space, time, energy considerations or semantic interpretations given to symbols by observers. Formal computation is, by definition, implementation-independent.

Modeling, however, is not entirely a formal process. The Hertzian modeling paradigm clearly relates formal, computational models to *measurements* of reality against which they must be validated. The measuring process transforms a physical interaction into a symbol – via a measuring device. The measuring process cannot be formalized as it ultimately depends on interacting with a specific (not implementation-independent) portion of reality. We can simulate a measurement process, but for that simulation to be a model we will need in turn to relate it to reality via another measurement. This important aspect of modeling is often forgotten in Artificial Life, when the results of simulations are interpreted without access to real world measurements.

---

[3] This section is indebted to many writings of Howard Pattee, including lecture notes and personal communications.

Likewise, a computer is a physical device that implements a particular abstract computational model as precisely as possible. Modern day computers are so successful because they can implement general-purpose computations almost independently of their specific physics. We do not have to worry about the specific physical architecture of the device as we compute, even though small errors in our computations do occur due to the physical elements of the computing device.

In summary, a computation is a process of rewriting symbol strings in a formal system according to a program of rules. The following characteristics are important: (1) Operations and states are syntactic. (2) Symbols follow syntactical rules. (3) Rate of computation is irrelevant. (4)Program determines result, not speed of machine (Physical implementation is irrelevant).

References

Aris, R. (1978). *Mathematical modelling techniques*. London: Pitman. (Reprinted by Dover).

Barbieri, M. [2003]. *The Organic Codes: An Introduction to Semantic Biology*. Cambridge University Press.

Bossel, J. (1994). *Modeling and simulation*. Wellesley, MA: A K Peters.

Cariani, P. (1989). *On the Design of Devices with emergent Semantic Functions*. PhD Dissertation, SUNY Bionghamton.

Hertz, H. [1894]. *Principles of Mechanics*. tr. D. E. Jones and J. T. Walley. New York: Dover 1956

Lin, C. C. and Segel, L. A. (1974). Mathematics applied to deterministic problems in the natural sciences. New York: Macmillan Publishing. (Reprinted by SIAM).

Polya, G. (1957). *How to solve it: A new aspect of mathematical method*. Pricenton, NJ: Princeton University Press.

Rocha, L.M. and S. Schnell (2005). “The Nature of Information”. Lecture notes for I101 – Introduction to Informatics. School of Informatics, Indiana University.

Rocha, L.M. and S. Schnell (2005b). “Modeling the World”. Lecture notes for I101 – Introduction to Informatics. School of Informatics, Indiana University.

Saari, D. G. (2001). *Chaotic elections! A mathematician looks at voting*. American Mathematical Society. Also you can read: “*Making sense out of consensus*” by Dana MacKenzie, *SIAM News*, Volume 33, Number 8, October 2000.

4. Self-Organization and Emergent Complex Behavior

Self-organization is usually understood as the process by which systems of many components tend to reach a particular state, a set of cycling states, or a small volume of their state space (attractor basins), with no external interference. This *attractor behavior* is often recognized at a different level of observation as the spontaneous formation of well-organized structures, patterns, or behaviors, from random initial conditions (emergent behavior). The systems used to study this behavior are referred to as dynamical systems or *state-determined systems*, since every trajectory is perfectly determined by its initial state. Dynamical systems are traditionally studied by continuous variables and sets of discrete-time difference equations (such as the logistic map) or continuous-time differential equations (such as models of the motion of bodies under gravitational forces). However, self-organization is more easily studied computationally with *discrete dynamical systems* (DDS) such as Boolean networks or cellular automata.

The state-determined transition rules of DDS are interpreted as the laws of some physical system [Langton, 1986] where the state of each component depends on the states of its neighbor (or related) components at the previous time instance. DDS possess a large number of components or variables, and thus very large state spaces. However, when started with random initial conditions (note: not from special initial conditions) they tend to converge, or self-organize, into small sets of attractor states in this space. Attractors may be chaotic in which case the emergent behavior is sensitive to initial conditions. But even chaotic attractors tend to be restricted to small volumes of their state-space (e.g. chaotic in a subset of dimensions of the state-space), therefore we still consider the convergence of a dynamical system into a chaotic basin of attraction to be a form of self-organization.

Since material systems are accurately modeled by dynamical systems, it follows from the observed attractor behavior [Wuensche and Lesser, 1992] of these systems that there is a propensity for matter to self-organize (e.g., [Kauffmann, 1992]). In this sense, matter is described by the (micro-level) dynamics of state transitions and the observed (emergent or macro-level) attractor behavior of self-organization. In general, attractors manifest or *emerge* as global patterns that involve many of components of the dynamical system, and are not easily describable in terms of their state-determined transition rules. For instance, the simple transition rules of the automata in Conway's *Game of Life* cannot describe what the emergent patterns of "blinkers" and "gliders" are. These emergent patterns pertain to a different, complementary level of observation of the same system [Pattee, 1978]. The process of self-organization is often interpreted as the *evolution* of order from random initial conditions. However, notice that this evolution is limited to the specific attractor landscape of a given dynamical system. Unless its parameters are changed (structural perturbation), no dynamical system can escape its own attractor landscape. This limitation will become more apparent when we approach the problem of self-replication.

Life on the Edge of Chaos?

Another interesting aspect of the behavior of dynamical systems concerns the concept of *bifurcation* or *phase transition*. When the parameters of a dynamic system are changed gradually its trajectories and attractors typically change gradually, however, for certain parameter values sudden changes in the dynamic behavior can occur. It is at this critical point that complicated spatio-temporal organization may occur (e.g. from a steady-state to a limit cycle attractor). Close to bifurcations the system also becomes increasingly more sensitive to parameter and initial condition changes. It is often proposed that bifurcations offer a selection mechanism [Prigogine, 1985], as a dynamical system may respond very differently to very small changes in their parameters.

However, if the parameter space is divided by many bifurcations, the system becomes increasingly sensitive to initial conditions and small parameter changes; in this sense its behavior becomes chaotic. It has been argued that the most useful behavior lies instead in between full order and chaos. Langton [1990, 1992] has shown (for one-dimensional cellular automata) that it is in this range of behavior that dynamical systems can carry the most complicated "computations". Computation here is used in a loose sense—not as the rate-independent, symbolic manipulation of Turing-machines—meaning that information exchange between elements of these systems is maximized in this range. In other words, Langton showed that the highest correlation among the automata in a cellular lattice occur at this stage.

Kauffman [1993,] likewise hypothesized that "living systems exist in the [ordered] regime near the edge of chaos, and natural selection achieves and sustains such a poised state". This hypothesis is based on Packard's [1988] work showing that when natural selection algorithms are applied to dynamical systems, with the goal of achieving higher discriminative power, the parameters are changed generally to lead these systems into this transitional area between order and chaos. This idea is very intuitive, since chaotic dynamical systems are too sensitive to parameter changes, that is, a single perturbation or mutation (structural perturbation) leads the system into another completely different behavior (sensitive to damage). By contrast, ordered systems are more resilient to damage, and a small parameter change will usually result in a small behavior change which is ideal for smooth adaptation. However, even though very ordered systems can adapt by accumulation of useful successful variations (because damage does not propagate widely), they may not be able 'step out' of their particular organization in the presence of novel demands in their environment.

It is here that systems at the edge of chaos were thought to enter the scene; they are not as sensitive to damage as chaotic systems, but still they are more sensitive than fully ordered systems. Thus, most mutations cause only minor structural changes and can accumulate, while a few others may cause major changes in the dynamics enabling a few dramatic changes in behavior. These characteristics of simultaneous mutation buffering (to small changes) and dramatic alteration of behavior (in response to larger changes) is ideal for evolvability [Conrad, 1983, 1990]. However, many of the real gene networks that have been successfully modeled with dynamical systems (e.g. the network of segment polarity genes in Drosophila melanogaster [Albert and Othmer, 2003]), exist in a very ordered regime, being very robust to structural changes [Chaves, Albert and Sontag, 2005; Willadsen&Wiles, 2007; kauffman et al, 2003]. Still, other genetic regulatory network models do operate close to criticality [Balleza et al, 2008]. It appears that evolution favors ordered, very robust regimes of self-organization in gene networks – at least the ones involved in very conserved regulatory pathways – though there is also evidence of near-critical regimes for increased evolvability.

Complex Self-organization

We have studied several computational systems said to be self-organizing in the sense described above. The discrete logistic equation observes several ranges of ordered behavior according to its parameter r. For $r \leq 3$, the system converges to a single point steady state (independently of its initial value). For $3 \leq r \leq 4$ the system enters a series of bifurcations, meaning that it changes its attractor behavior, first from a steady-state into a two-state limit cycle, and then progressively doubling the number of states in an attractor limit cycle as r increases. Close to $r = 4$, the limit cycle becomes chaotic. That is, in the chaotic range, the slightest change in the initial value, will lead to a completely different trajectory (though similarly chaotic). The system goes from being independent to strongly dependent of initial conditions, though, in each range, the attractor behavior of the equation is the same for random initial conditions. Thus, we can see the logistic equation as self-organizing.

But there is another aspect of the logistic equation that should be understood. In all of its ranges of behavior, from full order to full chaos, the system is (fairly) reversible. That is, I can always obtain a specific initial condition which caused some behavior, by formally running the system backwards. This means the system is deterministic in both temporal directions. Formally, this means the state transition function is invertible. (This is actually only true, if we decide to work on the lower half of its state space, since the logistic equation is a quadratic function, it has always two possible solutions for the previous value of the current state, these values are symmetric about the middle point of its state space). Some, resist calling this kind of reversible systems self-organizing because they are not sufficiently complex. They reason that if a system is self-organizing, when ran backwards it should be self-*dis*organizing, that is, it should lead to random initial conditions, or to an incomplete knowledge of possible initial states. Indeed, *complexity* is typically equated with the inability to describe the behavior of a system from the behavior of its components or predecessors. This way, we ought to reserve the term self-organization to those irreversible systems whose behaviors must be evaluated statistically. The logistic map shows "hints" of this backwards self-disorganization, but we can still work out effectively its backwards trajectory to an initial condition by restricting the quadratic solutions to half of its state space.

Random Boolean Networks are much more complicated than this. They are completely deterministic since a certain state will always lead to the same next state (state-determinacy), however, we cannot usually know exactly what the predecessor of a current state was. Systems like this are usually studied with statistical tools. Even though the rules that dictate the next state of its components are simple and deterministic, the overall behavior of the system is generally too complicated to predict and statistical analysis has to be performed. For instance, Kauffman has shown that when K=2 (number of inputs to each node), his networks will have on average $\sqrt{N}$ basins of attraction with a length of $\sqrt{N}$ states; if the output of one node is switched to the other boolean value (perturbation), the trajectory returns to that cycle 85% of the time, while on the remaining 15% of the time it will "jump" into a different basin of attraction. *Cellular automata* (CA) fall into this same category of deterministic, irreversible, self-organization.

Further Readings and References:

Albert, R., H.G. Othmer [2003]. "The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster." *J Theor Biol*. 223(1):1-18.

Balleza, E., E. R. Alvarez-Buylla, A. Chaos, S. Kauffman, I. Shmulevich, and M. Aldana. [2008]. "Critical Dynamics in Genetic Regulatory Networks: Examples from Four Kingdoms." *PloS One* **3** (6): e2456.

Chaves, M., R. Albert, and E.D. Sontag [2005]. "Robustness and fragility of Boolean models for genetic regulatory networks."*J Theor Biol*. 235(3):431-49.

Conrad. M. [1983], *Adpatability*. Plenum Press.

Conrad, M. [1990], "The geometry of evolutions". In *BioSystems* Vol. 24, pp. 61-81.

Forrest, S. (Ed.) [1990]. *Emergent Computation*. MIT Press/ North-Holland. Special Issue of Physica D. Vol. 42.

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

Kauffman, Stuart, Carsten Peterson, Björn Samuelsson, and Carl Troein [2003]. "Random Boolean Network Models and the Yeast Transcriptional Network." *Proceedings of the National Academy of Sciences of the United States of America* **100** (25): 14796–14799.

Klir, George J. [1991]. *Facets of Systems Science.* Plenum Press.

Langton, C. [1990]. "Computation at the edge of chaos: phase transitions and emergent computation". In Forrest [1990] pp12-37.

Langton, C. [1992], "Life at the edge of chaos". In *Artificial Life II.* C. Langton (Ed.) Pp 41-91. Addison-Wesley

Packard, N. [1988] "Adaptation toward the edge of chaos". In. *Complexity in Biologic Modelling*. S. Kelso and M. Shlesinger.

Pattee, Howard H. [1978]."The complementarity principle in biological and social structures." In: *Journal of Social and Biological Structures* Vol. 1, pp. 191-200.

Prigogine, I. [1985]. "New Perspectives on Complexity". In: *Sciences and Praxis of Complexity*. pp.107-118.

Willadsen, K. and Wiles, J. [2007] "Robustness and state-space structure of Boolean gene regulatory models". *Journal of Theoretical Biology*, **249** (4), 749-765.

Wuensche, A., & Lesser, M. [1992]. *The global dynamics of cellular automata: An atlas of basin of attraction fields of one-dimensional cellular automata*. Reading, MA: Addison-Wesley.

5. Reality is Stranger than Fiction

Updated from a presentation in the “Biocomplexity” discussion section at the 9th *European Conference on Artificial Life*, September 12, 2007 in Lisbon, Portugal

What can Artificial Life do about Advances in Biology?

> “By extending the empirical foundation upon which biology is based *beyond* the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be”*. [Langton, 1998, page 1]

From Langton’s original artificial life manifesto, the field was largely expected to free us from the confines of “life-as-we-know-it” and its specific biochemistry. The idea of “life-as-it-could-be” gave us a scientific methodology to consider and study the general principles of life at large. The main assumption of the field was that instead of focusing on the carbon-based, living organization, life could be better explained by synthesizing its “logical forms” from simple machines [Langton, 1989, page 11]—where, “fictional” machines substituted real biochemistry. The expectation was that this “out-of-the-box”, synthetic methodology would gain us a wider scientific understanding of life. We would be able to entertain alternative scenarios for life, challenge the dogmas of biology, and ultimately discover the design principles of life.

Interestingly, during the 20 years since the first artificial life workshop, biology witnessed tremendous advances in our understanding of life. True, biology operates at a completely different scale of funding and in a much larger community base than artificial life (the impact factors of key journals in both fields differ by an order of magnitude). But, still, it is from biology, not artificial life, that the strangest and most exciting discoveries and design principles of life arise today. Consider looking at the [September 6, 2007] number of *Nature*, with the quite apropos editorial title “Life as We Know it” [Vol. 449, 1], arguing for a comparative genomics approach, with articles, for instance, moving towards evolutionary principles of gene duplication [Wapinski et al, 2007]. Publications in the [September 2007 issue of] *PLoS. Biol.*, also presented new evidence towards updating or discovering general principles of life: for instance, Venter’s sequencing of his diploid genome, which updates our expectations of differences in chromosome pairs [Levy et al, 2007]), and the Ahituv et al [2007] study that challenges the idea that utraconserved DNA (across species) must be functional. Since then, many advances, often enabled by big data approaches of computational biology, keep being discovered; for instance, from large-scale comparative genomics, it has been found that retroviral genomic sequences account for 6 to 14% of host genomes—~8% of human DNA is from endogenous retroviruses, which comprises more DNA than the human proteome [Weiss & Stoye, 2013].

It is good to notice that this sort of work is not so much an exception, but has been a signature of research in the biosciences in the last couple of decades. Consider cases such as the discovery of DNA transfer from bacteria to the fly [Dunning Hotopp, 2007], extra-genomic inheritance in *Arabidopsis* [Lolle et al, 2005], or the profound importance of non-coding RNA in life which is a major player in, among other features, patterning [Martello et al, 2007] , essential gene regulation [Mattick, 2005], development [Mattick, 2007], epigenetic neural development and modulation [Mehler & Mattick, 2007; Mattick & Mehler, 2008], eukariotic complexity [Taft et al, 2007], etc. Moreover, advances such as these do not seem to be mere epiphenomena of a specific life form. Indeed, they point at important organization principles—as those that artificial life was supposed to provide. When we discover that non-transcribed RNA is involved in extra-genomic inheritance or that most of the evolutionary innovation responsible for differences between marsupials and placental mammals occurs in non-protein coding DNA [Mikkelsen et al, 2007], some fundamental principles of the living organization are to be re-thought: the simple, generalized genotype-phenotype mappings on which most of artificial life is based on, are just not enough to capture the principles of life as we know it. More intricate genomic structure, and its principles, need to be modeled and theories need to be built to understand life.

One could go on and on about many other advances in biology. We can also point to themes at the forefront of (bio)complexity theory that go largely overlooked in artificial life—though not completely (i.e. [Calabretta et al, 2000; Hintze & Adami, 2007]). Perhaps the key topic in complexity theory today is that of modularity in evolution [Schlosser & Wagner, 2004] and in networks [Newman, 2006; Guimerà et al 2007]. Nonetheless, looking at the papers accepted for the main sections of the latest Alife and ECAL conferences, it is easy to see that most papers, not only do not discover or even address such issues, but largely trade in biological and computational concepts that have not changed much since the field's inception (see list of top themes and terms in appendix). Is artificial life trapped in the (evolutionary) biology of twenty years ago? Why is reality stranger and more surprising than fiction?

Clearly, there has been very widely successful artificial life research. First and foremost, artificial life has been most successful as a means to study animal behavior, learning and cognition. Certainly, evolutionary robotics and embodied cognition have had an impact in cognitive science. But is artificial life simply a better way to do artificial intelligence? Moreover, one could argue that given the embodied nature of evolutionary robotics, it would seem that it is bound to some kind of material reality, rather than synthesized by constituent "logical forms" as Langton initially suggested.

But what to do about the organization of life itself? Surely the idea of explaining the living organization was behind the origin of the field. For the purposes of this discussion, we must question ourselves why artificial life does not produce more and surprising results about the living organization? Certainly, there is sound research in the field with impact outside of it [e.g. Adami, 2006; Hintze & Adami, 2007]. But even the most successful research in artificial life rarely goes beyond showing that artificial organisms can observe the same behaviors as their real counterparts (i.e. selective pressures, epistasis, etc.). A problem for the field is that as biotechnology gains more and more control of cellular processes, it is reasonable to ask what can one do with artificial organisms that one cannot do with real bacteria? For instance, recent studies of the evolutionary speed towards beneficial mutations were quite effectively done with E-coli [Perfeito et al, 2007], pointing to a much larger rate of beneficial mutations in bacteria than previously thought, and shedding new light on the general principal of clonal interference.

The point of this short statement is to discuss at this conference [ECAL 2007], how biocomplexity is dealt within artificial life, twenty years after the field's inception. Certainly the community can think of a variety of responses to this lack of new principles of life coming out of research in artificial life—even in theoretical biology. One concept that I venture may need updating in artificial life is its view of the genotype/phenotype relationship. Langton proposed that we generalize this relationship, but this meant that research in the field largely regarded the two as indistinguishable. While this move at fist glance seems appropriate to deal with the complexity of genomic-proteomic interaction, it prevents us from studying the specific roles each plays in the living organization. Genotype and phenotype are intertwined in a complex manner, but each operates under different principles that are often overlooked in artificial life. Thus, artificial life rarely approaches issues of genomic structure and regulation, or the co-existence of DNA and RNA as different types of informational carriers. This could well be because artificial life models seem to trade most often on the concept of Mendelian gene than on the molecular biology gene. In other words, artificial life models tend to regard genes solely as mechanisms of generational (vertical) inheritance, rather than as (informational) mechanisms of ontogenetic (horizontal) development, regulation, maintenance, phenotypic plasticity, and response to environmental change. This way, most artificial life models do not test, or even deal with, possible genomic structure architectures and their impact on development and evolution. This is a big shortcoming in the field since, as we have seen in the last two decades, the molecular biology gene and the genomic structure it implies are behind many essential principles of life—from hypersomatic mutation in vertebrate immunity to speciation.

Additionally, it is most often the case that artificial organisms in artificial life models are designed with many top-down features, rather than emerging out of artificial biochemical machines. For instance, typically the genes of artificial organisms encode pre-defined computer operations. Not only is the encoding pre-defined, but the function of individual genes is also pre-programmed, rather than emergent from some artificial chemistry—what is typically emergent is the behavior of a collection of such "atomic" genes and genotypes.

It is interesting to note that when biologists were looking for the location of genetic information for inheritance, they naturally assumed that it would reside in proteins. They knew of DNA chemically, but its sheer inertness deemed it unfit for the job. It took some time to realize that relative inertness was really the point--- from Griffith's experiment in 1928 to Avery's in 1944, the implications of which were only fully accepted much later , probably costing Avery a deserved Nobel [Judson, 2003]. This episode illustrates how reality very often surprises the best scientific expectations of the day—a big problem for Artificial Life, as long as it defines itself as the study of life-as-it-could-be, since it implies a science built on what scientists *think* life is and not on what experiments show it is. For instance, the biochemical difference between highly inert memory molecules and highly reactive, functional ones, while often overlooked in artificial life as a design principle, is ultimately the hallmark of life [Rocha and Hordijk, 2005; Brenner, 2012]. Indeed, Venter's achievement in successfully replicating a living cell with a "*prosthetic genome*" until the original organism's phenotype is fully re-programmed (see chapter 1), should lead Artificial Life scientists to ponder at least the question of what is it about life's *design principle* that makes it easier to synthesize a working prosthetic genome than a working "prosthetic" proteome or metabolome? Perhaps, Langton's view of artificial life being built-up from simple machines, may have clouded the fact that life as we know it is made of biochemical constituents with very different chemical and functional roles: chiefly, DNA (long-term, random-access memory), RNA (short-term memory and symbol processing) and proteins (functional machines). Perhaps more attention should be directed to the "logical forms" of these lower level, structural constituents that produce life, before we can tackle "life-as-it-could-be".

ACKNOWLEDGMENTS

I am extremely grateful to Karola Stotz at Indiana University for exciting conversations that directly lead to this text. I also thank Janet Wiles for performing the *Leximancer* analysis, and Peter Todd for patiently listening as some of the ideas here discussed were germinating.

Further Readings and References

Ahituv, N., et al. [2007]. "Deletion of Ultraconserved Elements Yields Viable Mice". *PLoS Biology* Vol. 5, No. 9, e234.

Brenner, S. [2012]. "Turing centenary: Life's code script." *Nature* **482** (7386): 461-461.

Calabretta, R., Nolfi, S., Parisi, D. and Wagner, G.P. [2000]. "Duplication of modules facilitates the evolution of functional specialization". *Artificial Life*, 6 , 69-84.

Hintze, A. and C. Adami [2007]. "Evolution of complex modular biological networks". arXiv.org:0705.4674.

Dunning Hotopp, J. C., et al. [2007] "Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes," *Science* doi:10.1126.

Guimerà, R, Sales-Pardo, M, Amaral, LAN. [2007] “Classes of complex networks defined by role-to-role connectivity profiles”. *Nature Phys*. 3, 63-69.

Judson, H.F. [2003]. "No Nobel Prize for Whining". *New York Times*, October 20, 2003.

Levy, S. et al. [2007] *PLoS Biol*. 5, e254.

Lolle, S. J., et al [2005]. “Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis”, *Nature* 434, 505-509.

Mikkelsen, T.S. et al [2007]. “Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences”. *Nature* 447, 167-177. ,

Martello, G. Et al [2007]. “MicroRNA control of Nodal signalling” *Nature*. doi:10.1038/nature06100;

Mattick, J.S. [2005] “The functional genomics of noncoding RNA”. *Science*. 309 (5740): 1570-3.

Mattick JS.[2007]. “A new paradigm for developmental biology”. *J Exp Biol*. ;210(Pt 9):1526-47

Mattick, JS & MF Mehler [2008] “RNA editing, DNA recoding and the evolution of human cognition”. *Trends in Neurosciences*, **31** (5):227-233.

Mehler MF, Mattick JS. [2007]. “Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease”. *Physiol Rev*. 87(3):799-823

Newman, M.E.J. [2006]. “Modularity and community structure in networks”. *PNAS*. 103: 8577.

Perfeito L, L. Fernandes, C. Mota and I. Gordo [2007]. “Adaptive Mutations in Bacteria: High Rate and Small Effects”. *Science*. **317** (5839):813 - 815.

Rocha, L.M. and W. Hordijk [2005]. "Material Representations: From the Genetic Code to the Evolution of Cellular Automata". *Artificial Life*. **11** (1-2):189 - 214

Schlosser, G. and G. P. Wagner (Eds.) [2004]. *Modularity in Development and Evolution*.

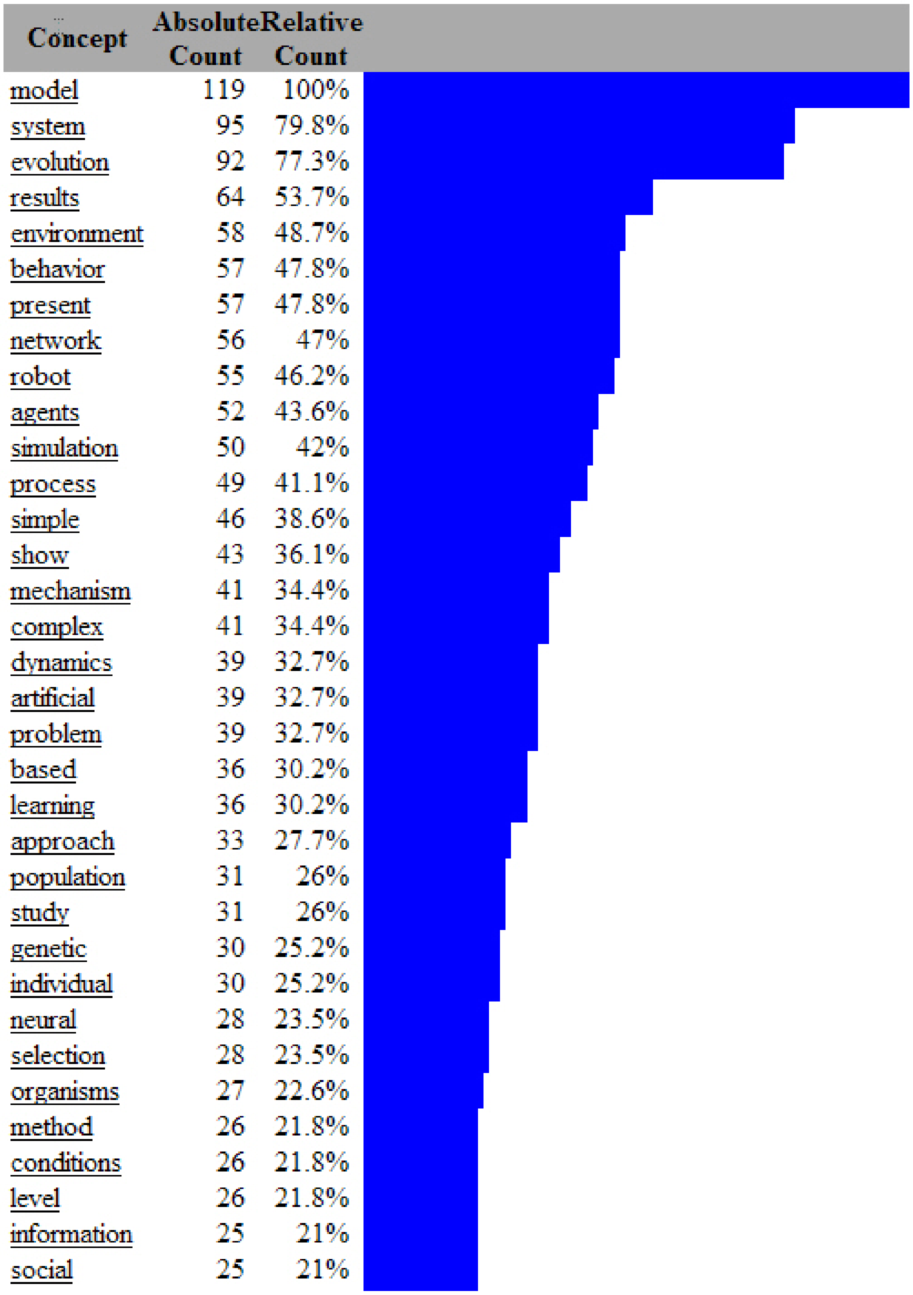
Taft RJ, Pheasant M, Mattick JS. [2007].”The relationship between non-protein-coding DNA and eukaryotic complexity”. *Bioessays*. 29(3):288-99.

Wapinski, I., A. Pfeffer, N. Friedman, and A. Regev [2007]. "Natural history and evolutionary principles of gene duplication in fungi". *Nature* 449, 54-61.1.

Weiss & Stoye [2013]. “Our Viral Inheritance.” *Science*.**340** (6134): 820-821

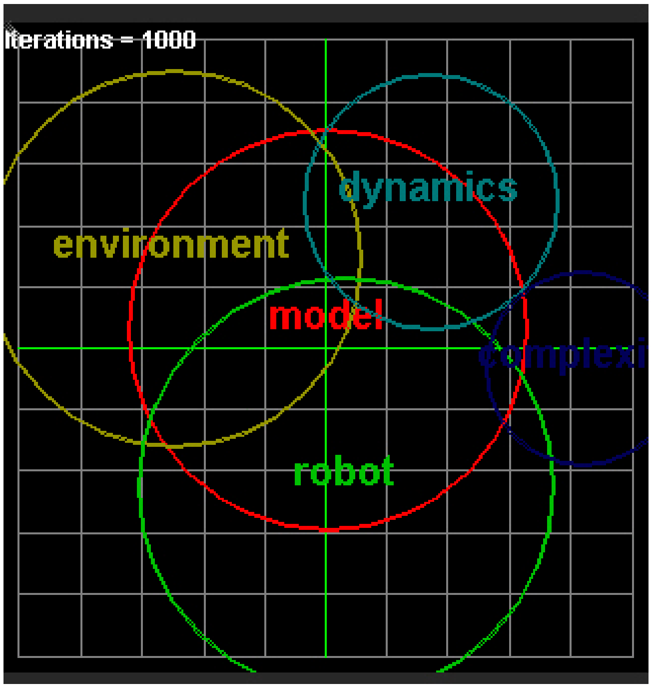
APPENDIX:

Top themes extracted from all abstracts accepted to ECAL 2007, produced the Leximancer (courtesy of Janet Wiles)



Concept	Absolute Count	Relative Count
model	119	100%
system	95	79.8%
evolution	92	77.3%
results	64	53.7%
environment	58	48.7%
behavior	57	47.8%
present	57	47.8%
network	56	47%
robot	55	46.2%
agents	52	43.6%
simulation	50	42%
process	49	41.1%
simple	46	38.6%
show	43	36.1%
mechanism	41	34.4%
complex	41	34.4%
dynamics	39	32.7%
artificial	39	32.7%
problem	39	32.7%
based	36	30.2%
learning	36	30.2%
approach	33	27.7%
population	31	26%
study	31	26%
genetic	30	25.2%
individual	30	25.2%
neural	28	23.5%
selection	28	23.5%
organisms	27	22.6%
method	26	21.8%
conditions	26	21.8%
level	26	21.8%
information	25	21%
social	25	21%

Top Themes produced from Leximancer set at 65% coverage themes (courtesy of Janet Wiles)



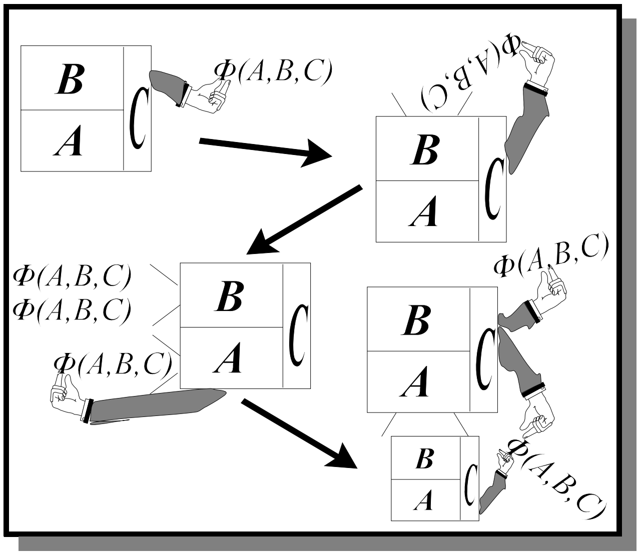
Top co-occurring (stemmed) word pairs in abstracts

neural--network
chang--environ
artifici--life
simul--result
autonom--robot
evolutionari--algorithm
evolutionari--robot
comput--simul
genet--algorithm
robot--mobil
cellular--automata
interact--between
artifici--chemistri
agent--adapt
pressur--select
neural--control

6. Von Neumann and Natural Selection

> “Turing invented the stored-program computer, and von Neumann showed that the description is separate from the universal constructor. This is not trivial. Physicist Erwin Schrödinger confused the program and the constructor in his 1944 book What is Life?, in which he saw chromosomes as “architect's plan and builder's craft in one”. This is wrong. The code script contains only a description of the executive function, not the function itself.” [Brenner, 2012]

6.1 Von Neumann’s Self-Reproduction Scheme



Von Neumann thought of his logical model of self-reproduction as an answer to the observation that, unlike machines, biological organisms have the ability to self-replicate while seemingly increasing their complexity without limit. Mechanical artefacts are instead produced via more complicated factories (as opposed to self-production) and can only degenerate in their complexity. He was searching for a complexity threshold beyond which systems may self-reproduce (with no outside control) while possibly increasing their complexity.

Von Neumann concluded that this threshold entails a memory-stored description Φ(*X*) that can be interpreted by a universal constructor automaton *A* to produce any automaton *X*; if a description of *A*, Φ(*A*), is fed to *A* itself, then a new copy of *A* is obtained. However, to avoid a logical paradox of self-reference, the description, which cannot describe itself, must be both copied (uninterpreted role) and translated (interpreted role) into the described automaton. This way, in addition to the universal constructor, an automaton *B* capable of copying any description, Φ(*X*), is included in the self-replication scheme. A third automaton *C* is also included to perform all the manipulation of descriptions necessary—a sort of operating system. To sum it up, the self-replicating system contains the set of automata (*A* + *B* + *C*) and a description Φ(*A* + *B* + *C*); the description is fed to B which copies it three times (assuming destruction of the original); one of these copies is then fed to *A* which produces another automaton (*A* + *B* + *C*); the second copy is then handled separately to the new automaton which together with this description is also able to self-reproduce; the third copy is kept so that the self-reproducing capability may be maintained (it is also assumed that *A* destroys utilized descriptions). Notice that the description, or program, is used in two different ways: it is both *translated* and *copied*. In the first role, it controls the construction of an automaton by causing a sequence of activities (*active* role of description). In the second role, it is simply copied (*passive* role of description). In other words, the *interpreted* description controls construction, and the *uninterpreted* description is copied separately, passing along its stored information (memory) to the next generation. This parallels the horizontal and vertical transmission of genetic information in biological organisms, which is all the more remarkable since Von Neumann proposed this scheme before the structure of the DNA molecule was uncovered by Watson and Crick[1953]—though after the Avery-MacLeod-McCarty [1944] experiment which identified DNA has the carrier of genetic information.

> "The concept of the gene as a symbolic representation of the organism–a code script–is a fundamental feature of the living world and must form the kernel of biological theory." [Brenner, 2012]

The notion of description-based self-reproduction implies a language. A description must be cast on some *symbol system* while it must also be implemented by some physical or a logical structure. When A interprets a description to construct some automaton, a *semantic* code is utilized to map instructions into construction commands to be performed. When B copies a description, only its *syntactic* aspects are replicated. Now, the language of this semantic code presupposes a set of primitives (e.g. parts and processes) for which the instructions are said to "stand for". Descriptions are not universal insofar as they refer to these building blocks which cannot be changed without altering the significance of the descriptions. The building blocks ultimately produce the dynamics, behavior, and/or functionality of the overall system, and may be material or computational. In Biology, we can think of the genetic code as instantiating such a language. Genes are descriptions that encode specific parts: amino acids chains. In a computational setting, parts are typically logical operations, but they can also be, for example, the building blocks of neural networks coded by genetic algorithms and L-Systems. Von Neumann [1966] (posthumously aided by Arthur Burks) produced a specification of a universal constructor using a 29-state cellular automaton. Implementations of this automaton appeared only fairly recently [e.g. Pesavento, 1995, see Sipper, 1998]

6.2 Open-ended evolution and natural selection

> "Biologists ask only three questions of a living organism: how does it work? How is it built? And how did it get that way? They are problems embodied in the classical fields of physiology, embryology and evolution. And at the core of everything are the tapes containing the descriptions to build these special Turing machines." [Brenner, 2012]

Perhaps the most important consequence of *separate* descriptions in Von Neumann's self-reproduction scheme (and Turing's Tape) is its opening the possibility for open-ended evolution [Rocha, 1998; McMullin, 2000]. As Von Neumann [1966] discussed, if the description of the self-reproducing automata is changed (mutated), in a way as to not affect the basic functioning of $(A + B + C)$ then, the new automaton $(A + B + C)`$ will be slightly different from its parent. Von Neumann used a new automaton $D$ to be included in the self-replicating organism, whose function does not disturb the basic performance of $(A + B + C)$; if there is a mutation in the $D$ part of the description, say $D`$, then the system $(A + B + C + D) + \Phi(A + B + C + D`)$ will produce $(A + B + C + D`) + \Phi(A + B + C + D`)$. Von Neumann [1966, page 86] further proposed that non-trivial self-reproduction should include this "ability to undergo inheritable mutations as well as the ability to make another organism like the original", to distinguish it from "naive" self-reproduction like growing crystals.

Notice that changes in $(A + B + C + D)$ are not heritable, only changes in the description, $\Phi(A + B + C + D)$, are inherited by the automaton's offspring and are thus relevant for evolution. This ability to transmit mutations (vertically) is precisely at the core of the principle of natural selection of modern Darwinism. Through variation (mutation) populations of different organisms are produced; the statistical bias these mutations impose on reproduction rates of organisms will create survival differentials (fitness) on the population which define natural selection. In principle, if the language of description is rich enough, an endless variety of organisms can be evolved: *open-ended evolution*.

The evolvability of a self-reproducing system is dependent on the parts used by the semantic code. If the parts are very simple, then the descriptions will have to be very complicated, whereas if the parts possess rich dynamic properties, the descriptions can be simpler since they will take for granted a lot of the dynamics that otherwise would have to be specified. In the genetic system, genes do not have to specify the functional characteristics of the proteins produced, but simply the string of amino acids that will produce that functionality "for free" [Moreno et al, 1994]. Furthermore, there is a trade-off between programmability and evolvability [Conrad, 1983, 1990] which grants some self-reproducing systems no evolutionary potential whatsoever. When descriptions require high programmability they will be very sensitive to damage. Low programmability grants self-reproducing systems the ability to change without destroying their own organization, though it also reduces the space of possible evolvable configurations [Rocha, 2001].

Turing and Von Neumann were the first to correctly formalize the required inheritance mechanism behind neo-Darwinian evolution by Natural Selection. This understanding of the most fundamental design principle of life, puts Turing and Von Neumann on the Parthenon of great thinkers in Biology, alongside Darwin and Mendel. The dovetailing of computational thinking and biology, inherent in the cybernetics movement of Turing, Von Neumann, Shannon, Wiener and others, emphasizes how (material) control of symbolic information is the hallmark of both computation and biocomplexity.

Further Readings and References:

Avery, Oswald T.; Colin M. MacLeod, Maclyn McCarty [1944]. "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III". *Journal of Experimental Medicine* **79** (2): 137–158.

Brenner, S. [2012]. "Turing centenary: Life's code script." *Nature* **482** (7386): 461-461.

Conrad. M. [1983], *Adpatability*. Plenum Press.

Conrad, M. [1990], "The geometry of evolutions". In *BioSystems* Vol. 24, pp. 61-81.

McMullin, B. [2000]. "John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards". *Artificial Life* 6(4):347-361.

Moreno, A., A. Etxeberria, and J. Umerez [1994], "Universality Without Matter?". In *Artificial Life IV*, R. Brooks and P. Maes (Eds). MIT Press. pp 406-410

Pesavento, U. [1995] An implementation of von Neumann's self-reproducing machine. *Artificial Life* 2(4):337-354.

Rocha, L.M. [1998]."Selected Self-Organization and the Semiotics of Evolutionary Systems". In: *Evolutionary Systems: The Biological and Epistemological Perspectives on Selection and Self- Organization*, . S. Salthe, G. Van de Vijver, and M. Delpos (eds.). Kluwer, pp. 341-358.[2]

Rocha, L.M. [2001]. "Evolution with material symbol systems". *Biosystems.* 60, pp. 95-121

Sipper, M. [1998]. "Fifty Years of Research on Self-replication: an Overview". *Artificial Life*, 4 (3), pp. 237-257.

von Neumann, John [1966]. *The Theory of Self-Reproducing Automata.* Arthur Burks (Ed.) University of Illinois Press.

Watson Jd, Crick Fh [1953]. "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". *Nature* **171** (4356): 737–8.

7. Modeling Evolution: Evolutionary Computation

> "How does evolution produce increasingly fit organisms in environments which are highly uncertain for individual organisms? How does an organism use its experience to modify its behavior in beneficial ways (i.e. how does it learn or 'adapt under sensory guidance')? How can computers be programmed so that problem-solving capabilities are built up by specifying '*what* is to be done' rather than '*how* to do it'?" [Holland, 1975, page 1]

These were some of the questions concerning John Holland when he thought of Genetic Algorithms (GA's) in the 1960's. All these questions were shown to be reducible to a problem of optimizing multi-parameter functions. Nature's "problem" is to create organisms that reproduce more (are more fit) in a particular environment: the environment-organism coupling dictates the selective pressures, and the solutions to these pressures are organisms themselves. In the language of optimization, the solutions to a particular problem (say, an engineering problem), will be selected according to how well they solve that problem. GA's are inspired by natural selection as the solutions to our problem are not algebraically calculated, but rather found by a population of solution alternatives which is altered in each time step of the algorithm in order to increase the probability of having better solutions in the population. In other words, GA's, or other Evolutionary Strategies (ES) such as Evolutionary Programming (EP), explore the multi-parameter space of solution alternatives for a particular problem, by means of a population of encoded strings (standing for alternatives) which undergo variation (crossover and mutation) and are reproduced in a way as to lead the population to ever more promising regions of this search space (selection) [Goldeberg, 1989; Mitchell, 1999; De Jong, 2006].

7.1 Evolutionary Strategies and Self-Organization

The underlying idea of computational ES is the *separation* of solutions for a particular problem (e.g. a machine) from *descriptions* of those solutions (memory). GA's work on these descriptions and not on the solutions themselves, that is, variation is applied to descriptions, while the respective solutions are evaluated, and the whole (description-solution) selected according to this evaluation. Such machine/description separation follows von Neumann's self-reproducing scheme (see chapter 6) which is able to increase the complexity of the (organization of) machines described. Therefore, the form of organization evolved by GA's is not self-organizing in the sense of a boolean network or cellular automata (see chapter 4). Even though the solutions are obtained from the interaction of a population of elements, and in this sense following the general rules usually observed by computationally emergent systems (e.g. Langton [1988], Mitchell [1992]), they do not ***self***-organize since they rely on the *selective* pressures of some environment (in ES, defined by an explicit or implicit fitness function). The order so attained is not a result of the internal dynamics of a collection of interacting elements, but is instead dictated by the *external* selection criteria. In this sense, ES follow an organizing scheme that is driven by external selection of encoded symbolic descriptions (a "Turing tape"). It is perhaps useful to think that ES are modeling the most fundamental design principle of biological systems: natural selection. While self-organizing systems model the dynamical characteristics of matter, ES model the existence of, external, selective pressures on populations of symbolic descriptions of some system. While self-organization models material dynamics, ES models the selection of information about dynamics.

7.2 Development and morphogenesis: self-organization and selection come together

Since the original introduction of GA's, many subsequent developments had to do with the inclusion of a developmental stage, or intermediate layers between genotype and phenotype; in other words, the creation of some artificial morphogenesis or regulation. The idea has been to encode rules that will themselves self-organize to produce a phenotype, rather than the direct encoding of the phenotype itself, or the introduction of gene regulation . As discussed in class, these rules often use L-System grammars which dictate production system programs [Wilson, 1988] leading to some phenotype. The most important advantage of this intermediate stage, as explored by Kitano [1990], Gruau [1993], Belew [1992] and others, is the ability to code for much larger structures than a direct encoding allows. In practical terms, they have solved some of the scalability problems of encoding (e.g.) neural networks in GA's, by reducing the search space dramatically.

L-system grammars are higher-level descriptions of self-organizing developmental processes. However, these first approaches used solely context-free, state-determined, L-System grammars, compromising epistasis (or mutual, non-linear, influence of genetic descriptions amongst each other) in the simulation of self-organizing development. Dellaert and Beer [1994] and Kitano [1994], for instance, used Boolean networks to simulate genetic epistasis and self-organization. In other words, the GA encodes rules which construct Boolean networks whose nodes stand for aspects of the phenotypes we wish to evolve on some physical simulation. In Dellaert and Beer's model, the nodes stand for cell mitosis and other characteristics. This way, the solutions of the GA are self-organizing systems whose attractor behavior dictates pre-defined phenotypic traits.

These approaches in effect offer an emergent morphology, that is, they encode rules which will themselves self-organize into some phenotype (instead of strict programming of morphology). The indirect encoding further allows the search to occur in a reduced space, amplified through development. An interesting side effect has to do with the appearance of modularity traits on the evolved phenotypes [Wagner, 1995]. Subsequent developments paid even more attention to the contextual regulation that indirect encodings afford to the search [Rocha 1995, 1997]. More recently, given our expanded view of genomics, other intermediate layers between genotype and phenotype have been explored, such as transcription regulation [Reil, 1999; Hallinan & Wiles, 2004] and RNA Editing [Rocha et al, 2006]. The inclusion of more sophisticated regulation of genetic information prior to translation, while not necessarily including a self-organizing component, allows us to model a much more realistic genotype/phenotype/environment interaction. Instead of genotypes used exclusively for Mendelian inheritance (see chapter 5) of (directly encoded) phenotypic traits, ES with genotype regulation allow us to model the contextual, plastic development of phenotypes we have come to understand via modern Genomics—thus also learning additional design principles for bio-inspired computation [Huang et al, 2007].

The most important aspect of GA's with emergent morphologies is the utilization in the same model of an external selection engine (the GA) coupled to a particular self-organizing dynamics (e.g. Boolean networks) standing for some materiality. Such schemes bring together, computationally, the two most important aspects of evolutionary systems: self-organization and selection. These models belong to a category of self-organization referred to as *Selected Self-Organization* which is based on symbolic memory [Rocha, 1996, 1997, 1998]. Selected Self-Organization with distributed memory is also possible in autocatalytic structures, though its evolutionary potential is much smaller than the local memory kind [Rocha, 2001][Vasas, 2010]. The reason lies in Von Neumann's notion of self-reproduction (see chapter 6). The introduction of symbolic descriptions allows a much more sophisticated form of communication: structures are constructed from static descriptions and do not have to reproduce through some complicated, and limited process of self-inspection. In other words, separate descriptions can be used to reliably construct any kind of structure in an open-ended manner, while self-inspection relies on only those structures that happen to be able to make copies of themselves. As an example, a non-genetic protein-based life form, would have to rely only on those proteins that could make direct copies of themselves [Rocha, 2001].

Further Readings and References:

Belew, R.K. [1993]."Interposing an Ontogenic Model Between Genetic Algorithms and Neural Networks." In: *Advances in neural information processing (NIPS5).* J. Cowan (Ed.). Morgan Kaufmann.

De Jong, K.A.[2006]. *Evolutionary Computation: A Unified Approach*. MIT Press.

Dellaert, F. and R.D. Beer [1994]."Toward an evolvable model of development for autonomous agent synthesis." In: *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. R. Brooks and P. Maes (Eds.). MIT Press.

Goldberg, D. E. [1989]. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley.

Gruau, Frédéric [1993]."Genetic Sythesis of Modular Neural Networks." In: *Proceedings of the fifth international conference on Genetic Algorithms*. S. Forrest. Morgan Kauffman, pp 318-325.

Hallinan, J. And J Wiles [2004]. “Asynchronous Dynamics of an Artificial Genetic Regulatory Network”. *9th International Conference on Artificial Life*. MIT Press.

Holland, John H. [1975]. *Adaptation in Natural and Artificial Systems.* University of Michigan Press.

C. Huang, J. Kaur, A. Maguitman, L.M. Rocha[2007].""Agent-Based Model of Genotype Editing". *Evolutionary Computation*, 15(3).

Kitano, H. [1990]."Designing Networks using Genetic Algorithms with Graph Generation System." In: *Complex Systems* Vol. 4, pp 461-476.

Kitano, Hiroaki [1994]."Evolution of Metabolism for Morphogenesis." In: *Artificial Life IV: proceedings of the fourth international workshop on the synthesis and simulation of living systems*. R. Brooks and P. Maes (Eds.). MIT Press.

Mitchell, M. [1992]. “Genetic algorithms”. In Lectures in Complex Systems. L. Nadel and D. Stein (Eds.). SFI Studies in the Science of Complexity Vol. V, Addison-Wesley. pp 3-87.

Mitchell, M. [1999]. *An Introduction to Genetic Algorithms*. MIT Press.

Reil, T. 1999. “Dynamics of Gene Expression in an Artificial Genome - Implications for Biological and Artificial Ontogeny”. In *Proceedings of the 5th European Conference on Advances in Artificial Life* (September 13 - 17, 1999). D. Floreano, J. Nicoud, and F. Mondada, Eds. Lecture Notes In Computer Science, vol. 1674. Springer-Verlag, London, 457-466

Rocha, Luis M. [1995]."Contextual Genetic Algorithms: Evolving Developmental Rules." In: *Advances in Artificial Life*. F. Moran, A. Moreno, J.J. Merelo, and P. Chacon (Eds.). Series: Lecture Notes in Artificial Intelligence, Springer-Verlag. pp. 368-382.

Rocha, Luis M. [1996]."Eigenbehavior and symbols." In: *Systems Research* Vol. 12, No. 3, pp. 371-384.

Rocha, Luis M. [1997]. *Evidence Sets and Contextual Genetic Algorithms: Exploring Uncertainty, Context, and Embodiment in Cognitive and Biological Systems*. PhD. Dissertation. SUNY Binghamton.

Rocha, Luis M. [1998]." Selected Self-Organization and the Semiotics of Evolutionary Systems." In: *Evolutionary Systems: The Biological and Epistemological Perspectives on Selection and Self- Organization*. S. Salthe, G. Van de Vijver, and M. Delpos (eds.). Kluwer Academic Publishers, pp. 341-358.

Rocha, Luis M. [2001]. "Evolution with Material Symbol Systems". *Biosystems*. Vol. 60, pp. 95-121.

Rocha, L.M., A. Maguitman, C. Huang, J. Kaur, and S. Narayanan. [2006].""An Evolutionary Model of Genotype Editing". In: *Artificial Life 10: Tenth International Conference on the Simulation and Synthesis of Living Systems*. L.M.Rocha, L. Yaeger, M. Bedau, D. Floreano, R. Goldstone, and A. Vespignani (Eds.). MIT Press, pp. 105-111.

Vasas, Vera, Eors Szathmary, and Mauro Santos [2010]. " Lack of evolvability in self-sustaining autocatalytic networks: A constraint on the metabolism-first path to the origin of life. ". *Proceedings of the National Academy of Sciences of the United States of America*: 0912628107.

Wagner, Gunter [1995] “Adaptation and the modular design of organisms”. In: *Advances in Artificial Life*. F. Moran, A. Moreno, J.J. Merelo, and P. Chacon (Eds.). Series: Lecture Notes in Artificial Intelligence, Springer-Verlag. pp. 317-328.

An Introduction to Evolutionary Computation

Talib S. Hussain

Department of Computing and Information Science
Queen’s University, Kingston, Ont. K7L 3N6

hussain@qucis.queensu.ca

Abstract

Researchers in many fields are faced with computational problems in which a great number of solutions are possible and finding an optimal or even a sufficiently good one is difficult. A variety of search techniques have been developed for exploring such problem spaces, and a promising approach has been the use of algorithms based upon the principles of natural evolution. This tutorial will introduce the basic principles underlying most evolutionary algorithms, as well as some of the key details of the four most popular methods: genetic algorithms, genetic programming, evolutionary strategies, and evolutionary programming. The aim of the tutorial is to introduce the participants to the jargon and principles of the field of evolutionary computation, and to encourage the participants to consider the potential of applying evolutionary optimization techniques in their own research.

1. Introduction

An important area in current research is the development and application of search techniques based upon the principles of natural evolution. Most readers, through the popular literature and typical Western educational experience, are probably aware of the basic concepts of evolution. In particular, the principle of the '*survival of the fittest*' proposed by Charles Darwin (1859) has especially captured the popular imagination. We shall use this as a starting point in introducing evolutionary computation.

The theory of natural selection proposes that the plants and animals that exist today are the result of millions of years of adaptation to the demands of the environment. At any given time, a number of different organisms may co-exist and compete for the same resources in an ecosystem. The organisms that are most capable of acquiring resources and successfully procreating are the ones whose descendants will tend to be numerous in the future. Organisms that are less capable, for whatever reason, will tend to have few or no descendants in the future. The former are said to be more *fit* than the latter, and the distinguishing characteristics that caused the former to be more fit are said to be *selected for* over the characteristics of the latter. Over time, the entire population of the ecosystem is said to *evolve* to contain organisms that, on average, are more fit than those of previous generations of the population because they exhibit more of those characteristics that tend to promote survival.

Evolutionary computation techniques abstract these evolutionary principles into algorithms that may be used to search for optimal solutions to a problem. In a search algorithm, a number of possible solutions to a problem are available and the task is to find the best solution possible in a fixed amount of time. For a search space with only a small number of possible solutions, all the solutions can be examined in a reasonable amount of time and the optimal one found. This *exhaustive search*, however, quickly becomes impractical as the search space grows in size. Traditional search algorithms randomly sample (e.g., *random walk*) or heuristically sample (e.g., *gradient descent*) the search space one solution at a time in the hopes of finding the optimal solution. The key aspect distinguishing an evolutionary search algorithm from such traditional algorithms is that it is *population-based*. Through the adaptation of successive generations of a large number of individuals, an evolutionary algorithm performs an efficient directed search. Evolutionary search is generally better than random search and is not susceptible to the hill-climbing behaviors of gradient-based search.

2. Basic Evolutionary Computation

In an evolutionary algorithm, a *representation scheme* is chosen by the researcher to define the set of solutions that form the search space for the algorithm. A number of individual solutions are created to form an *initial population*. The following steps are then repeated iteratively until a solution has been found which satisfies a pre-defined *termination criterion*. Each individual is evaluated using a *fitness function* that is specific to the problem being solved. Based upon their fitness values, a number of individuals are chosen to be *parents*. New individuals, or *offspring*, are produced from those parents using *reproduction operators*. The fitness values of those offspring are determined. Finally, survivors are selected from the old population and the offspring to form the new population of the next *generation*.

The mechanisms determining which and how many parents to select, how many offspring to create, and which individuals will survive into the next generation together represent a *selection method*. Many different selection methods have been proposed in the literature, and they vary in complexity. Typically, though, most selection methods ensure that the population of each generation is the same size.

The remainder of the paper presents the traditional definitions of the four most common evolutionary algorithms: genetic algorithms (Holland, 1975), genetic programming (Koza, 1992, 1994), evolutionary strategies (Rechenberg, 1973), and evolutionary programming (Fogel et al., 1966). The traditional differences between the approaches involve the nature of the representation schemes, the reproduction operators, and the selection methods.

3. Genetic Algorithms

The most popular technique in evolutionary computation research has been the *genetic algorithm*. In the traditional genetic algorithm, the representation used is a *fixed-length bit string*. Each position in the string is assumed to represent a particular feature of an individual, and the value stored in that position represents how that feature is expressed in the solution. Usually, the string is "evaluated as a collection of *structural* features of a solution that have little or no interactions" (Angeline, 1996, p. 4). The analogy may be drawn directly to genes in biological organisms. Each gene represents an entity that is structurally independent of other genes.

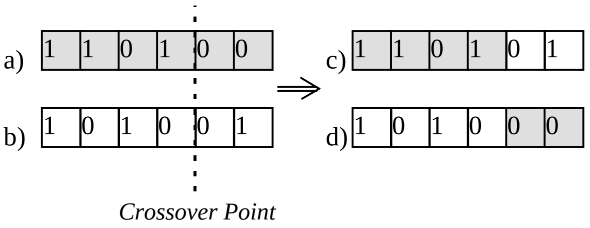


Figure 1: Bit-String Crossover of Parents a & b
to form Offspring c & d

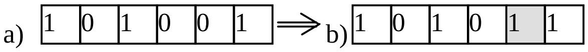


Figure 2: Bit-Flipping Mutation of Parent a
to form Offspring b

The main reproduction operator used is *bit-string crossover*, in which two strings are used as parents and new individuals are formed by swapping a sub-sequence between the two strings (see Figure 1). Another popular operator is *bit-flipping mutation*, in which a single bit in the string is flipped to form a new offspring string (see Figure 2). A variety of other operators have also been developed, but are used less frequently (e.g., *inversion*, in which a subsequence in the bit string is reversed). A primary distinction that may be made between the various operators is whether or not they introduce any new information into the population. Crossover, for example, does not while mutation does. All operators are also constrained to manipulate the string in a manner consistent with the structural interpretation of genes. For example, two genes at the same location on two strings may be swapped between parents, but not combined based on their values.

Traditionally, individuals are selected to be parents *probabilistically* based upon their fitness values, and the offspring that are created replace the parents. For example, if N parents are selected, then N offspring are generated which replace the parents in the next generation.

4. Genetic Programming

An increasingly popular technique is that of *genetic programming*. In a standard genetic program, the representation used is a variable-sized tree of functions and values. Each leaf in the tree is a label from an available set of value labels. Each internal node in the tree is label from an available set of function labels. The entire tree corresponds to a single function that may be evaluated. Typically, the tree is evaluated in a left-most depth-first manner. A leaf is evaluated as the corresponding value. A function is evaluated using as arguments the result of the evaluation of its children.

Genetic algorithms and genetic programming are similar in most other respects, except that the reproduction operators are tailored to a tree representation. The most commonly used operator is *subtree crossover*, in which an entire subtree is swapped between two parents (see Figure 3). In a standard genetic program, all values and functions are assumed to return the same type, although functions may vary in the number of arguments they take. This *closure* principle (Koza, 1994) allows any subtree to be considered structurally on par with any other subtree, and ensures that operators such as sub-tree crossover will always produce legal offspring.

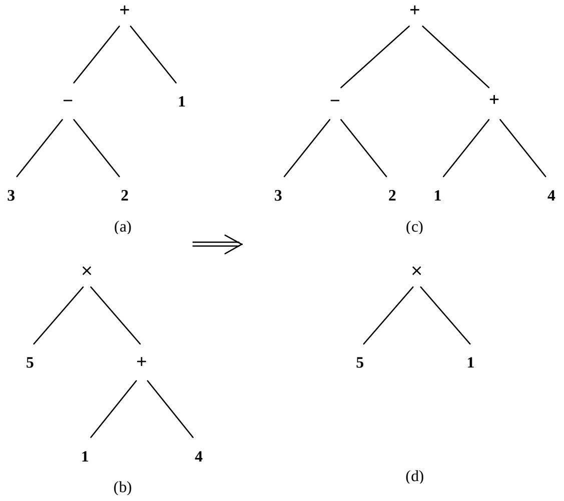


Figure 3: Subtree Crossover of Parents a & b
to form Offspring c & d

5. Evolutionary Strategies

In evolutionary strategies, the representation used is a fixed-length real-valued vector. As with the bit-strings of genetic algorithms, each position in the vector corresponds to a feature of the individual. However, the features are considered to be behavioral rather than structural. "Consequently, arbitrary non-linear interactions between features during evaluation are expected which forces a more holistic approach to evolving solutions" (Angeline, 1996, p. 4).

The main reproduction operator in evolutionary strategies is *Gaussian mutation*, in which a random value from a Gaussian distribution is added to each element of an individual's vector to create a new offspring (see Figure 4). Another operator that is used is *intermediate recombination*, in which the vectors of two parents are averaged together, element by element, to form a new offspring (see Figure 5). The effects of these operators reflect the behavioral as opposed to structural interpretation of the representation since knowledge of the values of vector elements is used to derive new vector elements.

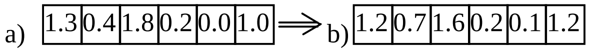


Figure 4: Gaussian Mutation of Parent a
to form Offspring b

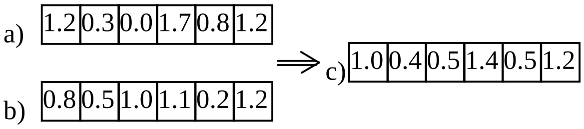


Figure 5: Intermediate Recombination of Parents a & b
to form Offspring c

The selection of parents to form offspring is less constrained than it is in genetic algorithms and genetic programming. For instance, due to the nature of the representation, it is easy to average vectors from many individuals to form a single offspring. In a typical evolutionary strategy, N parents are selected uniformly randomly (i.e., not based upon fitness), more than N offspring are generated through the use of recombination, and then N survivors are selected deterministically. The survivors are chosen either from the best N offspring (i.e., no parents survive) or from the best N parents and offspring (Spears et al., 1993).

6. Evolutionary Programming

The representations used in evolutionary programming are typically tailored to the problem domain (Spears et al., 1993). One representation commonly used is a fixed-length real-valued vector.

The primary difference between evolutionary programming and the previous approaches is that no exchange of material between individuals in the population is made. Thus, only mutation operators are used. For real-valued vector representations, evolutionary programming is very similar to evolutionary strategies without recombination.

A typical selection method is to select all the individuals in the population to be the N parents, to mutate each parent to form N offspring, and to probabilistically select, based upon fitness, N survivors from the total 2N individuals to form the next generation.

7. Current Issues

In current research, the line distinguishing these different approaches has started to blur. Researchers in each technique have begun to examine more complex representation schemes and to apply a variety of selection methods. Many genetic algorithm researchers are examining the use variable-length representations and analyzing how such representations grow in size over the course of evolution (Wu & Lindsay, 1996). Many genetic algorithms now use selection methods, such as *elitist recombination*, in which parents compete with their offspring for survival into the next generation (Thierens, 1997). Some genetic programming researchers have begun to examine the effects of allowing multiple types of functions and values into the representation. The benefits of such strongly typed genetic programming are only beginning to be explored (Haynes et al., 1996).

References

Angeline, P.J. (1996) "Genetic programming's continued evolution," Chapter 1 in K.E. Kinnear, Jr. and P.J. Angeline (Eds.), *Advances in Genetic Programming 2*. Cambridge, MA: MIT Press, p. 1 -20.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection*. London: John Murray.

Fogel, L.J., Owens, A.J. & Walsh, M.J. (1966) *Artificial Intelligence through Simulated Evolution*. New York: John Wiley.

Haynes, T.D., Schoenefeld, D.A. & Wainwright, R.L. (1996) "Type inheritance in strongly typed genetic programming," Chapter 18 in K.E. Kinnear, Jr. and P.J. Angeline (Eds.), *Advances in Genetic Programming 2*. Cambridge, MA: MIT Press. p. 359-376.

Holland, J.H. (1975) *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Koza, J.R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Koza, J.R. (1994) *Genetic Programming II: Automatic Discovery of Reusable Programs*. Cambridge, MA: MIT Press.

Rechenberg, I. (1973) *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Stuttgart: Frommann-Holzboog Verlag.

Spears, W.M., DeJong, K.A., Back, T., Fogel, D.B., & deGaris, H. (1993) "An overview of evolutionary computation," *Proceedings of the 1993 European Conference on Machine Learning*.

Thierens, D. (1997) "Selection schemes, elitist recombination, and selection intensity," *Proceedings of the Seventh International Conference on Genetic Algorithms*. San Francisco, CA: Morgan Kauffman, p. 152-159.

Wu, A. & Lindsay, R.K. (1996) "A survey of intron research in genetics," In Voigt, H., Ebelin, W., Rechenberg, I., & Schwefel, H. (Eds.), *Parallel Problem Solving from Nature - PPSN IV*. Berlin: Springer-Verlag, p. 101-110.

3.4 EVOLUTIONARY BIOLOGY

Evolutionary biology is a science concerned, among other things, with the study of the diversity of life, the differences and similarities among organisms, and the adaptive and non-adaptive characteristics of organisms. Its importance are manifold, from the health sciences to the understanding of how the living organisms adapt to the environment they inhabit. For instance, evolutionary biology helps in the understanding of disease epidemics, population dynamics, and the production of improved cultures. Over roughly the last 60 years, computer scientists and engineers realized that evolutionary biology has various interesting ideas for the development of theoretical models of evolution (some of them being rather abstract models) that can be useful to obtain solutions to complex real-world problems.

The word *evolution* is originated from the Latin *evolvere*, which means to unfold or unroll. Broadly speaking, evolution is a synonym for 'change'. But what type of change? We do not usually employ the word evolution to refer to the changes suffered by an individual during its lifetime. Instead, an evolving system corresponds to the one in which there is a descent of entities over time, one generation after the other, and in which characteristics of the entities differ across generations (Futuyma, 1998). Therefore, *evolution* can be broadly defined as *descent with modification* and often *with diversification*. Many systems can be classified as evolutionary: languages, cellular reproduction in immune systems, cuisines, automobiles, and so on.

Any evolutionary system presents a number of features:

- *Population(s)*: in all evolutionary systems there are populations, or groups, of entities, generally termed *individuals*.
- *Reproduction*: in order for evolution to occur, the individuals of the population(s) must reproduce either sexually or asexually.
- *Variation*: there is variation in one or more characteristics of the individuals of the population(s).
- *Hereditary similarity*: parent and offspring individuals present similar characteristics. Over the course of generations, there may be changes in the proportions of individuals with different characteristics within a population; a process called *descent with modification*.
- *Sorting of variations*: among the sorting processes, it can be emphasized *chance* (random variation in the survival or reproduction of different variants), and *natural selection* (consistent, non-random differences among variants in their rates of survival and reproduction).

Adaptation as a result of variation plus natural selection leads to improvement in the function of an organism and its many component parts. "Biological *or* organic evolution *is change in the properties of populations of organisms, or groups of such populations, over the course of generations*." (Futuyma, 1998; p. 4). Note that according to this definition of evolution, individual organisms do not evolve and the changes of a population of individuals that are assumed to be

Copyrighted material

evolutionary are those resultant from inheritance, via the genetic material, from one generation to the other.

The history of evolutionary biology is marked by a number of hypotheses and theories about how life on earth appeared and evolved. The most influential theory to date is the one proposed by Charles Darwin and formalized in his book *On the Origins of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (Darwin, 1859). Historically, Alfred Wallace is also one of the proponents of the theory of evolution by means of natural selection, but it was Darwin's book, with its hundreds of instances and arguments supporting natural selection, the landmark for the theory of evolution.

Among the many preDarwinian hypotheses for the origin and development of beings, the one proposed by Jean Baptist Pierre Antoine de Monet, chevalier de Lamarck, was the most influential. According to Lamarck, every species originated individually by spontaneous generation. A 'nervous fluid' acts within each species, causing it to progress up the chain over time, along a single predetermined path that every species is destined to follow. No extinction has occurred: fossil species are still with us, but have been transformed. According to Lamarck, species also adapt to their environments, the more strongly exercised organs attract more of the nervous fluid, thus getting enlarged; conversely, the less used organs become smaller. These alterations, acquired during an individual's lifetime through its activities, are inherited. Like everyone at that time, Lamarck believed in the so-called *inheritance of acquired characteristics*.

The most famous example of Lamarck's theory is the giraffe: according to Lamarck, giraffes need long necks to reach the foliage above them; because they are constantly stretching upward, the necks grow longer; these longer necks are inherited; and over the course of generations the necks of giraffes get longer and longer. Note that the theory of inheritance of acquired characteristics is not Lamarck's original, but an already established supplement to his theory of 'organic progression' in which spontaneous generation and a chain of beings (progression from inanimate to barely animate forms of life, through plants and invertebrates, up to the higher forms) form the basis. Lamarck's theory may also be viewed as a *transformational theory*, in which change is programmed into every member of the species.

3.4.1. On the Theory of Evolution

Darwin's studies of the natural world showed a striking diversity of observations over the animal and vegetal kingdoms. His examples were very wide ranging, from domestic pigeons, dogs, and horses, to some rare plants. His research that resulted in the book *Origin of Species* took literally decades to be concluded and formalized.

In contrast to the Lamarckian theory, Darwin was certain that the direct effects of the conditions of life were unimportant for the variability of species.

> "Seedlings from the same fruit, and the young of the same litter, sometimes differ considerably from each other, though both the young and the

Copyrighted material

> parents … have apparently been exposed to exactly the same conditions of life; and this shows how unimportant the direct effects of the conditions of life are in comparison with the laws of reproduction, and of growth, and of inheritance; for had the action of the conditions been direct, if any of the young had varied, all would probably have varied in the same manner." (Darwin, 1859; p. 10)

Darwin starts his thesis of how species are formed free in nature by suggesting that the most abundant species (those that range widely over the world) are the most diffused and which often produce well-marked varieties of individuals over the generations. He describes some basic rules that promote improvements in organisms: reproduce, change and compete for survival.

Natural selection was the term used by Darwin to explain how new characters arising from variations are preserved. He starts thus paving the ground to his theory that slight differences in organisms accumulated over many successive generations might result in the appearance of completely new and more adapted species to their environment. As defended by himself

> "… as a general rule, I cannot doubt that the continued selection of slight variations … will produce races differing from each other …" (Darwin, 1859; p. 28) and "… I am convinced that the accumulative action of Selection, whether applied methodically and more quickly, or unconsciously and more slowly, but more efficiently, is by far the predominant Power." (Darwin, 1859; p. 35)

In summary, according to Darwin's theory, evolution is a result of a population of individuals that suffer:

- Reproduction with inheritance.
- Variation.
- Natural selection.

These very same processes constitute the core of all evolutionary algorithms. Before going into the details as to how reproduction and variation happen within individuals and species of individuals, some comments about why Darwin's theory was so revolutionary and 'dangerous' at that time (and, to some people, until nowadays) will be made.

3.4.2. Darwin's Dangerous Idea

Darwin's theory of evolution is controversial and has been refuted by many because it presents a sound argument for how a "Nonintelligent Artificer" could produce the wonderful forms and organisms we see in nature. To D. Dennett (1991), *Darwin's dangerous idea* is that evolution, thus life, can be explained as the product of an *algorithmic process*, not of a superior being (God) creating everything that might look wonderful to our eyes. But the reason there is a section on Dennett's book here is not to discuss particular beliefs. Instead, to discourse about some key interpretations of evolution, from a computational perspective, presented by D. Dennett in his book *Darwin's Dangerous Idea: Evo-*

Copyrighted material

lution and the Meanings of Life. These are not only interesting, but also useful for the understanding of why the theory of evolution is suitable for the comprehension and development of a class of search techniques known as evolutionary algorithms.

Dennett defines an algorithm as a certain sort of formal process that can be counted on (logically) to yield a certain sort of result whenever it is run or instantiated. He emphasizes that *evolution* can be understood and represented in an abstract and common terminology as an *algorithmic process*; it can be lifted out of its home base in biology. Evolutionary algorithms are thus those that embody the major processes involved in the theory of evolution: a population of individuals that *reproduce with inheritance*, and suffer *variation* and *natural selection*.

Dennett also discusses what can be the outcomes of evolution and its probable implications when viewed as an *engineering process*. He stresses the importance of genetic variation and selection, and quotes an interesting passage from M. Eigen (1992).

> "Selection is more like a particularly subtle demon that has operated on the different steps up to life, and operates today at the different levels of life, with a set of highly original tricks. Above all, it is highly active, driven by an internal feedback mechanism that searches in a very discriminating manner for the best route to optimal performance, not because it possesses an inherent drive towards any predestined goal, but simply by virtue of its inherent non-linear mechanism, which gives the appearance of goal-directedness." (Eigen, 1992; quoted by Dennett, 1991, p. 195)

Another important argument is that evolution requires *adaptation* (actually it can also be seen as adaptation plus selection, as discussed in the previous chapter). From an evolutionary perspective, adaptation is the reconstruction or prediction of evolutionary events by assuming that all characters are established by direct natural selection of the most adapted state, i.e. the state that is an 'optimum solution' to a 'problem' posed by the environment. Another definition is that under adaptation, organisms can be viewed as complex adaptive systems whose parts have (adaptive) functions subsidiary to the fitness-promoting function of the whole.

The key issue to be kept in mind here is that evolution can be viewed as an algorithmic process that allows - via reproduction with inheritance, variation and natural selection - the most adapted organisms to survive and be driven to a state of high adaptability (optimality) to their environment. These are the inspiring principles of evolutionary algorithms; the possibility of modeling evolution as a search process capable of producing individuals (candidate solutions to a problem) with increasingly better 'performances' in their environments.

3.4.3. Basic Principles of Genetics

The theory of evolution used in the development of most evolutionary algorithms is based on the three main aspects raised by Darwin as being responsible

Copyrighted material

for the evolution of species: reproduction with inheritance, variation, and selection. However, the origins of heredity along with variations, which were some of the main ingredients for the natural selection theory, were unknown at that time. This section explores the genetic basis of reproduction and variation in order to provide the reader with the necessary biological background to develop and understand evolutionary algorithms, in particular *genetic algorithms*. The union of genetics with some notions of the selection mechanisms, together with Darwin's hypotheses led to what is currently known as neo-Darwinism.

Gregor Mendel's paper establishing the foundations of *genetics* (a missing bit for a broader understanding of the theory of evolution) was published only in 1865 (Mendel, 1865), but it was publicly ignored until about the 1900. He performed a series of careful breeding experiments with garden peas. In summary, Mendel selected strains of peas that differed in particular *traits* (characteristics). As these differences were clearly distinguishable, their *phenotypes* (measurable attributes, or observable physical or biochemical characteristics of an organism) were identified and scored. For instance, the pea seeds were either smooth or wrinkled, the pod shape was either inflated or constricted, and the seed color was either yellow or green. Then, Mendel methodically performed crosses among the many pea plants, counted the progeny, and interpreted the results. From this kind of data, Mendel concluded that phenotypic traits were controlled by *factors*, later called *Mendelian factors*, and now called *genes*. *Genotype* is the term currently used to describe the genetic makeup of a cell or organism, as distinguished from its physical or biochemical characteristics (the phenotype). Figure 3.4 summarizes the first experiment performed by Mendel.

The basic structural element of all organisms is the *cell*. Those organisms whose genetic material is located in the *nucleus* (a discrete structure within the cell that is bounded by a nuclear membrane) of the cells are named *eukaryotes*. *Prokaryotes* are the organisms that do not possess a nuclear membrane surrounding their genetic material. The description presented here focuses on eukaryotic organisms.

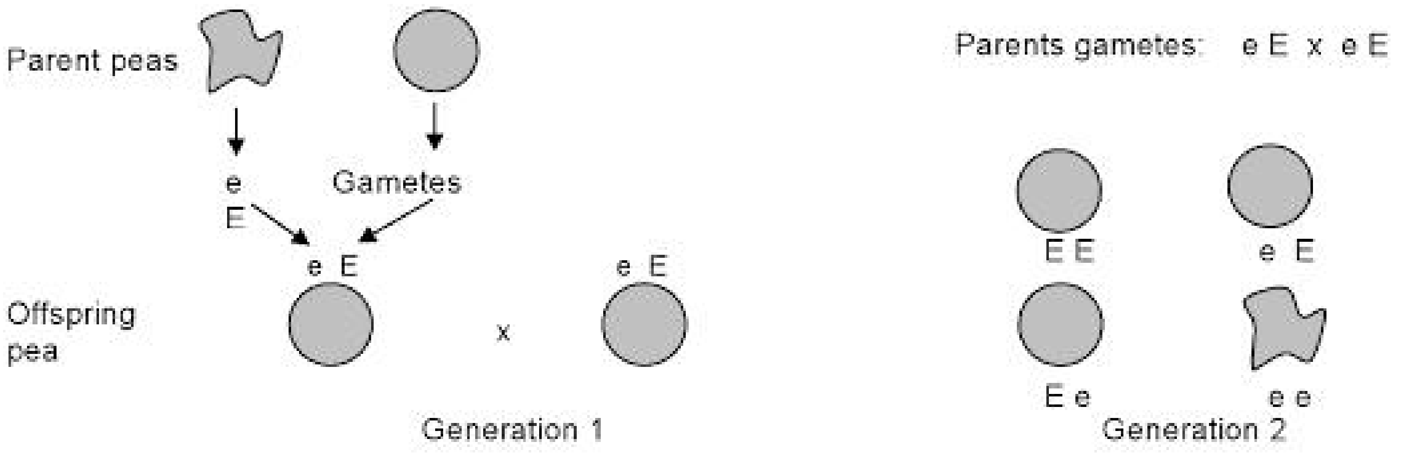


Figure 3.4: First experiment of Mendel. When crossing a normal pea with a wrinkled pea, a normal pea was generated (generation 1). By crossing two daughters from generation 1, three normal peas were generated plus one wrinkled pea. Thus, there is a recessive gene (e) that only manifests itself when there is no dominant gene together. Furthermore, there is a genetic inheritance from parents to offspring; those offspring that carry a factor that expresses a certain characteristic may have offspring with this characteristic.

Copyrighted material

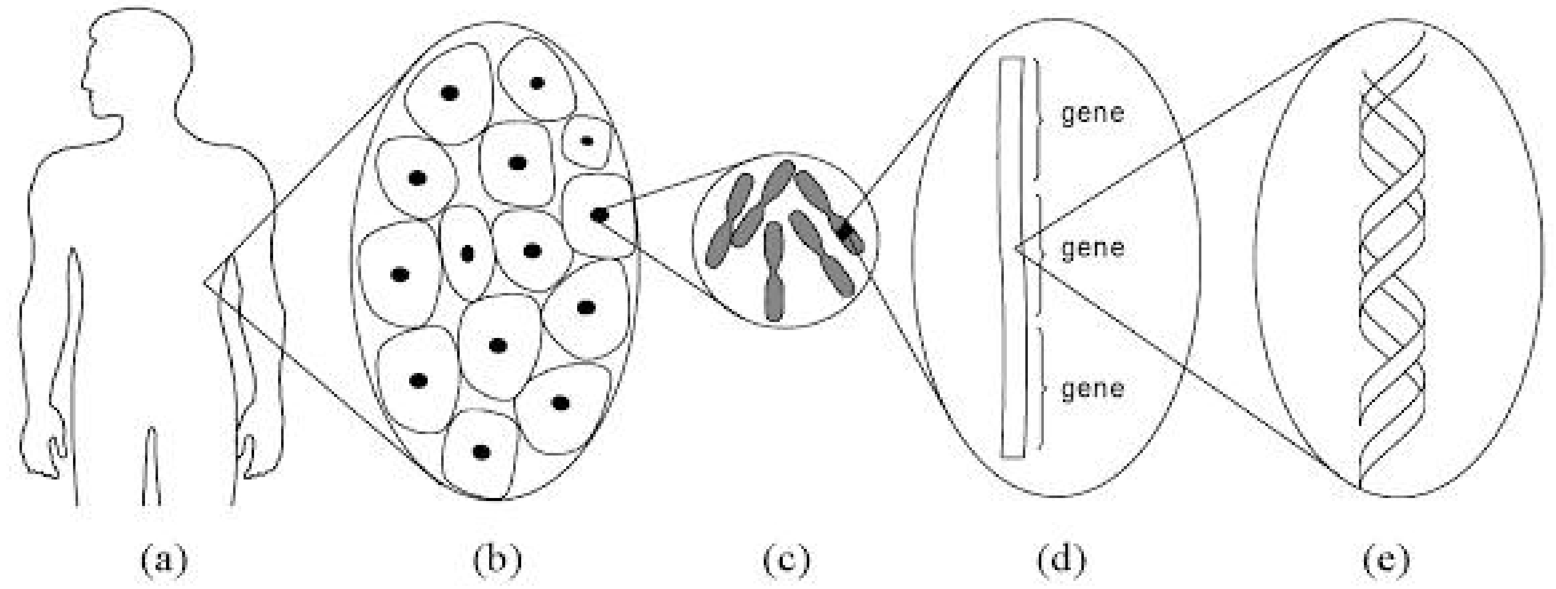


Figure 3.5: Enlargement of an organism to focus the genetic material. (a) Human organism. (b) Cells composing the organism. (c) Each cell nucleus contains chromosomes. (d) Each chromosome is composed of a long DNA segment, and the genes are the functional portions of DNA. (e) The double helix of DNA. (Modified with permission from [Griffiths et al., 1996], © W. H. Freeman and Company.)

In the cell nucleus, the genetic material is complexed with protein and is organized into a number of linear structures called *chromosomes*, which means, 'colored body', and is so named because these threadlike structures are visible under the light microscope only after they are stained with dyes. A *gene* is a segment of a helix molecule called *deoxyribonucleic acid*, or *DNA* for short. Each eukaryotic chromosome has a single molecule of DNA going from one end to the other. Each cell nucleus contains one or two sets of the basic DNA complement, called *genome*. The genome itself is made of one or more chromosomes. The *genes* are the functional regions of DNA. Figure 3.5 depicts a series of enlargements of an organism to focus on the genetic material.

It is now known that the DNA is the basis for all processes and structures of life. The DNA molecule has a structure that contributes to the two most fundamental properties of life: reproduction and development. DNA is a double helix structure with the inherent feature of being capable of replicating itself before the cell multiplication, allowing the chromosomes to duplicate into *chromatids*,

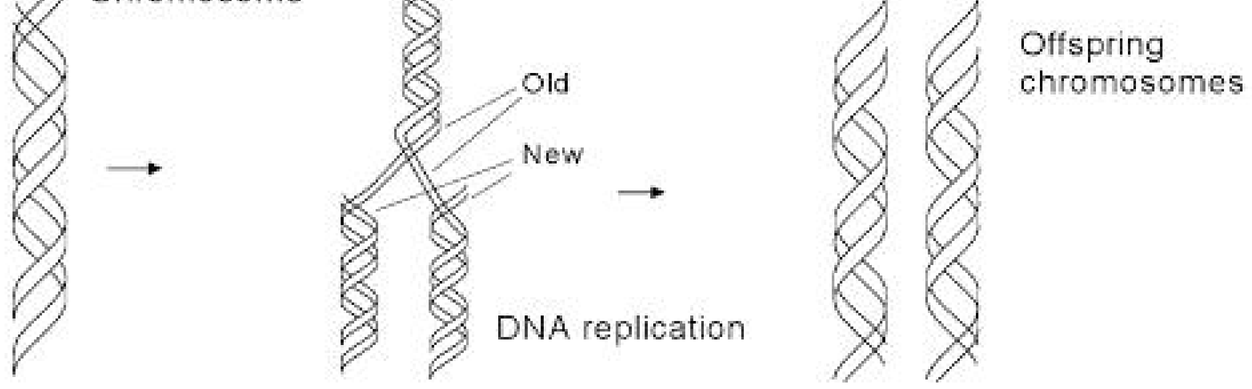


Figure 3.6: When new cells are formed, the DNA replication allows a chromosome to have a pair of offspring chromosomes and be passed onto the offspring cells.

Copyrighted material

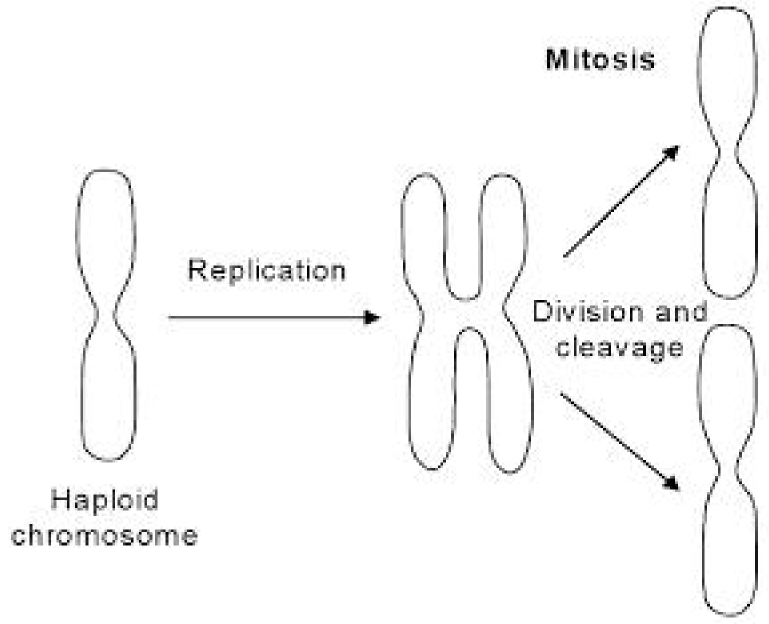


Figure 3.8: Asexual reproduction in haploids. The chromosome replicates itself, the cell nucleus is divided through a process named mitosis, and then the cell is divided into two identical progeny.

Sexual reproduction is the fusion of two *haploid gametes* (sex cells) to produce a single *diploid zygote* cell. An important aspect of sexual reproduction is that it involves *genetic recombination*; that is, it generates gene combinations in the offspring that are distinct from those in the parents. Sexually reproducing organisms have two sorts of cells: *somatic* (body) cells, and *germ* (sex) cells. All somatic cells reproduce by a process called *mitosis* that is a process of nuclear division followed by cell division. Figure 3.9 illustrates the process of sexual reproduction.

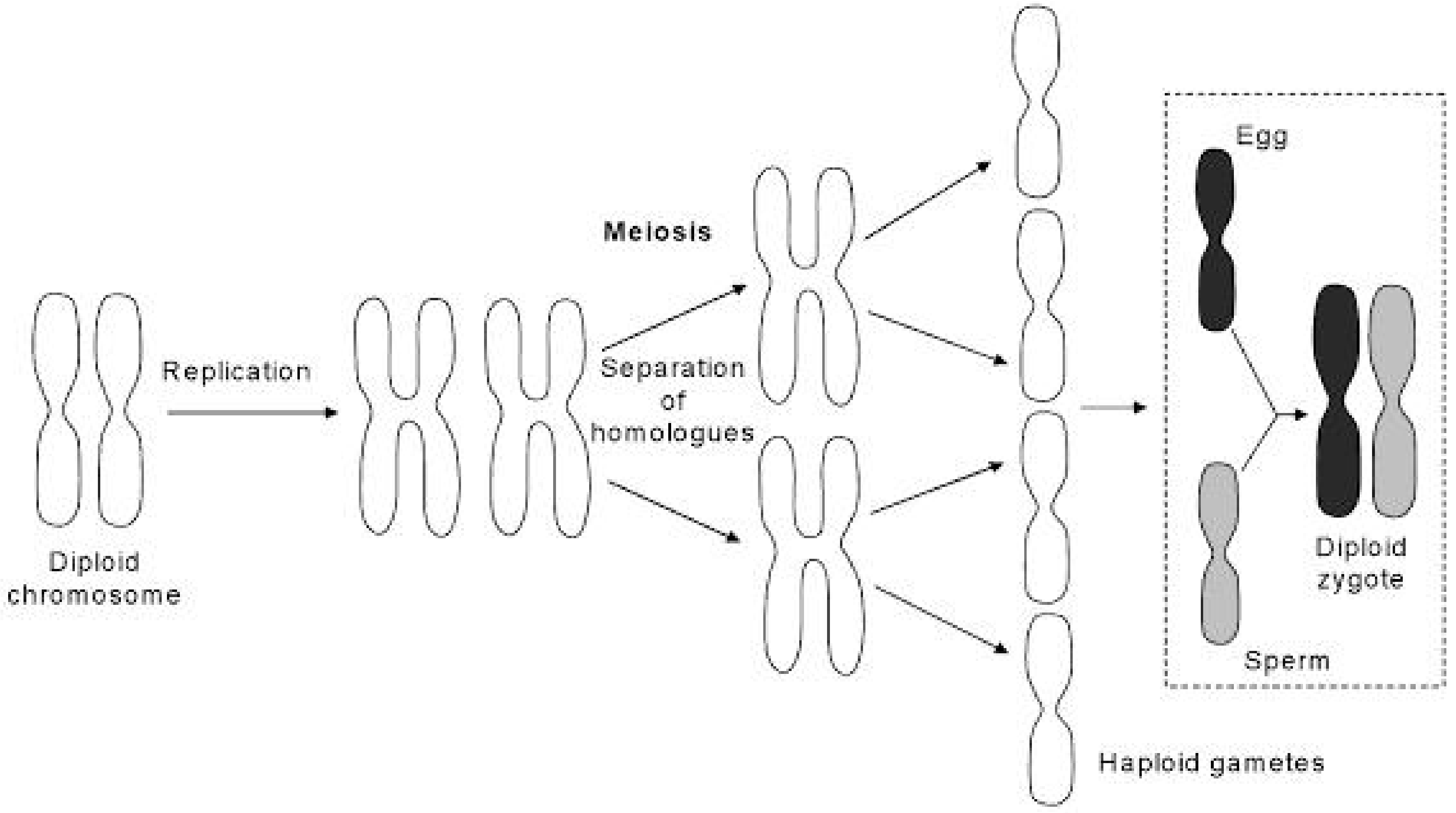


Figure 3.9: Sexual reproduction. A diploid chromosome replicates itself, then the homologues are separated generating haploid gametes. The gametes from each parent are fused to generate a diploid zygote.

Copyrighted material

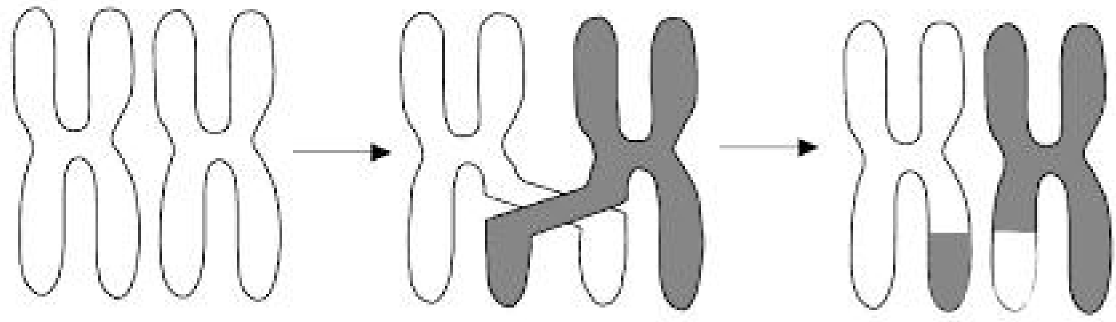


Figure 3.10: Crossing over between two loci in a cell undergoing the first meiotic division. Of the four chromatids, two will have new combinations and two will retain the parental combination of alleles.

In the classical view of the meiosis process in sexual reproduction, homologous chromosomes first undergo the formation of a very tight association of homologues, and then the reciprocal physical exchange of chromosome segments at corresponding positions along pairs of homologous chromosomes, a process termed *crossover* (Russel, 1996). Crossing-over is a mechanism that can give rise to *genetic recombination*, a process by which parents with different genetic characters give birth to progeny so that genes are associated in new combinations. Figure 3.10 depicts the crossing-over process.

The differences among organisms are outcomes of the evolutionary processes of *mutation* (a change or deviation in the genetic material), *recombination* or *crossover* (exchange of genetic material between *chromosomes*; see Figure 3.10), and *selection* (the favoring of particular combinations of genes in a given environment). With the exception of gametes, most cells of the same eukaryotic organism characteristically have the same number of chromosomes. Further, the organization and number of genes on the chromosomes of an organism are the same from cell to cell. These characteristics of chromosome number and gene organization are the same for all members of the same species. *Deviations* are known as *mutations*; these can arise spontaneously or be induced by chemical or radiation mutagens. Several types of mutation exist, for instance point mutation, deletion, translocation, and inversion. Point mutation, deletion and inversion are illustrated in Figure 3.11.

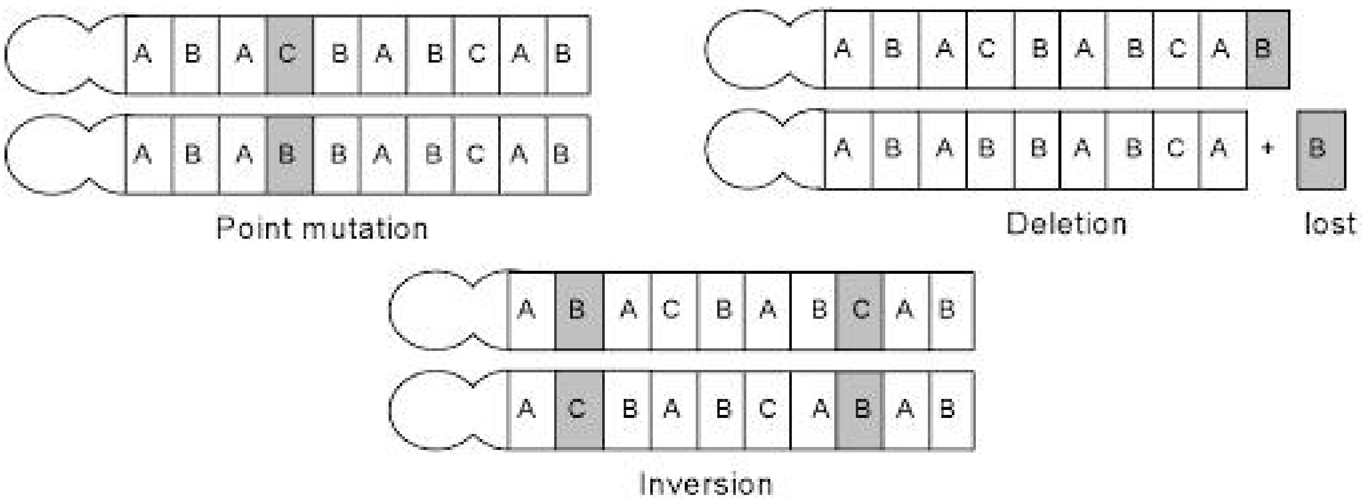


Figure 3.11: Some types of chromosomal mutation, namely, point mutation, deletion, and inversion.

Copyrighted material

3.4.4. Evolution as an Outcome of Genetic Variation Plus Selection

So far we have seen the two types of reproduction, sexual and asexual, and some of the main mechanisms that alter the genetic makeup of a population of individuals, emphasizing crossover and mutation. It still remains to discuss the process by which these altered individuals survive over the generations.

Populations of individuals change over time. The number of individuals may increase or decrease, depending on food resources, climate, weather, availability of breeding areas, predators, and so forth. At the genetic level, a population may change due to a number of factors, such as mutation and selection. These processes not only alter allele frequencies, but also result in changes in the adaptation and diversity of populations, thus leading to the evolution of a species (Gardner et al., 1991).

The viability and fertility of an individual are associated with *fitness*, a term that is used to describe the overall ability of an organism to survive and reproduce. In many populations, survival and reproductive ability are variable traits. Some individuals die before they have a chance to reproduce, whereas others leave many progeny. In a population of stable size, the average number of offspring produced by an individual is one.

Variation in fitness is partially explained by the underlying genetic differences of individuals. The crossing-over of parental genetic material and mutation can increase or decrease fitness, depending on their effects on the survival and reproductive capabilities of the individuals. Thus, genetic recombination and mutation can create phenotypes with different fitness values. Among these, the most fit will leave the largest number of offspring. This differential contribution of progeny implies that alleles associated with superior fitness will increase in frequency in the population. When this happens, the population is said to be undergoing *selection*.

As Darwin made a series of observations of domestic animals and plants, and also those existing free in nature, he used the term natural selection to describe the latter in contrast to men's selection capabilities of domestic breeds. To our purposes, the more general term *selection* is assumed in all cases, bearing in mind that selection under nature has been originally termed natural selection, and selection made by men has been sometimes termed artificial selection.

Under the evolutionary biology perspective, *adaptation* is the process by which traits evolve making organisms more suited to their immediate environment; these traits increase the organisms' chances of survival and reproduction. Adaptation is thus responsible for the many extraordinary traits seen in nature, such as eyes that allow us to see, and the sonar in bats that allow their guidance through the darkness. Note however, that, more accurately speaking, adaptation is a result of the action of both, variation and selection. Variation by itself does not result in adaptation; there must be a way (i.e., selection) of promoting the maintenance of those advantageous variations.

S. Wright (1968-1978) introduced the concept of *adaptive landscapes* or *fitness landscapes*, largely used in evolutionary biology. In his model, each po-

Copyrighted material

pulation of a *species* (reproductively isolated group) is symbolized by a point on a *topographic map*, or *landscape*. The contours of the map represent different levels of adaptation to the environment (fitness). Populations at high levels (peaks) are more adapted to the environment, and populations at low levels (valleys) are less adapted. At any one time, the position of a population will depend on its genetic makeup. Populations with alleles that improve fitness will be at a higher peak than populations without these alleles. Consequently, as the genetic makeup of a population changes, so will its position on the adaptive landscape. Figure 3.12 depicts a landscape representing the different levels of adaptation of the populations in relation to the environment.

The adaptive (fitness) landscape corresponds to the response surface discussed in Section 3.2.1 in the context of problem solving via search in a search space. Note that, under the evolutionary perspective, the search performed is for individuals with increased survival and reproductive capabilities (fitness) in a given environment (fitness landscape).

A *niche* can thus be defined as the region consisting of the set of possible environments in which a species can persist; members of one species occupy the same ecological niche. In natural ecosystems, there are many different ways in which animals may survive (grazing, hunting, on water, etc.), and each survival strategy is called an ecological niche. However, it is generally recognized that the niche of a single species may vary widely over its geographical range. The other fundamental concept of niche was proposed by Elton (1927) "The niche of an animal means its place in the biotic environment, its relations to food and enemies;" where the term *biotic* refers to life, living organisms. Thus, niche in this case is being used to describe the role of an animal in its community (Krebs, 1994).

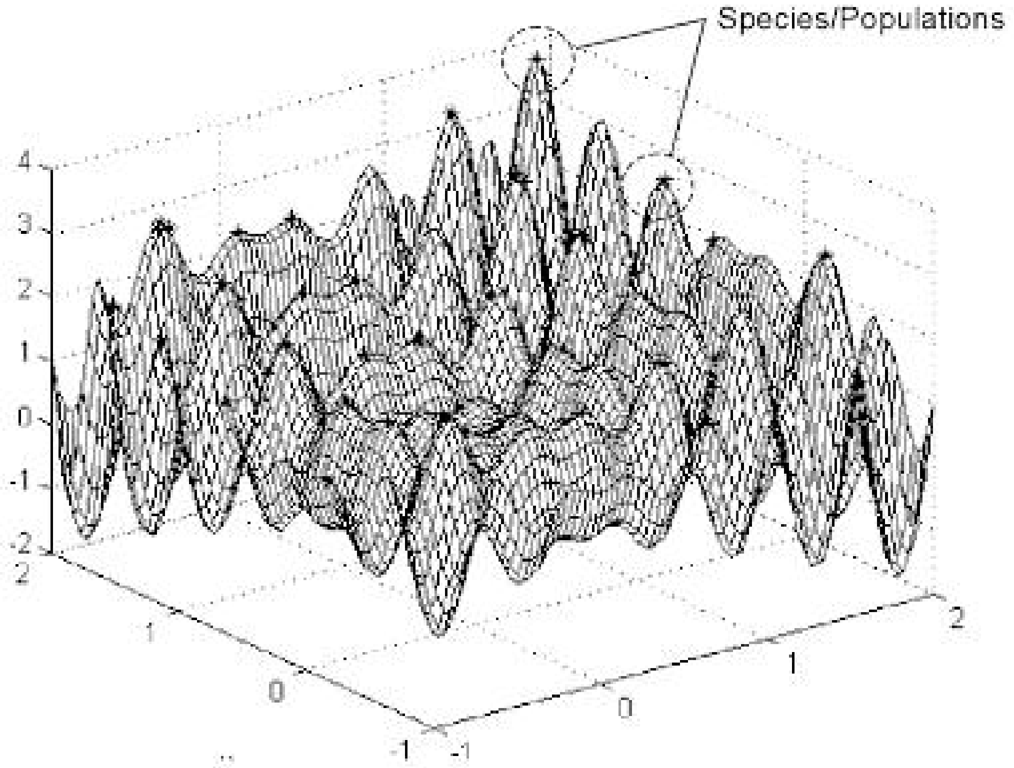


Figure 3.12: An example of an adaptive landscape. The topographic map (landscape or surface) corresponds to the different levels of adaptation of the populations (points in the landscape) to the environment. The populations or individuals at each peak are assumed to be reproductively isolated, i.e., they only breed with individuals in the same peak, thus forming species inhabiting distinct niches.

Copyrighted material

It has been discussed that evolution is an outcome of genetic variation plus selection. In order for continuing evolution to occur, there must be mechanisms that increase or create genetic variation and mechanisms that decrease it. We have also seen that recombination and mutation cause differences among (variations in) organisms. Other two important mechanisms of evolution are the so-called *genetic drift* (chance fluctuations that result in changes in allele frequencies) and *gene flow* (spread of genes among populations via *migration*). It is known that selection and genetic drift decrease variation, while mutation, recombination, and gene flow increase genetic variation (Colby, 1997).

Natural selection sifts through the genetic variations in the population, preserving the beneficial ones and eliminating the harmful ones. As it does this, selection tends to drive the population uphill in the adaptive or fitness landscape. By contrast, the random genetic drift will move the population in an unpredictable fashion. The effect of all these mechanisms (mutation, recombination, genetic drift, gene flow, plus selection) will bring the population to a state of 'genetic' equilibrium, corresponding to a point near or at a peak on the adaptive landscape. Actually, the population will hover around a peak because of fluctuations caused by genetic drift. Note also, that, under nature, the environment is constantly changing, hence the population is also adapting to the new landscape resultant from the new environment, in a never-ending process of variation and selection.

3.4.5. A Classic Example of Evolution

A classic example of evolution comes from species that live in disturbed habitats. In the particular example of the evolution of melanic (dark) forms of moths, human activity has altered the environment and there has been a corresponding change in the species that inhabit this environment. The peppered moth, *Biston betularia*, is found in wooded areas in Great Britain, where it exists in two color forms, light and dark; light being the typical phenotype of this species. The difference between the two forms is believed to involve a single gene. Since 1850, the frequency of the dark form has increased in certain areas in England, in particular in industrialized parts of the country. Around heavily industrialized cities, such as Manchester and Birmingham, the frequency of the dark form has increased drastically from 1 to 90% in less than 100 years. In other areas of England, where there is little industrial activity, the dark form has remained very rare (Gardner et al., 1991).

The rapid spread of the dark form in industrialized areas has been attributed to natural selection. Both, light and dark forms are active at night. During the day, the moths remain still, resting on tree trunks and other objects in the woodlands. Since birds may find the moths and eat them during their resting period, camouflage is their only defense against predation. On white or gray tree bark, the light moths are protectively colored, especially if the bark is overgrown with lichens. However, in industrialized areas most of the lichens have been killed by pollution and the tree bark is oftenest darkened by soot. Such conditions offer little or no cover for the light moths, but make ideal resting spots for the dark

Copyrighted material