Marcello Buiatti[1] and Marco Buiatti[2]

Chance vs. Necessity in Living Systems: A False Antinomy

Abstract. *The concepts of order and randomness are crucial to understand 'living systems' structural and dynamical rules. In the history of biology, they lay behind the everlasting debate on the relative roles of chance and determinism in evolution. Jacques Monod [1970] built a theory where chance (randomness) and determinism (order) were considered as two complementary aspects of life. In the present paper, we will give an up to date version of the problem going beyond the dichotomy between chance and determinism. To this end, we will first see how the view on living systems has evolved from the mechanistic one of the 19th century to the one stemming from the most recent literature, where they emerge as complex systems continuously evolving through multiple interactions among their components and with the surrounding environment. We will then report on the ever increasing evidence of "friendly" co-existence in living beings between a number of "variability generators", fixed by evolution, and the "spontaneous order" derived from interactions between components. We will propose that the "disorder" generated is "benevolent" because it allows living systems to rapidly adapt to changes in the environment by continuously changing, while keeping their internal harmony.*

Rivista di Biologia / Biology Forum 101 (2008), pp. 29-66.

1. INTRODUCTION

All through the history of Biology one of the most relevant sources of discussion has been the debate on the relative impact on living systems of chance and necessity. In this paper we shall propose a new vision of this critical dilemma derived from an up to date conception of the peculiar features of the "living state of matter". For this purpose we shall start by better defining the meaning, in living systems, of the two apparently opposite concepts.

At variance with non living systems, living ones can be considered as "individuals", as it is always possible to distinguish in them the "internal" from the "external".

All biological systems are composed of a number of elements, often compartmentalized and connected to one another, isolated from the exterior by some membrane, but continuously exchanging energy and matter with the environment. In the case of ecosystems, they are not limited by physical barriers but by the connections between different organisms. In other words all connected components belong to one system, organisms coming from "outside" being defined as invasive individuals and species.

Components within individual living systems (molecules, cells, organisms, populations, species, ecosystems) can assume a number of configurations.

The level of "order" of a system can be measured through the evaluation of the number of possible configurations and of their relative probabilities. Within this frame, only one configuration is allowed in wholly ordered system, maximum disorder being reached when all possible configurations have the same probability to occur. The most direct way to measure the level of constraints to randomness consists in estimating the ratio between maximum entropy corresponding to equi-probable configurations (total entropy according to Atlan [1979]) and the entropy observed in the systems under study (the observed entropy of Atlan). An observed entropy significantly lower than the maximum entropy signals that the different components of the system are correlated one with each other, i.e. there is a significant amount of constraints to randomness.

What is the role of order and randomness in the evolution of living systems? What is the functional meaning of constraints to randomness in the continuous fight for the permanence in the "living state of matter" (as previously defined in Buiatti [2000]; Buiatti *et al.* [2004]; Buiatti and Buiatti [2004])? These questions have always been at the core of biology, and since we, the humans, are indeed living systems, they have always had an impact on our general conception of life and of ourselves. This has provoked a permanent and often misleading interaction between science and ideology. On one hand biological data have been often used to support ideological concepts and socio-political choices. On the other, scientists have been themselves unconsciously influenced by the changing "spirit of times" ("ZeitGeist") acting as a filter or as a pair of eyeglasses to be used when observing nature.

For a long time, the "spirit of times", influenced by the industrial revolutions and a dominating deterministic ideology, has pushed for the attribution of a positive "value" only to constraints and "rules" or, in other terms to "seeds of order", considering variability as disturbing noise. This has led to a rather unfortunate bias in the discussion on the significance of randomness in life, often reduced to an irrational fight over a supposed order/disorder antinomy.

In this paper, we shall try to overcome this historical antinomy by showing that living systems are endowed both with ordered features and a relevant amount of plasticity (disorder), and that both components are fundamental for living systems to adapt to the surrounding environment. To this aim, we will focus in the first place on DNA, the hereditary material and therefore the "living memory" of biological dynamics.

We will first start with a brief overview of the history of the discussion on the order/disorder antinomy since the discovery of heredity laws by Mendel to our days. We will then propose a "solution" to this antinomy, based on the increasing experimental evidence of "friendly" co-existence in living systems between a number of "variability generators", fixed by evolution, and the "spontaneous order" derived from interactions between components.

2. A HISTORICAL REVIEW OF THE ORDER/RANDOMNESS DEBATE

1. *From Mendel to the "Central Dogma of Molecular Genetics"*

The relative contribution of order to the organization of living systems has been for the aforementioned reason the object of a long lasting debate particularly since the beginning of the industrial revolution and the explosion of "modern" science. The temporal connection between these two events should not be considered a mere coincidence. Human artifacts and particularly "machines" (the products of industry), are highly ordered systems, as their components can only assume the configuration allowed by the human project following which they have been constructed. Moreover, machines are an ordered assemblage of independent components in the sense that the removal of one of them does not have any effect on the structure and function of the others. In other words, a part of the machine as for instance a wheel in a car is exactly the same when assembled with the other components as it is before the assemblage. The situation is obviously different in the case of living systems where all components are parts of networks where the extent of non additive interactions depends, as we shall see later, on the number and the weight of links between any given part and the others. Now, one of the basic methodological innovations of "modern" science has been the simplification of a complex system into its components to be studied one by one, with the aim of obtaining the knowledge of the whole system through the mere assemblage of the information thus gained. The application of the method of "reduction" of the knowledge of a whole to that of its components has been successfully applied to non living systems since the 18th century in Physics and Chemistry but was officially introduced in Biology only in the second half of the 19th century when the "Manifesto of medical materialists" was published (1848) by a group whose leaders were Von Helmholtz, Du Bois Raymond, Brücke. The rational basis of the extension to life of the use of the reductionist method has been, coherently with the modern "spirit of time", the principle of substantial equivalence of living and non living system. This principle implies that

the laws of physics and chemistry are necessary and sufficient for the total understanding of living systems which may therefore be considered essentially as "living machines". It is implicit in this conception that Biology students should, whenever possible, dissect the systems to be investigated into components and analyze them, eventually aiming, as physicists and chemists, at the development of mathematical hopefully "universal" laws. The mathematization of life becomes therefore the ultimate goal of Biology as it had already happened since a long time in the case of other disciplines.

Now, in many cases mathematization of natural systems laws, has been developed with the aim of predicting system dynamics, an extremely useful exercise for humans, whose adaptation strategy is largely based on the modification of nature possibly without unwanted and unpredicted "collateral effects". Given the self-reproductive capacity of living systems, long term prediction can only be achieved through the discovery of mathematical laws of heritability. Certainly this has not been the conscious aim of Gregor Mendel when he started his experiments with peas. However, he was a physicist, studied in Wien and had some loose but relevant relationships with some of the founders of reductionism in Biology. With that cultural background (Orel [1983]), he decided to analyze in peas the behavior of single characters discrete variants (red or white color of the flower, rough or smooth seeds etc.), this being the best way to obtain the numerical ratios of the variants themselves appearing in subsequent generations of self-fertilization of a cross between "pure lines". These numbers may then open the way to the mathematization in probabilistic terms of heritable processes. The results obtained were perfectly adequate to Mendel's expectations and seemed to confirm also in the field of heredity the high heuristic power of the reductionist method. Mendel unfortunately never realized the real relevance of his data. The reason for his lack of understanding of the value of his own work was again his training as a physicist inducing him to look for universal mathematical laws, thus underestimating the complexity of his object of study. With the aim of understanding whether the laws developed for peas could be applied to other organisms and not being a biologist, Mendel submitted his results to the critical re-

view of a famous botanist of his period, von Naegeli. The answer of the "referee" was moderately positive but he suggested repeating the experiment on a plant different from the pea. Unfortunately von Naegeli was not aware of the fact that the plant proposed for the new experiments (Hyeracium) is not endowed with sexual reproduction. Therefore Mendel failed to obtain the same results found in the previous experiments, considered them as essentially irrelevant and unfortunately never got to know that he had started one of the most important revolutions ever happened in Biology. For this reason Mendel completely abandoned his studies of heredity and worked for the rest of his life on the behavior of bees and on the mathematical laws of winds dynamics, as can be found written on his grave. Mendel's work, although known to a fair number of scientists, was only reconsidered at the beginning of the twentieth century, his "laws" being confirmed almost at the same time by three biologists, a Dutch (Hugo De Vries), a German aristocrat living in the Austrian empire (Erich Tschermak von Seissenegg) and a Belgian (Carl Correns).

Let us now discuss with a little more detail the main concepts developed by the Moravian scientist. According to Mendel's results, heredity is controlled by discrete factors ("anlage") present each in two copies in every individual heritable complex. Due to what De Vries later called "mutations", two or more discrete variants ("alleles") of each factor can be found in each population and species. Alleles can be "dominant" or recessive and individuals "homozygotes" (bearing two copies of only one allele) or "heterozygotes" (with one copy of two alleles). As every factor ("gene") was thought to carry, in a wholly deterministic way, the information for one character, the genotype (the assemblage of the alleles and genes) of an organism would dictate without ambiguity the corresponding phenotype.

Now, if we call "A" the dominant allele of one character (for instance the red color of a flower) and "a" the recessive allele of the same gene (for instance white), a heterozygous genotype would be Aa, the two homozygous ones being AA and aa. For reproduction the organisms produce "gametes" (spermatozoa if males and ovules if females in animals) containing now only one copy of every gene and therefore only one allele. The ovules of heterozy-

gote individuals may then be either A or a and the same would hold true for the spermatozoa of a male. At fertilization one sperm and one ovule would meet and give rise to the "zygote" from which one member of the progeny would develop. Then, without interferences by other factors, the probability of a sperm or an ovule from our heterozygote containing A or a being the same (50%), the probabilities of the three possible genotypes of progeny members (AA, Aa, aa) would be respectively 1/4; 2/4; 1/4. Phenotypes, accordingly, given the dominance rule, will be then, in our case 3/4 dominant (red) and 1/4 recessive (white). Summarizing those results, the distribution of alleles to the progeny is for Mendel completely random, the phenotypes of each individual of the new generation being univocally determined by the randomly acquired genotypes. To be more precise, every individual of a species has the same gene complement (the same "functions") although in different variants (every pea has a flower color, white or red), random assortment referring to the alleles, that is to gene variants. In other words, the genetic complement of one individual is for Mendel and his followers, as envisaged by Ernst Mayr, like a "bean bag" (Mayr [1982]) from which alleles are randomly chosen and consequently an organism is supposed to be like a machine determined by one of the many equi-probable combinations of the "bean" variants deriving from parental genotypes.

It is also noteworthy that the alleles whose distributions have been analyzed by Mendel and the Mendelians had on purpose been chosen to be discrete (red or white, rough or smooth etc.) for the simple reason that discrete objects are easy to count in subsequent generations. Therefore all results and the derived theories did not take into account the possible existence of genes controlling quantitative characters showing a continuous distribution of values like for instance weight, height, length etc. This fact started a very harsh debate in the early years of the twentieth century between the Mendelians and the school of a British statistician, the Darwin relative Sir Francis Galton, an orthodox Darwinian and a student of quantitative characters. The debate was particularly centered on the relevance of Mendelian results for the understanding of evolution, the Galtonians strongly supporting the Darwinian view of evolution as a continuous process. The only compre-

hensive theory opposed to the Galtonian view was proposed by one of the "discoverers" of Mendel's work at the beginning of the twentieth century, Hugo De Vries, in his two-volume treatise, *Die Mutations Theorie,* first published in 1902. In his masterpiece he was the first to confirm in a large number of experiments Mendel's data and to extrapolate from them in a clear way the two concepts of major relevance for evolutionary dynamics of the Mendelian view, namely randomness and discreteness. He however, integrated Mendel's view with the new concept of "mutation" defined as a discrete "jump" from one allele to another.

According to De Vries, mutations were totally random events occurring all with the same frequency in time. Evolution therefore would derive from the linear accumulation of mutations with time without any interference of the environment. De Vries even wrote a simple linear regression function correlating mutation numbers with time, thus extending to evolution the mathematization started by Mendel with his probabilistic treatment of genes and characters behavior throughout generations. Evolutionary dynamics was then characterized as a cyclic occurrence of total "disorder" (maximum entropy) due to the random assortment of alleles in gamete production and fertilization followed by total order, the development during life cycles. Development was envisaged as a unique path of configurations changing from birth to death, completely determined "a priori" by the genotype. In synthesis, life processes were thought to be controlled by a blind "bricoleur" (Jacob [1970]) giving rise to random changes in otherwise deterministic machines.

This vision was in open contradiction with Darwin's conception which attributed evolution to three interacting factors: environment controlled selection, use and disuse of organs and correlated variation. This last factor introduced the very interesting concept of constraints to randomness deriving from the interaction between the parts of an organism and from the "rules" needed for the maintenance of a sufficient level of dynamic harmony between components. According to this concept, clearly stated by Darwin but not by neo-Darwinians, only mutations coherent with the desired amount of harmony would be compatible with life and therefore fixed by selection. The author of the "Origin", wrote in

1876 in "Variation in animals and plants under domestication":

> Although every variation is directly or indirectly caused by changes in external conditions, we should never forget that the nature of the organization on which we act is by far the most important factor influencing the result.

The Mendelian-Galtonian controversy on the two sets of opposite concepts chance/necessity and continuous/discrete was solved few decades after its inception, when a compromise was attempted by the "Modern synthesis" as it was defined in the forties of the twentieth century by Julian Huxley [1942]. In this theory, only one of the Darwinian factors of evolutionary dynamics was re-introduced, namely natural selection. According to the neo-Darwinists, selection operates on single genes and alleles, still believed to be as independent as Mayr's beans in a bag, the "best" alleles being chosen out of those present in a population. The choice is based on the relative reproductive capacity conferred to individuals by the different alleles. The reproductive capacity (average progeny size of individuals carrying a given allele) is called "fitness" and used as the only parameter for the evaluation of the level of adaptation. To take a classical example, in a population where the three possible genotypes (AA, Aa, aa) are present, phenotypes are either A or a due to the presence of dominance and, at least in the classical version of the theory where interaction between alleles was not taken into consideration, the fitness of all individuals with A phenotype is in principle supposed to be the same, and different from those with phenotype a. Starting from a very simple equation written by Hardy and Weinberg at the beginning of the twentieth century, describing the probabilities of the three genotypes in the absence of all possible factors interfering with reproduction in the so-called "equilibrium population", genotype frequencies in subsequent generations can be predicted once allele fitness values have been calculated. In this operation the environment is implicitly taken to be constant and to have therefore the same effect on selective levels throughout generations. Although fitness values are quantitative and the fitness function is implicitly assumed to be continuous with time, the effect of environment is wholly deterministic as fitness values are constant in time for each allele, the

phenotype being considered simply as a carrier, irrelevant for evolution. Implicit in this extreme view clearly stated in Fisher's "The genetic theory of selection", is the Mendelian principle of lack of "ambiguity" in the passage from gene to character. Again, as all calculations are carried out on single genes and the results extrapolated to the whole genotype, this is considered as an assemblage of independent particles, the Mendelian factors.

However, while selection is thought to be essentially deterministic, neo-Darwinism envisages the presence of two other factors of allele frequency changes due to pure random events. The first is mutation, defined as the change of A into a and of a into A, in an updated version of Hugo de Vries's concepts. Obviously, mutations will only modify gene frequencies if the probabilities of the two possible changes are different. The second purely random factor, called random drift, is essentially based on the occurrence of "sampling errors" in the relative proportions of alleles throughout generations. Let us take the simplest example of this process, namely the so-called "founder effect" or "migration". Suppose we analyze the behavior of a butterfly population where wing color is controlled by one gene and two alleles (A = red, a = white). Suppose furthermore that allele frequencies in the home area of this species where one million butterflies live, is 0.2 for A and, consequently 0.8 for a. Now, let's imagine that a very small group of 10 butterflies migrates into an island where no individual inter-fertile with our species exists. As every butterfly has two copies of the gene for wing color, they have all together 20 alleles. Relative numbers of A and a, to repeat the situation present in the home area, should then be 16a vs. 4A. This however happens only in theory, and only if allele frequencies in the original one million individuals population have been maintained in the migrant group. The probability to obtain the same result in the colonized island, however is very low as, due to the small dimensions of the migrant sample, small random deviations from the base value will significantly change future allele frequencies. Therefore, the general neo-Darwinian conception of evolution was based on stochastic determinism, mutations and random drift being fully stochastic processes, development during single life cycles being solely determined by heritable information.

Surprisingly, although all this evidence suggested that living systems were controlled by key heritable factors, the critical question of their material nature was not asked by geneticists. It was a physicist instead, Erwin Schrödinger, who, extremely intrigued by the results obtained by the newly born discipline of Genetics, set the ground for a new theory of life in a short series of lectures, assembled in a famous book by the exciting title *What is life?* published in 1942.

Being a physicist, Schrödinger tried to find analogies between his discipline and modern Biology. Within this frame, he assimilated the random mutation events to quantum jumps and at the same time offered a far-sighted explanation of the extraordinary level of complexity and order of an organism developing from a single and apparently simple object like a fertilized egg, using for its "organization" "disordered" matter. It was Schrödinger who proposed that living systems were able to absorb what he called "negative entropy", "extracting" it from the environment through the exchange of the molecules of food and energy used in metabolism. This way entropy (disorder) would only locally be maintained low, at the expenses of the context where, according to the second principle of thermodynamics, entropy would be still steadily increasing.

According to Schrödinger the whole process of the reproduction of the ordered organization of life ("order from order" as he wrote) had to be controlled by an "a-periodic crystal", the hereditary material, capable of self reproduction and liable to change with a low frequency due to quantum "jumps". The control would be achieved through a "code", a term with which Schrödinger introduced the concept of "genetic code", the basis of the "informational theory of life" (Buiatti [1998]) which has been dominating Biology ever since.

As Schrödinger wrote:

> The chromosomes, or probably only a fiber giving the fundamental backbone of what we see under the microscope as a chromosome, are the elements containing in a sort of encrypted code the entire design of the future development of the individual and of its functions at maturity……When we compare the chromosome fiber structures to a coded text, we mean that the universal

> mind mentioned by Laplace, to which every casual connection is immediately known, could tell from those structures whether the egg will develop, in given conditions, into a black cock or a freckled one, a fly, a corn plant, a beetle. …… The expression "code text" has, alone, a too narrow meaning. Chromosome structures are, at the same time, the tools used to achieve the development of which they are symbols. They are coded laws with extensive power or, to use a different metaphor, they are the project architect and at the same time skilled builders.

For Schrödinger, the unique nature of life, different from other physical systems, is defined by the "order from order" process, deterministic and based on the instruction contained in the "aperiodic solid".

As he says:

> The physicist knows that the laws of classical physics have been modified by quantum theory particularly at low temperatures. There are many examples of this fact. Life seems to be governed by a behavior of matter, ordered and controlled by exact laws, not based only on its tendency to go from order to disorder but, partially, on the conservation of order.

So far Schrödinger, who left to "biochemists" and other biologists the job of the isolation of the "aperiodic crystal" and of understanding the processes leading to the absorption of "negative entropy" (of order).

We have been discussing and quoting at length Schrödinger because his thoughts had and still have a relevant influence on Biology. Three main concepts, proposed by Schrödinger, have in fact dominated modern Biology for many decades. The first is the relevance of order in living systems. In Schrödinger's theory, an essentially negative meaning had been attributed to randomness as in the case of mutations, whose frequency, according to Schrödinger himself, has to be maintained low to avoid a too high level of instability of the systems. This concept of a negative value of variation, present in Mendelian genetics as in the neo-Darwinian "Modern synthesis", is fully coherent with the double helix unique structure of DNA called later the "fundamental invariant" by Jacques Monod, representing the transmissible key molecule of life

"necessity". Also in Monod's book, probably the most complete "modern synthesis" of molecular genetics, chance (randomness, noise) was attributed only to mutations, in general rather rare and with negative effects, and to the assortment of alleles at reproduction as foreseen by the Mendelians. The two other concepts which changed Biology have been the notions of "code" and "information", both related to DNA sequences, whose knowledge should allow the complete prediction of the phenotype. From this conceptual "corpus" the "informational metaphor" (Buiatti [1998a, b]; Gagliasso [2000]) and the "reductionist ideology" were later derived (Buiatti [2004]).

The apex of the success of this very powerful theoretical system can be considered to have been reached by the proposal of the so called "Central dogma of molecular genetics" describing the flow of the information needed for the self construction of organisms, from DNA to characters. This flow, neatly described in a paper in *Nature* by Francis Crick [1958], is considered to be devoid of mistakes and ambiguity and thought to proceed from DNA to a similar molecule, the RNA, through direct transcription, both being nucleic acids using a four letter (A, T, G, C) code organized into triplets. The message is then "translated" into proteins, molecules "written" using a 20 letter code of amino-acids. The translation uses 61 out of the available 64 triplet combinations of DNA letters, the remaining three being used as "stop" signals for reading. The fact that the code has been deciphered three years after the "Central dogma" does not change the value of that proposal but, rather, helps understanding how it works.

It is worth noting here that not all the concepts contained in Schrödinger's theory necessarily lead to a deterministic view of the "living state of matter". Particularly, negentropy by itself does not imply any choice between total order and disorder.

This concept has been utilized by whole sectors of interdisciplinary studies involving at the same time physicists and biologists, and has opened the way to a series of significant achievements in the understanding of self-organization processes of "out of equilibrium systems", both living and not living (see Nicolis and Prigogine [1977], Buiatti [2000], Kauffman [1995], Strogatz [2003]). This is the area of investigation of contemporary non linear Phys-

ics whose influence on Biology will be the subject of our next paragraphs. Another difference between Schrödinger's view and Modern Biology is his firm opinion that living systems, although obeying the general laws of Physics, are also regulated by some specific biological laws like those related with the "order from order" and negative entropy dynamics concepts. This is the reason why he was also aware that Physics and Chemistry alone would not be able to completely explain the behavior and dynamics of life, thus implicitly suggesting the existence of a "living state of matter" (Buiatti [2001], Buiatti and Buiatti [2004], Buiatti *et al.* [2004]).

2. *The "Phenotype Paradigm" and the Theoretical Challenge to the Chance/necessity Antinomy*

As mentioned before, classical neo-Darwinists and the molecular biologists of the central dogma attributed only a minor role, in the history of organisms, to adaptation during their life cycles. The rationale behind the conceptual removal of any role of the phenotype in evolution was based on the concepts of gene independence and lack of ambiguity as stated in the central dogma, both implying the view of the phenotype itself as a single, unambiguous product, of genotype reading without any influence of the external and internal contexts. It is worth reminding that most of the often complex mathematical treatments of population dynamics have been based on the analysis and simulations of changes of frequencies of single gene alleles, consequent rules and "laws" deriving from the extrapolation to whole genomes, envisaged as resulting from the sum of heritable components. Although, however, the conception represented by the Central Dogma has been largely prevalent in the twentieth century, the premises for the future, dramatic changes occurred in the last twenty years in contemporary Biology, were set by a number of scientists already in the days when the "Modern synthesis" was built, and a new vision of the living state of matter, although often disregarded, was developed.

The first field in which a role of environmental effects and gene interactions has been introduced and discussed is a nowadays almost forgotten "sub-discipline" of population genetics, quantita-

tive genetics, initially based on the discovery of poly-gene systems controlling quantitative characters by Johannsen and Nilsson-Ehle. More or less in the same years when the general neo-Darwinist theory was developed essentially by three geneticists (J.B.S. Haldane, R.A. Fisher, S. Wright), a solution was also formulated for the other aspect of the Darwinian-Mendelian controversy, the discrete vs. continuous problem. The compromise was reached when Nilsson-Ehle [1908], working on a quantitative character in barley (seed color), demonstrated that a continuous distribution of color intensities in a plant population could be explained with the segregation of a number of Mendelian genes, all giving a positive or negative quantitative contribution to the chosen character. Color expression according to this hypothesis was determined by the numbers of "colored" alleles present in each genotype in a linear fashion. At the beginning of the new discipline, quantitative genetics, it looked as genetic determinism had won another key battle and that life remained, as in the early years of the century, mechanistic and controlled by chance and necessity. It is worth noting here that the success of quantitative genetics was due mainly to the need of new biometrical methods for the prediction of the effects of intra-specific crossing and artificial selection in domestic animals and plants for productive reasons. Inevitably however, this goal directed the interest of geneticists to the phenotype and its dynamics during life cycles. The experimental observation of the dynamics of production relevant characters in subsequent generations showed that the prediction of phenotype productivity of individuals on the basis of the performances of the parents was not reliable as expected due to the effects of the environment. This observation led to the development of a statistical parameter called "heritability" used to estimate the relative impact of genetic variability on phenotypic variation. This parameter, generally based on the use of a hierarchical analysis of variance as well as on parent-offspring regressions and correlations, was found to provide a good index of predictability of the phenotypic outcome of genotypes. Conceptually, all this means that single genotypes may lead to more phenotypes each moving during its life along a path, dependent on the environment and its modifications, within a phase-space whose limits are set by the genotype and its plasticity (capac-

ity of phenotypic change).

It was also from the genetics of quantitative characters that emerged another relevant concept, a new version of the Darwinian correlated variation.

The Darwinian need, for adaptation and survival, of a certain level of harmony between the parts of an organism, was proposed again by quantitative geneticists when they found out that in many cases selection for one character (for instance meat production in cows) would have negative effects on another (milk production) while in others the interaction was found to be positive. Fine statistical tools were therefore developed by breeders to measure the relative impact of genotype on correlations allowing modulation of selection pressure to avoid negative side effects and efficiently exploit positive ones.

While the interactions between parts of a living network and between a living system and the environment were the main discoveries of quantitative genetics, the relevance of non additive interaction between alleles was introduced by the observation in maize and, later on, in other out-crossing plants and in animals, that individual "vigor" was highly enhanced in heterozygous organisms derived from crosses between highly homozygous parental lines. The phenomenon was called "heterosis" and its discovery led to an abrupt change in plant and animal breeding methodologies until then aiming at the development through selection of "optimal", highly homozygous individuals putatively containing in a "pure" state the "good" alleles. It was the complete failure of a selection program based on self-fertilization in maize that showed the need to maintain high heterozygosity levels for the improvement of production. Seen with contemporary eyes, this has probably been the first observation of a positive value of variability. The concept was further supported by the innovative work by Michael Lerner [1954], one of the most prominent scientists of neo-Darwinism but also involved in animal breeding, who showed that hybrid individuals were more stable, during their life cycles, than the more homozygous counterparts. This result was interpreted by Lerner in his book *Genetic Homeostasis* [1954] as due to a sort of "buffering capacity" of variability at the gene level favoring individuals carrying different alleles in a relevant number

of genes. In other words, the presence of two different forms of a gene would increase the homeostatic capacity of individuals thereby enhancing their phenotypic vigor and fitness. As, obviously, in a given population heterozygosity would be higher in the presence of a high general level of genetic variation, the natural extension of the new concept to evolution would lead to the suggestion of a positive value for high variability also at that level. The concept of "Genetic homeostasis", the term coined by Lerner for this process, has later been one of the ground tenets which induced a number of scientists like Richard Lewontin [1972] to introduce the concept of "balanced selection" in Population Genetics emphasizing the positive value of high amounts of genetic variability. This particular development, took place almost twenty years later of Lerner's initial work and was also supported by a large number of experimental data obtained with new biochemical methods of protein analysis contradicting the classical neo-Darwinist hypothesis. According to that conception, natural selection should lead to highly homozygous genotypes containing only the "optimal" allele. Therefore, one would expect to find in present day species, coming from million years of natural selection, a very low remaining genetic variability. The opposite was found to be true with the new methods as most enzymes were found to be present in the same organism in more than one form. Lewontin and his followers interpreted this result as deriving from positive interactions between alleles giving a selective advantage to individuals carrying both of them, homozygotes for each allele showing a lower fitness. The concept of interaction had been extended to whole genotypes by Sewall Wright, one of the founders of neo-Darwinism, who, in a famous debate with R.A. Fisher, proposed them as the unit of selection, clearly stating that fitness values cannot be calculated on single genes as they derive from the effect of the non additive combination of many if not all the genes of a genotype. He then coined the term "genetic landscape" whose graphic representation was an artificial landscape (a phase space) with fitness values attributed to each of the many possible gene combinations. In such a landscape, more than one combination can be optimal in terms of fitness, and the value of this parameter changes according to the environmental context.

A completely different explanation of the unpredicted high genetic variability in populations was offered by a group of Japanese workers led by M. Kimura [1961]. The Japanese interpreted the high variability observed as being due to the lack of selective forces acting on a complex set of completely "neutral" variants. The Japanese "Neutral theory of Evolution" was a new version of the previously discussed "Mutation theory" of H. de Vries, as Kimura and his followers, at least in their first period of work, envisaged again evolution as strictly happening only through mutation and random drift. The role of the phenotype in this case was not considered, organisms being fully determined by independent genes once they had been randomly assorted. Needless to say, Kimura's theory was in open contradiction with the aforementioned data of quantitative genetics, not only for what regards the roles of environment, phenotypic variation and selection, but for the refusal of the concept of interaction itself.

Probably the single person who better described and discussed the living systems as being in a state somewhat intermediate between order and variation has been Conrad Hal Waddington. He was an embryologist by formation, with a strong theoretical tendency and interest in mathematics as a source of tools for the description and interpretation of living systems. In the period 1940-1975, when, due to the successful use of the reductionist method and to the fast improvement of experimental tools, the pace of accumulation of analytical data had become very fast, he proposed what he called (Waddington [1959]) the foundation of "synthetic Biology" aiming at developing "a coherent intellectual picture of the whole kingdom of life".

The most relevant achievements of Waddington's synthetic Biology program are the results of a series of interdisciplinary meetings which produced four volumes by the title "Towards a theoretical biology" from 1969 to 1972. Many of the concepts and ideas in Biology proposed at that time were anticipatory of the discoveries of the last fifteen years or so and resulted from intense and fruitful collaboration and discussions between geneticists, physicists, embryologists, and mathematicians. It was through these discussions that Waddington developed what he called the "Phenotype paradigma" aimed at complementing the neo-Darwin-

ist view and at a better characterization of the specific character of the "living matter".

He started considering that "one of the most basic characters of living systems, considered as active machines, is the fact that they produce a local enhancement of order through the assumption of simple molecules and their organization into complex compounds, orderly structured. At first sight one could think that those systems act in opposition to the second law". Although surprisingly Waddington very seldom quoted Schrödinger, this statement is fully coherent with the "negentropy" theory proposed by the physicist. However, Schrödinger's concepts are further implemented by Waddington with a thorough discussion on the real value of the application of information's theory to life.

His attitude was highly critical towards the direct use of information theory particularly in the case of information transfer from DNA to protein and to characters during development:

> In this case we find that Biology has developed more flexible mechanisms than those used by telephone technicians. There is a system through which information can be changed in the sense that a gene can mutate so that what the progeny receives is not exactly what was present in the parents… It's in the later phases of phenotype formation that information theory fails to deal with the situation. It is obvious that one organism's phenotype is not simply a collection of all the proteins corresponding to all the genes of the genotype. It is true that the phenotype is the result of the assemblage of parts highly heterogeneous, in each of which some but not all the proteins are present which would derive from the genes, but also many structures and substrates other than primary proteins corresponding to specific genes …. It appears completely obvious following the "common sense" that a rabbit running in a field contains an amount of "variation" higher than that contained in a rabbit's egg immediately after fertilization. Now, how can we explain such a situation within the frame of information theory whose fundamental principle states that information cannot be increased… in fact the main problem is that throughout the passage from a zygote to an adult organism, "information" is not simply transcribed and translated but works in terms of "instruction" or, to say it in a more elegant manner, of "algorithms". DNA produces RNA which in turn builds a protein.

> The protein then affects the surrounding components in different ways, whose result is the production of a higher variety of molecules than that present at the beginning".

Of course, this is possible, as observed by many authors, and also by Waddington himself, because living systems are open in the sense that they continuously exchange energy and matter with the context around them. The number, quality and configurations of the components of biological systems are continuously modified following complex non additive interaction processes within the system itself and with the ever changing environment. Complex, non linear functions are needed to describe organismal paths of self-organization based on the Waddingtonian principle of homeorrhesis, that is the dynamic counterpart of homeostasis or, from the Greek, a flow in which the system maintains its "self". The paths followed by single individuals have been called by the British scientist "chreods", each of them being one of the potential life cycles of the organisms allowed by the "inherited tools" and by the contexts. While homeostasis can be defined as the capacity of a system to maintain constant its main features, responding through a series of buffering processes to the noise coming from the external context, homeorrhesis is its dynamical version, that is the property of living beings to avoid diverting their paths from the system-specific attractor through a continuous exercise of adaptation through change. While an organism absorbs "order" from the context, as foreseen by Schrödinger, it also increases disorder/information all through its life cycle.

Seen this way, developmental but also evolutionary paths do not seem directed in any way towards the stable acquisition of an optimal order as it happens with crystals and human-made machines. Rather, they appear to obey to homeorrhetic laws maintaining specific equilibria between order and change within the frame of the requirements of the second principle of thermodynamics.

Implicit in the non additive nature of the interactions among components of living systems and between organisms and the environment, is the presence of non linear components in functions describing them. It is worth noting that a relevant role in this re-

spect, has been played by the friendship and cooperation between Waddington and René Thom. The French mathematician described from a topological point of view the dynamics of developmental evolutionary processes on the basis of Waddington's biological suggestions, thus rendering the whole theory clearer and inducing a new interest in the use of mathematics in biological studies beyond the probability theory and statistics dear to many neo-Darwinists. Particularly, the reciprocal exchange of views and data between the developmental geneticist and the mathematician had probably an influence on the development by Thom of the so-called "Catastrophe Theory" and the introduction of the notion of bifurcation, describing sudden changes in systems reaching a critical point, directly reminiscent of chaos theory. A well known heuristic representation of these concepts by Waddington and Thom can be seen in the "epigenetic landscape", i.e. the attractor within which a living system moves in development and evolution. Such a landscape shows the minima in which a path will be running, allowing the "jump" from one valley to another when the dynamic system reaches a bifurcation point.

It is worth noting that, while Waddington and the relatively small group of theoretical biologists who followed him seemed to be rather isolated from the mainstream of experimental life sciences, their conceptual framework was enriched, in the seventies and in the eighties of the twentieth century, by the work of physicists and mathematicians like those of the groups led by Prigogine (see Prigogine and Nicolis [1972]), Henry Atlan (Atlan [1979]), Maturana and Varela [1980], Goodwin, and many others. A particularly interesting result of their work was the extension of Schrödinger's intuitions and the demonstration and application to life of the concept of self-organization. Prigogine's work was based on the observation coming from chemistry, that systems out of equilibrium, or, in other words, embedded in a matter/energy flow, can undergo a transition to order which is maintained until the recovery of an equilibrium state or a chaotic transition. A first mathematical modeling of such systems goes back to Alan Turing [1952] but has been carried out much further by Prigogine and his co-workers who introduced and modeled the concept of dissipative systems and structures. The concept of self-organization, implicit

in Waddington's work, had been introduced quite early by J. Needham in his famous book "Order and life" and was later taken up by Waddington himself and the students of biological complex systems. It was then suggested that dynamical interactions among components may be the main source of constraints, keeping the variation of systems during development and evolution within the attractors as defined earlier on in this paper. A particular attention was later on given to this area of thought by the whole group of Santa Fe starting from S. Kauffman's work (Kauffman [1995]). Kauffman's approach is essentially based on simulations of the dynamics of N/K Boolean systems, containing N components with K connections. The simulations have be implemented by introducing connection functions and the result of computer simulations can be treated with statistical methods describing the dynamics of fluctuations. Kauffman's simulations generally did not involve experimental validation. Recently, a different approach to the study of what has been called "spontaneous order" involving "Sync" (synchronous behavior) has been used (see Strogatz [2003]). This new way of looking at self-organization of living matter was introduced by a pioneering work by Watts and Strogatz (Watts and Strogatz [1998], [1999]) and is based on the study of the networks formed by the interactions among the components of real living systems. According to Strogatz, their research had been prompted by the observation of the mixture of order and randomness of web sites network and of the strange way of spreading of infection diseases.

Strogatz and Watts and almost at the same time the group of Barabasi (Barabasi [2000]) started looking at a large amount of heterogeneous networks studying the distribution of numbers of connections of the different components. The networks analyzed included a variety of examples such as the systems of actors present contemporarily in movies, of genes in the genotypes of a worm and a yeast, nervous systems etc. and they surprisingly showed always the same basic structure. All these networks comprise a very high number of components each connected with others with a variable number of links. When one studies the shape of the curve describing the statistical distribution of the number of links per node, he will always find that it is not Poisson, as it would be ex-

pected for random interactions, but a power law, denoting a scale-free structure of the network. This implies that these networks have a very inhomogeneous distribution of links, allowing the simultaneous presence of a few nodes linked to many other nodes and a large number of poorly connected elements (Boccaletti *et al.* [2006]). Such scale-free structure allows a fast information flow even between distantly related components, because the number of links between any two components is always very low.

A parallel current of thought (H. Atlan, J. Wicken, J. Brooks and J. Wiley etc.) looked at the problem from another point of view, namely the effect of variation and constraints on entropy functions. Thus, for instance, H. Atlan, proposed to consider the calculated "observed" Shannon entropy of experimentally analysed biological systems as the result of constraints determined by self-organisation processes reducing the "potential" entropy expected in the case of a totally random distribution of the n components) These concepts contradicted the false alternative between chance and necessity in an explicit way and opened the way to a very fruitful area of studies aiming at better statistically defining the "living state of matter" (Buiatti and Buiatti [2004]) and understanding the functional dynamics of constraints.

All the just discussed interdisciplinary work along with earlier theories such as those by Lewontin, Lerner, Waddington, Wright and many others including those implicitly derived from quantitative genetics, have been often emarginated from the mainstream of discussion in experimental Biology until the end of the last decade of the twentieth century and the beginning of the third millennium. Surprisingly, the debate on self-organization, and in general on the just discussed concepts was, on the other hand, very wide in non biological areas such as those of mathematics, physics, philosophy and involved also a part of the large public, interested in the problem of the level of predictability of the dynamics of complex systems. Unfortunately, a large majority of hard core biologists for a long time either did not read or psychologically removed the major inputs by physicists and mathematicians working in the area of non linear dynamics. The experimental area was in those times completely dominated by what we would call the "Central dogma ideology", whose biased interpretation induced

life science students to privilege studies of the "ordered side" of life and led to the interpretation of variations as disrupting white-noise.

As it happens in science, however, the experimentalists had soon to cope with the extraordinary amount of data coming from the use of the powerful reductionist method often showing an increasing amount of "exceptions" to the common paradigm.

In the next paragraph therefore we will try to summarize some relevant achievements of the new, contemporary, Biology, certainly not with the aim to build a new paradigm but to update the state of the art of the present vision of the "living state of matter" (Buiatti and Buiatti [2004]). Particularly we shall report on the ever increasing evidence of "friendly" co-existence in living beings of a number of "variability generators", fixed by evolution, and the "spontaneous order" derived from interactions between components. We will then discuss the thesis that the "disorder" generated but contained is "benevolent" in the sense that it allows the maintenance of the connections necessary for life, contrasting the effects of the disrupting internal and external random noise.

3. THE CONTEMPORARY REVOLUTION IN BIOLOGY AND "THE BENEVOLENT DISORDER"

Before entering into the discussion of recent experimental data leading in our opinion to the solution of the chance vs. necessity antinomy, we shall introduce a few concepts that we consider relevant to the comprehension of the general framework of the debate on the "living state of matter" of the last decade of the XXth century and the beginning of the third millennium.

In the first place, recent data show that different groups of organisms, namely prokaryotes, plants and animals and humans display different adaptation strategies although they are all "explorative".

The term "Explorative strategies" has been proposed in 1997 by Gerhart and Kirchner meaning by that adaptive strategies based on the exploration of variability generated by specific "generators" (Buiatti *et al.* [2004], Buiatti and Buiatti [2004]) and on selective processes, developed during evolution, allowing the "choice" of the

specific variants needed to answer to incoming, disruptive noise and in general to changes in the context. In all cases therefore, a certain level of "disorder" (variability) is needed but shall always be limited by constraints due to the need of a certain level of "internal harmony" in living dynamical networks. The variability induced by generators can be heritable or limited to the phenotype. Heritable variation, as proposed and discussed at length in Jablonka and Lamb [2004] may be genetic, epigenetic, behavioural, symbolic. In all cases heritable variation of populations, species, and ecosystems will contribute to a measurable parameter also proposed by Gerhard and Kirschner called "Evolvability" or, following Earl and Deem, "propensity of evolve" (Earl and Deem [2004]), that is the capacity to adapt throughout generation through structure-function changes. Evolvability levels are determined by inborn processes leading to genetic variation fixed through evolution for their value "per se" as shown by the fact that variations increasing this parameter bear in themselves a selective advantage.

However, epigenetic variation, is, for animals and plants the major source of phenotypic adaptation and can be in some cases transmitted to the progeny without changes in DNA sequences.

The term "epigenesis" has been introduced, as mentioned earlier, by C.H. Waddington, and epigenetic plasticity in a very broad definition, includes all processes allowing variation during life cycles not already present at birth as information contained in the DNA. As shown by several authors also plasticity can be selected as such thus increasing adaptation capacity based on the development of a series of tools for the exploitation of "ambiguity" (more than one meaning of information stored in nucleic acids and proteins, utilized according to different codes).

Finally, as discussed before, it is now universally accepted that living systems are dynamic networks endowed with different levels of *robustness* (the capacity of a system to resist to incoming random noise through the use of its internal plasticity), *redundancy* (the presence of more than one copy of living system components) and *vicariance* (the presence of alternative pathways allowing the achievement of the same result). The very notion of network immediately recalls the Darwinian intuition partially confirmed by quantitative genetics, of the presence of correlations between the

parts of an organism (the correlated variation). The meaning of the Darwinian principle was that the parts are not independent one from another in the sense that none of them is free to assume all the "configurations" possible (sizes, weight, shapes and so on) but that only those changes are allowed which do not harm the overall "harmony" of the system. To translate this concept into modern terms, we can simply extend it to all components of living systems at all levels of the hierarchical organization of life. Communication within networks is in fact the reason why not all the variability induced by the aforementioned "generators" can be utilized by living systems, the "benevolent" part of it being composed of elements harmonically inserted in the communication network. One may recall here the concept of "internal selection" proposed by Bateson (Bateson [1980]). By internal selection he meant that, as the components of living systems are part of a connected network with its harmonic quantitative and qualitative rules, even a single mutation may damage not only the affected component but a number of others connected with it thus eventually decreasing the fitness of the organism.

Having thus defined some basic interdisciplinary concepts of contemporary Biology, let us now start discussing, using the existing evidence, our proposal of solution of the chance vs. necessity antinomy based on the contemporary presence of variability on one hand, and the choice of different portions of it on the other, through communication between components of living systems and with the environment. In other words, in our opinion, life is maintained by ever changing equilibria between the "benevolent order" and the needed amount and quality of the "benevolent disorder". We will now provide evidence on how benevolent order and benevolent disorder emerge in DNA, the hereditary material and therefore the "living memory" of biological dynamics, liable to be statistically characterized as a string of information. This is now possible as whole genome sequencing has being carried out in hundreds of eukaryote and prokaryote species and has completely changed our conception of genome structures and functions. Genome size has been shown to vary from a few thousand to billions of nucleotides and to be much smaller in prokaryotes (in the order of millions) than in eukaryotes (in the order of billions). The in-

crease in DNA amounts in plants and animals moreover has been shown to be due to a far higher amount of non-coding sequences than coding ones. We know now that coding sequences cover 80% or more of prokaryote genomes but less than 2% for instance in the humans. A bacterium like *E. coli* has about 4000 genes, in a genome of four millions while the human genomes only have 23000 according to the last estimates although their total DNA content amounts to more than 3 billions nucleotides. So, there is in this case an inverse relationship between the complexity of the organisms and the ratio coding/non coding DNA.

Before providing evidence on generators of "benevolent disorder", we first focus on the widespread 'benevolent order' in non-coding DNA sequences, since as we will explain later on, the two benevolent components are probably tightly linked. "Order" in non-coding DNA sequences emerges from several studies on the extent of constraints to randomness in DNA strings. As it often happened in the history of life sciences, physicists and mathematicians involved in the general discussion on this problem in nature were the first to enter in this research field using a range of analytical methods such as the study of periodicities, short and long range correlations, or information theory measures (Li [1997]). Interestingly, the picture coming from the results obtained with the different methods used in the analysis of the fast increasing DNA sequences available from gene Banks is quite coherent. In general, constraints to randomness are higher in eukaryotes than in prokaryotes and, particularly in the first class of organisms, in non coding sequences than in coding ones. While until the end of last century the prevalent opinion was that non coding DNA was thought to be devoid of any significant functional meaning and for this reason was called "junk DNA", the analyses of whole genomes transcribed RNA products (the "transcriptome") carried out in the first years of the third millennium have shown that a large part of non coding DNA including repeated sequences is indeed transcribed into RNA and therefore has to be considered functionally active. The fact that the distribution of nucleotides is far from randomness in eukaryotic non coding DNA can be attributed to several experimentally observed causes. As discussed more at length in Buiatti *et al.* [2004], eukaryote genomes contain large numbers of

short and long sequences present in many copies (repeated sequences) whose function is often as yet largely unknown. These are generally classified into long and short repeats according to their length. A large part of the long repeated sequences in eukaryotes are "transposons", the mobile elements discovered in 1951 by Barbara Mc Clintock in maize, and afterwards forgotten for a very long time, mainly due to the contradiction between their existence and the dogma of DNA stability We know now that they often cover more than 50% of eukaryotic genomes and are responsible, for instance in maize, for "catastrophic events" of genome shuffling recently discovered (Brunner *et al.* [2005]). Short repeated sequences are composed of a variable number of identical nucleotides, of purine or pyrimidine, "strong" (G, C) or "weak" (A, T) nucleotides, or of repeats of 1-7 nucleotides called "microsatellites". All these sequences have been called "homogeneous" or "low complexity" and their distribution in the genomes has been found not to be random (Buiatti *et al.* [2004]). Moreover, again in eukaryotes, genes themselves are not randomly distributed along the chromosomes but concentrated in specific regions, and the GC/AT ratio is heterogeneous as shown long time ago by Bernardi and his group (Bernardi *et al.* [1985]), who called the homogeneous areas "isochores". Finally, regions upstream and downstream of eukaryotic exons (the eukaryote coding sequences), contain specific "motifs" highly conserved throughout evolution, often at regular distances from the beginning or the end of transcription units.

The main question to be asked at this point is whether or not constraints to randomness found in the non-coding eukaryotic sequences are "signaling" specific functions, and therefore their nucleotide composition and order acts as a sort of "hidden code", as suggested by Trifonov [1998]. We will now describe experimental data suggesting that this is probably often the case, and that many of these functions are generators of "benevolent disorder" or tools for choosing the portion of variation to be utilized for adaptation in different contexts.

Before discussing these data, however, it should be noted that the different amounts and distributions of constraints to randomness in prokaryotes and eukaryotes may be related to the different adaptation strategies adopted by the two groups of organisms. We

now know that while in prokaryotes adaptation mainly relies on genetic variability, processes leading to phenotypic plasticity are by far more frequent in plants and animals. The different strategy seems to derive from the very structures/functions of the prokaryotes and eukaryotes. Particularly, bacteria have, as individuals, very short life cycles and are haploid, that is endowed with only one copy of each gene. Now, short life cycle means a very high capacity to spread in a bacterial population a mutation conferring selective advantage. This process is also easier because of the haploid nature of the bacterial genome, allowing the immediate expression of the new allele, independently of its recessive or dominant nature. On the other hand, eukaryotes are endowed with genomes at higher ploidy levels, have by far longer lives, and therefore need several long-lasting generations for a favorable, new mutation to be expressed and fixed in a population.

Consequently, bacteria have developed during evolution a number of very sophisticated processes for the exchanges of significant portions of genomes both at the intra-specific and at the inter-specific levels. In fact, exchanges are so frequent that the very concept of species is rightly being questioned in the case of microorganisms. Moreover, bacteria are also endowed, as shown recently (Lombardo *et al.* [2004]), with systems for stress perception and consequent activation of "mutator" genes increasing mutation frequencies and therefore the probability of occurrence of mutations favorable for response to stress. This is the case of the so-called RpOs system: upon stress, the RpOs gene is activated, inducing the expression of mutator genes like those coding for DNA polymerase IV, an inducer of DNA replication with mistakes. In other words, in bacteria, a process has been developed during evolution through selection leading not to the fixation of specific mutations, but increasing the probability of obtaining them through higher mutation frequencies (new "benevolent disorder" available).

It should be noted that in this case, the process of variability generation is initiated by recognition of stress signals by the cell and eventually by the non-coding DNA sequences upstream of the RpOs gene. RpOs then produces a protein which behaves like a signal to the mutator genes again through recognition. Recogni-

tion is necessary for the activation of all DNA functions, both in prokaryotes and eukaryotes. For instance, both replication and transcription are induced and regulated by signals, whose recognition generally occurs through the formation of complexes with DNA in well localized positions of the string. Complex formation occurs whenever local DNA conformations are complementary with signaling proteins "shapes" and charge distribution, complementarity being the limiting factor of recognition at all levels in living systems from the enzyme-substrate complexes, to RNA-protein, RNA-RNA, RNA-DNA, protein-DNA, protein-protein complex formation. All this happens because all macromolecules, DNA included, are endowed with "conformation landscapes", that is the "ensemble" of possible conformations that a molecule can assume, including those placed at "energy minima". So, DNA, far from being always a neat double Watson Crick helix, can have many shapes each of which being determined by the specific DNA sequence composition, by the physico-chemical context, by possible pre-existing complexes with other molecules of all sorts. And, indeed, regulation occurs because many local conformations and the relative local sequences have been selected for, to allow localized and function-specific complex formation. In other words, at least a part of DNA sequences constraints (deviations from randomness) are determined by the functional role of complex formation with other molecules equally constrained. In the case of RpOs, constraints in DNA sequences leading to recognition are needed for the release of adaptive "benevolent disorder".

In eukaryotes, although genetic variability is certainly a tool for long term evolution at the population and species level, adaptation during life cycles relies on phenotypic plasticity, a part of which, however is due to somatic genetic changes. It has been shown, in fact, that mutation frequencies are not random in eukaryote genomes, being higher in the already mentioned "low complexity sequences", and that these sequences are often localized in parts of the genomes where high mutation rates are requested. This is the case, for instance of the sequences coding for antigen/antibody recognition in immuno-globulin genes, which mutate by shuffling of repeated tracts during cell development. This process leads to the production of an extremely high number of different antibodies in single

individuals conferring the capacity to recognize an extraordinary number of "enemies" (antigens) (see for instance Lewin [2007]). Moreover, it is known that transposons also in eukaryotes may be activated in the presence of stress (Madlung and Comai [2004]) and will thus increase somatic mutation frequencies, by "jumping" from one part to another of the genome, thus determining shuffling of fairly long sequences. All these processes, generally occurring at the somatic level in eukaryotes, are examples of the apparent paradox of constraints to randomness in DNA sequences leading to increases in variability.

However, the main processes leading to phenotypic plasticity in eukaryotes are the differential activation and modulation of gene sets in different internal and external contexts and gene "ambiguity" obtained through signal recognition.

To give an idea of the levels of plasticity that can be reached by gene differential activation and modulation, humans are endowed with about 23.000 genes, around 2000 of which are thought to be active in every single cell. However different cells, differently located in the human body "use" different gene combinations and the same is true for single cells in different periods of their life cycles. Moreover, gene activity far from being a binary, yes or no, process is modulated through changes in the stability of the so-called "transcription complex" between DNA and a large number of proteins. To obtain a stable transcription complex DNA has to curve, twist and roll at specific points, these conformation changes depending on DNA flexibility, only possible in the presence of specific sequences in the right positions along the string. Activation patterns can be transient or stable for even long periods of time. The stabilization mechanism generally involves methylation or de-acetylation processes induced by one or more protein-DNA recognition events through shape complementarity. Gene ambiguity, that is synthesis of more than one protein from single genes, on the other hand derives from the use of different starting and ending sites in the DNA "reading" process as in the case of the so-called "overlapping genes" discovered in prokaryotes but much more frequent in eukaryotes, or through "alternative-splicing" of transcribed RNA. The last process occurs only in eukaryotes where genes are discontinuous in the sense that they are endowed both

with coding (exons) and non coding (introns) sequences.

Gene expression starts with the transcription into RNA of the whole sequence but later transcribed introns are dismissed and destroyed, exons being assembled to yield the "mature DNA" on the basis of which proteins will be synthesized. However, exons can be assembled in different orders leading to a number of different sequences and, consequently, to the synthesis of more than one protein. Recent data suggest that an average of 10-15 proteins per gene are synthesized on the basis of the 23,000 human genes but the ambiguity level varies according to protein function. For instance neurexins, proteins contributing to the connection between human neurons, are more then 2000, but they are coded by only three genes; similarly, in *Drosophyla* only one gene has been shown to produce 38000 different proteins. A higher ambiguity capacity in proteins of the human brain is coherent with the general human adaptation strategy mainly based on "inventions". In all these processes, the "choice" of the proteins to be produced in every spatially and temporally variable context relies on signal recognition and transduction through intermolecular complex formation.

Processes leading to ambiguity are not limited to transcription and protein synthesis. Functional proteins are often the result of the assemblage of a number of polypeptide chains allowing more than one combination, and therefore different "meanings" (functions). Moreover, each protein has its own conformational landscape, resulting in different minima and functions. One typical example of this is the "mad cow" illness (BSE), where two different conformations of exactly the same protein make the difference between life and death for the cow itself and the humans eating it. Calmodulin is another often quoted example as it can form at least 30 different complexes with other proteins leading to different conformations and functions. Finally, an already assembled protein may complex with other compounds such as for instance sugars and lipids and again hence perform different functions. Again, inter-molecular communication and complex formation increase the "benevolent disorder" at the somatic level and at the same time limit the range of combinations possible between different molecules, allowing the choice of the proper ones for each specific temporal or spatial context. Communication therefore is a key process for

all living systems at all levels of complexity, from the molecular one to cells, organisms, populations, species, ecosystems and the biosphere, the robustness of each level of organization being correlated with the variability potential of the lower one and at the same time with the efficiency of communication between components of the modular networks previously discussed allowing the exploitation of vicariance and redundancy capacity of the systems.

How close is the link between variability generators and the choice of the necessary part of variation for adaptation on one hand and constraints to randomness in the DNA code on the other?

Biologically functional constraints leading to reduction in entropy measures and, particularly, to short and long range correlations in DNA sequences are manifold and therefore we shall just mention a few examples.

Causes of constraint fixation can be structural and functional. To the first class belong periodicities deriving from DNA strings compression through DNA/protein complex formation into nucleosome and different levels of winding of the double helix as discussed for instance by Arneodo *et al.* [2001]. At the functional level it is known that four dimensional local DNA conformational landscapes are critical for signal recognition and modulation of gene expression. Particularly, specific "motifs" are located in all eukaryotic genes upstream, in the non coding internal part (introns) and downstream; in mammals GC rich regions are near to transcription initiation sites and in all cases strings located at different distances from the mRNA production starting point, must be endowed with specific patterns of twisting, bending and rolling to obtain optimal levels of expression according to the contexts, in coherence with the work by Audit *et al.* [2001] who showed a link between the distribution of DNA bending profiles and long-range correlations in eukaryotic DNA sequences. Finally, levels of sequence stability and conservation play a relevant role in modulating both plasticity and homeostasis in different regions of the genome leading to the formation of AT and GC patches on the strings. It is plausible to imagine therefore that such constraints may extend for long tracts in the sequence, and give a statistical signature in terms of correlations of that sequence. However, the number of processes leading to constraints to randomness is enor-

mous and many of them are still to be discovered. Future studies building a large-scale framework that includes a large number of these processes into a biologically plausible physico/chemical model of DNA may improve our insight on the complexity of the DNA system.

4. CONCLUSIONS

The data just reported seem to us to solve the antinomy between chance and necessity showing that living systems are always at an intermediate state between total randomness of the distribution of the components and total order. All living beings need therefore the "benevolent disorder" (genetic variability and phenotypic plasticity) to actively react to internal and external noise during single life cycles and throughout evolution. Both evolvability and plasticity are critical for survival, the first being more relevant for prokaryote adaptation, while the second being the key for eukaryote lives. During evolution, refined and different tools have been developed in the two domains through selection of both evolvability and plasticity.

Genes and alleles increasing phenotypic plasticity and evolvability then, at the population level, are a part of the gene pool or, to say it with the words of Sewall Wright, of the genetic landscape (the genetic space) in which populations and species are situated. The larger the genetic space of a population the higher its evolvability, the larger the phenotypic space of individuals in that population the higher their individual fitness, leading as mentioned before, to the increase in plasticity related coding and non coding sequences during evolution.

The second point worth being mentioned here is the fact that the network structure of living systems and the consequent "internal selection" pressure may involve acceleration of the evolutionary dynamics when a non-lethal but major change occurs in a network. This is the case of the "hopeful monsters" proposed by R. Goldschmidt, that is organisms where such a mutation has occurred and selection will be acting (this at least is the "hope" of the "monsters") to adapt the rest of the genome to the effect of the

mutation by selecting a new, more harmonic combination of genes and alleles.

It should be stressed here that the proposed new vision of evolutionary processes does not by any means disprove neo-Darwinism and its all currents of thought because certainly it is true that allele frequencies vary due to selection, mutation and random drift. However, the new data show that the Modern Synthesis does not take into account several key factors and processes not sufficiently considered so far, suggesting the need to return to the ideas of Charles Darwin himself although with a much deeper knowledge of the material basis of the different processes.

Finally, we think that what we have discussed is only part of the present day revolution and that in the future it will be necessary to take into account the "four dimensions" of evolution introduced by Jablonka and Lamb [2004], namely the genetic, epigenetic, behavioral, symbolic.

[1] *Dept. of Evolutionary Biology, Firenze University, Firenze (Italy)*
[2] *Center for Mind Brain Sciences, Trento University, Trento (Italy)*

REFERENCES

Atlan, H. [1979], *Le crystal et la fumée,* Seuil, Paris.

Audit, B., C. Thermes, C. Vaillant, Y. d'Aubenton-Carafa, J.-F. Muzy and A. Arneodo [2001], Long-range Correlations in Genomic DNA: A Signature of the Nucleosomal structure. *Phys. Rev. Lett.* **86**: 2471-2474.

Barabasi, A.L. [2002], *Linked, the New Science of Networks.* MIT Press, Cambridge (Mass.).

Bateson, G. [1980], *Mind and Nature.* McMillan Publ., New York.

Bernardi, G., B. Olofsson and J. Filipsky [1985], The Mosaic Genome of Warm Blooded Vertebrates. *Science* **228**: 953-958.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang [2006], Complex Networks: Structure and Dynamics. *Physics Reports* **424**: 175-308.

Brooks, D.R. and E.O. Wiley [1988], *Evolution as Entropy.* Chicago U.P., Chicago.

Brunner, S., K. Fengler, M. Morgante, S. Tingey, A. Rafaski [2005], Evolution of DNA Sequence Nonhomologies among Maize Imbreds. *The Plant Cell* **17**: 343-360.

Buiatti, M. [1998], L'analogia informatica del dogma centale e le conoscenze attuali della biologia. In *L'informazione e le scienze della vita.* Angeli, Milano.

Buiatti, Marcello, Marco Buiatti [2001], The Living State of Matter. *Riv. Biol./Biol. Forum* **94**: 59-83.

Buiatti, M. [2004], *Il benevolo disordine della vita.* UTET, Torino.

Buiatti, M., P. Bogani, C. Acquisti, G. Mersi, L. Fronzoni [2004], The Living State of Matter: between Noise and Homeorrhetic Constraints. In *Nonextensive Entropy - Interdisciplinary Applications.* M. Gell-Mann and C. Tsallis (eds.), Oxford U.P., Oxford.

Buiatti, Marco, Marcello Buiatti [2004], Towards a Statistical Characterisation of the Living State of Matter. *Chaos, solitons and fractals* **20**: 55-61.

Crick, F. [1958], Central Dogma of Molecular Biology. *Nature* **227**: 561-563.

De Vries, H. [1901-1903], *Mutations Theorie.* Veit, Leipzig.

Earl, D.J. and M.W. Deem [2004], Evolution is a Selectable Trait. *Proc. Natl. Acad. Sci.* **101**: 11531-11536.

Gagliasso, E. [2001], *Verso un'epistemologia del mondo vivente.* Guerini, Roma.

Gerhart, M. and J. Kirschner [1997], *Cells, Embryos and Evolution.* Blackwell Science, Malden (Mass.).

Goldschmidt, R.A. [1940], *The Material Basis of Evolution.* Yale U.P., Yale.

Goodwin, B. [1995], *How the Leopard Changed Its Spots.* Weidenfeld and Nicholson, London.

Huxley, J. [1942], *Evolution, the Modern Synthesis.* Harper, New York.

Jablonka, E. and M. Lamb [2004], *Evolution in Four Dimensions.* MIT Press, Cambridge (Mass.).

Jacob, F. [1970], *La logique du vivant.* Gallimard, Paris.

Kauffman, S. [1995], *At Home in the Universe.* Penguin Books, London.

Kimura, M. [1961], A Mathematical Problem in Population Genetics. *Proc. Camb. Phil. Soc.* **57**: 574-582.

Lerner, I.M. [1954], *Genetic Homeostasis.* Oliver and Boyd, London.

Lewin, B. [2007], *Gene IX.* Prentice Hall, New York.

Lewontin, R.D. [1974], *The Genetic Basis of Evolutionary Change.* Columbia U.P., New York.

Li, W. [1997], The Study of Correlation Structures of DNA Sequences - A Critical Review. *Computers & Chemistry* **21**(4): 257-272.

Lombardo, M.J., I. Aponyi, S.M. Rosenberg [2004], General Stress Response Regulator RpoS in Adaptive Mutation and Amplification in *Escherichia coli*. *Genetics* **166**: 669-680.

Madlung, A. and L. Comai [2004], The Effect on Stress on Genome Regulation and Structure. *Annals of Botany* **94**: 481-495.

Maturana, H.R. and F.J. Varela [1980], *Autopoiesis and Cognition.* Reidel Publishing Company, Dordrecht.

Mayr, E. [1982], *The Growth of Biological Thought, Diversity, Evolution and Inheritance.* Harvard U.P., Cambridge (Mass.).

Mc Clintock, B. [1951], Chromosome Organisation and Gene Expression, Cold Spring Harbor. *Symp. Quant. Bio.* **16**: 13-47.

Monod, J. [1970], *Le hazard et la nécessité.* Seuil, Paris.

Needham, J. [1986] (1941), Evolution and Thermodynamics. In *Time, the Refreshing River.* Russell Press, Nottingham, pp. 207-233.

Nicolis, G.G. and I. Prigogine [1977], Self-organisation in *Non Equilibrium Systems.* John Wiley, New York.

Nilsson-Ehle, J. [1908], Einige Ergebnisse von Kreuzungen bei Hafer und Weizen.

Bot. Notiser **1**: 257-294.

Orel, V. [1983], Mendel's Achievements in the Context of the Cultural Peculiarities of Moravia. In *Gregor Mendel and the Foundations of Genetics.* Mendelianum, Brno, pp. 23-47.

Schrödinger, E. [1942], *What is Life? The Physical Aspects of the Living Cell.* Cambridge U.P., Cambridge (Mass.).

Strogatz, S.H. [2001], Exploring Complex Networks. *Nature* **410**: 268-272.

Thom, R. [1979], *Stabilité structurelle et morphogenèse.* Inter-Editions, Paris.

Trifonov, E.N. [1989], The Multiple Codes of Nucleotide Sequences. *Bull. Math. Biol.* **51**: 417-432.

Waddington, C.A. [1974], A Catastrophe Theory of Evolution. *Ann. N.Y. Acad. Sci.* **231**: 32-42.

Wicken, J. [1987], *Evolution, Thermodynamics and Information.* Oxford U.P., Oxford.

Wright, S. [1988], Surfaces of Selective Value Revisited. *Am. Naturalist* **131**: 115-123.