

Randomness in biology

THOMAS HEAMS

*INRA - UMR 1313 - Génétique Animale et Biologie Intégrative,
Domaine de Vilvert Bâtiment 211, 78352 Jouy-en-Josas cedex,
AgroParisTech - Département Sciences de la Vie et Santé,
16 rue Claude Bernard, 75231 PARIS cedex 05, France
Email: thomas.heams@agroparistech.fr*

Received 1 January 2012; revised 1 June 2012

Biology has contradictory relationships with randomness. First, it is a complex issue for an empirical science to ensure that apparently random events are truly random, this being further complicated by the loose definitions of unpredictability used in the discipline. Second, biology is made up of many different fields, which have different traditions and procedures for considering random events. Randomness is in many ways an inherent feature of evolutionary biology and genetics. Indeed, chance/Darwinian selection principles, as well as the combinatorial genetic lottery leading to gametes and fertilisation, rely, at least partially, on probabilistic laws that refer to random events. On the other hand, molecular biology has long been based on deterministic premises that have led to a focus on the precision of molecular interactions to explain phenotypes, and, consequently, to the relegation of randomness to the marginal status of ‘noise’. However, recent experimental results, as well as new theoretical frameworks, have challenged this view and may provide unifying explanations by acknowledging the intrinsic stochastic dimension of intracellular pathways as a biological parameter, rather than just as background noise. This should lead to a significant reappraisal of the status of randomness in the life sciences, and have important consequences on research strategies for theoretical and applied biology.

1. Introduction

‘The large number of cell states that are present in the lifetime of an organism and the reproducibility with which they are generated indicates the existence not just of programmes but also of mechanisms that ensure their reliable execution’ (Arias and Hayward 2006).

Reading such a statement, which accurately describes the dominant belief currently held in the life sciences, it might seem surprising for us to raise the issue of randomness in biology. Recent advances in experimental biology, such as cloning, stem cell research, cell therapy, genetic engineering and synthetic biology, appear to have in common the rewiring of pre-existing connected ‘programs’ at the cell (*genetic program*) or organism level (for example, a *developmental program*). Ambitious research programs of the ‘-omics’ era involving large scale collaborations between biologists and bioinformaticians were aimed at deciphering the complexity of the networks formed by genes (and/or gene products). To this end, ever more sophisticated tools are continuously being developed to understand a biological reality that, in turn, seems ever more complex. However, as

advanced as these tools can be, they nearly all rely on what is now a fairly old deterministic view of cellular biology, according to which cells, tissues and organs are fine-tuned systems that have evolved by means of natural selection, and are consequently adapted to their immediate environment and, to a certain extent, to its variations. Their genome, and its different features (gene content, topology, gene order, gene repetition and coding and non-coding regions), is supposed to be either the data or the software, or even part of the hardware, that allows the execution of the ‘program’, which leads to adaptations to constant variations in the environment (Atlan and Koppel 1990).

Significantly, these variations are often called ‘signals’, which cells would be able to interpret so that they can respond properly. Complex networks of gene regulations, even though they may involve hundreds of genes and proteins, might be transiently difficult to disentangle, but are ultimately expected to rely on such precise relationships between their elementary units. Developmental biologists, for example, can decipher complex networks of genetic regulations with outstanding precision, and represent them as printed circuit boards (Davidson *et al.* 2002), here again merging informatics and biology. Furthermore, genetic engineering relies on the substitution, addition or deletion of genes, as could be done with parts of such a circuit board, or with lines of code. This seems *a priori* to leave little room for randomness, but a number of important factors lead us to question this common belief.

One is that the life sciences have a long history of dealing with chance. One of Darwin’s early sources of inspiration was the English botanist Robert Brown, who, in 1827, was the first to describe the erratic movement in water of particles ejected from pollen grains of *Clarkia pulchella*, which later became known as Brownian motion. He went on to observe such movements in inorganic situations, thereby ruling out any biological forces as an explanation. Beyond this example, as we will demonstrate below, several of the most important milestones in biology reached during the last two centuries were achieved by acknowledging the importance of random dynamics. In fact, the degree to which this is the case may even mean that the deterministic status of molecular biology described above may come to be seen as an intriguing exception. Another reason for looking more closely at randomness is that in recent decades this determinism has itself been challenged to its very core: new tools, as well as new theoretical approaches, which have been substantiated by an increasing number of experimental results, have revealed intrinsic sources of unpredictability at various scales, calling into question the deterministic premises. ‘Regulation networks’, as many systems of molecular relationships within cells are called, were initially thought of as cybernetic loops, but this representation appears to face so many sources of variability that it might be poorly suited to describing reality. There seems to be an important paradox here: while evolutionary theory has successfully ruled out any hidden design or program as an explanation for the living world, molecular biology seems, despite a growing body of factual contradictions, to focus on the reliability of systems, and to do so by continuing a dependence on ‘program’ metaphors to explain how cells or organisms work.

This paradox, which will be our theme in this paper, is based on at least two important foundations. First, it must be stressed that the definitions of randomness in living systems are often equivocal. In many ways, biology is an empirical science, and in the face of

unpredictable events, biologists often lack the tools to determine whether they are truly random, or chaotic, noisy, stochastic, probabilistic or just uncertain (Merlin 2009). Such distinctions may mean a lot to mathematicians, but they are seldom made by biologists themselves (Ruvinsky 2009), though more so by philosophers of science (Werndl 2012). The issue is made all the more puzzling since, for example, it is well known that deterministic processes can appear to be random, and *vice versa* (Lenormand *et al.* 2009). This semantic discussion is of importance because it is related to the philosophical issue of the very nature of indeterminism in biology, and even in science more generally (Rosenberg 2001): when randomness is observed, we need to ask whether it is only apparent randomness resulting from our ignorance of a very complex global set of deterministic causes, or the result of a fundamental indeterminism. In a way, one of the basic processes of science is the attempt to discover order or regularities that may hide behind apparent disorder. In biology, many such causes still remain elusive and, furthermore, many sources of potential indeterminism have also been identified, so the debate is ongoing. However, the extent to which this has a practical impact on biology may be questioned. Take the example of spontaneous genetic mutations: there is no doubt that they can ultimately be explained by physical laws, but this does not change the fact that they seem to appear randomly at the DNA scale (whether they are subsequently corrected or not is an independent issue here). So biologists find it more useful to take this randomness for granted, and describe the occurrence of mutations using probabilistic laws, rather than waiting for some elusive Laplacian supercomputer that could integrate all subatomic particles interactions to predict where and when a mutation will appear. In doing this, they acknowledge a practical form of randomness, which does not rule out a possible hidden determinism.

The second foundation for the paradox is the fact that biology is both multistep and multiscale. 'Multistep' here refers to the concatenation of events that take place during biochemical processes. For example, gene expression is a complex process that includes activation of DNA promoters, transcription of DNA in RNA, translation of RNA into proteins and downstream post-translational modifications. This unidirectional flow of information is known as the 'central dogma of molecular biology' (Crick 1958). However, each of these steps displays different efficiencies, along with associated 'error' rates. Moreover, to understand gene expression, we should also mention the various possible outcomes at each step (for example, RNA splicing, editing. . .), and the feedback (positive or negative) each step can impose on the previous ones. Obviously, this leaves room for random variations. The 'multiscale', or multilevel aspect (Buiatti and Buiatti 2008; Buiatti and Longo 2011) arises from the fact that biology spans a size range from biomolecules through organisms to populations. Each unit at a given scale is the result of units at a smaller scale, giving a structure like a Russian doll. 'Multiscale' also refers to different timescales: some biological phenomena need to be studied at the microsecond scale, whereas others take place over millions of years. These two orthogonal features imply that living beings embed unpredictability at all levels, and are thus resistant to any attempts to describe them comprehensively, let alone to monitor them. On the other hand, we shall see later that these multilayer structures are also one of the reasons that a biological order is possible, through Darwinian variation/selection processes. Depending on what aspect we are interested in, the resulting description

of such systems can emphasise either randomness or order, and consequently fuel the paradox.

Rather than trying to provide a solution to this paradox, this paper will present a review of the current status of ‘randomness’ *largo sensu* in different fields of biology. We will begin, in the next section, by taking evolutionary biology and genetics as examples of disciplines where there has been some constitutive and long lasting, though sometimes inconsistent, attention paid to this notion. Then, in later sections we will consider the equivocal relationship between molecular biology and chance: though, at first neglected, if not completely disregarded, random dynamics has progressively earned its place in the field, even if it has been interpreted in various ways.

2. Randomness as a constitutive dimension of biology: evolutionary biology and genetics

It was a natural tendency in the prescientific era to look for hidden causes and elusive organising forces to explain the structure and shape of living beings, their resemblance within groups, their adaptation to the environment and the apparent perfection of their development. Because this included humans beings, the general framework for these explanations was generally theistic. In Europe, the Enlightenment gradually created the conditions for a switch to new categories of explanation. Linneus published his *Systema Natura* in 1735, and this provided a methodology for the general classification of species. Although he was a religious man, and a devoted fixist for most of his life, his intellectual legacy led to the discovery of evidence for the relatedness of species, and was therefore a first significant step towards modern conceptions of evolution. Jean-Baptiste Lamarck, one of the very first scientists to acknowledge the evolution of living beings in his *Philosophie Zoologique* of 1809, postulated a kind of vital force to explain it, leaving little if any space for chance in the process.

One of the main contributions by Charles Darwin, who, along with Alfred Russel Wallace, triggered biology’s entry into modernity, was to take into account the natural variability that occurs in every natural population (Darwin 1859; Darwin and Wallace 1858). He posited that these variations were the default state of every natural population and regarded them as the necessary playing field on which natural selection (later complemented by other evolutionary forces, such as migration or drift) can act. This opened the door to integrating chance into biology, though variability is not reducible to randomness. Darwin himself, in the opening of chapter five of *On the Origins of Species*, had a balanced view on this issue. He stated that variations ‘are due to chance’, but at the same time he describes ‘chance’ as a ‘wholly incorrect expression’ as it just describes ‘the ignorance of the cause of each particular variation’. Nonetheless, and more importantly, Darwin stated the fundamental independence of variation and selection, making the theory incompatible with any kind of predetermined design. In doing this, he undoubtedly gave credit to chance as a generator of diversity. Random events were for the first time seen not as challenging biological order, but as useful, and forming parts of the process that leads to such order. Thus the Darwinian theory, at least in its initial form, dealt with an unstable equilibrium between variation, if not chance, and selection.

This set the stage for debates on the relative importance of variation and selection, and, indeed, it seemed at the turn of the twentieth century that there was a major conflict between randomness and Darwinism. After the rediscovery of the Mendelian laws of heredity (which had been established in 1865 by Gregor Mendel, but remained generally unknown until 1900) and the subsequent birth of genetics, the ‘mutationist’ (or ‘geneticist’) school, led by De Vries and Bateson, put the emphasis on chance, leading to mutations, as the main evolutionary force, and consequently overlooked selection. The mechanisms of mutation were starting to be understood, which then led to a rather discrete conception of evolutionary processes. In doing this, they challenged the ‘gradualists’ or ‘biometricians’, who were then seen as the true heirs of Darwin, since they insisted on gradual change, thereby understating the actual role of each random mutation. This controversy only came to an end in the 1930s, when population and quantitative genetics reconciled the two approaches by demonstrating that complex phenotypes often resulted from many genes, each with a very small influence: gradualism, when observed, was thus actually compatible with elementary mutations, thereby solving the contradiction. Moreover, this explanation was compatible with spectacular, but rare, single mutations that may sometimes bring novelty in discrete ways.

This historical episode illustrates the fact that debates in evolutionary biology often deal with the importance of random events with different origins. More recently, in 1972, Niles Eldredge and Steven Jay Gould challenged the gradualism of the evolutionary synthesis, and defended, *inter alia*, the view that contingency, which we shall see here as an external form of randomness based on apparently unpredictable historical events, had a far more important influence on the evolution of life forms than was admitted at that time (Eldredge and Gould 1972). Robust debates followed, and intensified when Gould later teamed up with Richard Lewontin to dispute the central role of adaptation in the variation/selection process (Gould and Lewontin 1979), thereby diminishing the role of genetic programs as the sources (and memory) of the evolutionary successes of a lineage. Here again, after strong opposition, these critics have been absorbed into the evolutionary synthesis, and are no longer seen as challenging it: when studying a given evolutionary pathway, the question is no longer if, but how much, contingency has influenced it, with a large spectrum of answers being theoretically possible with respect to the large spectrum of possible local situations.

Beyond these examples, it can be said that the various contributions of randomness, even if far from completely understood, are now carefully taken into account in evolutionary biology. Lenormand *et al.* (2009) provides an extensive review of this issue and distinguishes three categories of stochasticity: at the gene, individual life histories and environment levels. In particular:

- The first category refers to mutation, where stochasticity affects the occurrence of mutation itself – see also Ruvinsky (2009). DNA mutations are diverse, from elementary nucleotide substitutions, through insertions or deletions of nucleotides or sequences, to more global ones such as chromosomal translocations. If mutations occur in coding regions, they may have an affect on the resulting protein structure, and consequently change the phenotype.

Each of the categories of mutation has its own probability of occurrence (roughly speaking, the simpler ones are the most frequent), and repair mechanisms exist for most of them.

As we have already mentioned, the random dimension of these mutations is mainly empirical, and obviously does not exclude underlying physical causes: for example, nucleotide substitution may be a result of tautomeric shifts in nucleotides due to thermodynamic or quantum effects. Such a shift can lead to a mispairing of nucleotides, and if not corrected by the enzymatic repair system, will finally result in a heritable mutation. Again, because there are no tools to give accurate predictions of such a chain of events, it is better described as a stochastic event. When occurring in coding regions, most mutations are deleterious, and, consequently, from an evolutionary point of view, not only the occurrence, but also the order in which rare beneficial mutations occur in a chronological sequence, has to be accounted for to give an accurate description of the influence of this level of stochasticity.

It should be mentioned here that the independence of mutation and selection has been questioned recently, at least in the case of prokaryotes (Bjedov *et al.* 2003; Brisson 2003; Martincorena *et al.* 2012). Various situations in which natural stress-induced directed mutations occur have been presented, even if their extent remains controversial (Badyaev 2005). Even if the appearance of such stress-induced mutation systems can be explained as the product of Darwinian selection, it might reintroduce a certain form of determinism in the way organisms face episodic stress, thus preventing modelling by strict Darwinian dynamics. Several cases of stress-induced transposons (mobile genetic elements) have also been described in eukaryotes (Madlung and Comai 2004).

- The second category refers to the chain of unpredictable events in individuals of a given population that lead to changes in its genetic structure from one generation to the next. For multicellular organisms that use sexual reproduction, several random processes occur. During gametogenesis, chromosomes from the maternal and paternal origins of a given pair are separated during the cell division called meiosis so that each gamete randomly inherits one of them, each pair being independent in this process. Furthermore, a given pair of chromosomes may exchange homologous portions of their genetic material before separation, and the distribution of the breaking points follows a probabilistic law (the longer the space between two genes on a given chromosome, the more likely it is that a so-called ‘crossing-over’ will occur between them). These sources of variation among and within chromosomes are called recombination, and result in a huge range of combinations of genes in gametes. Then, the fertilisation step introduces additional levels of probabilistic events: the choice of mate (if any), and, ultimately, the specific pair of gametes involved in the fertilisation. As a consequence, the frequency of alleles in an offspring generation depends on all these stochastic events, as well as on the number of offspring per individual, which is also a random variable. The resulting frequency of such a variable in the new generation is, therefore, also a random variable. This source of stochasticity is called genetic drift, and may have major evolutionary influences: in particular, in small populations.
- The third category of stochasticity is concerned with the environment, whether it is biotic or abiotic. Environmental variations can be regular (for example, seasonal) but

may also have a stochastic component (for example, volcanic eruptions), which has to be taken into account when modelling the evolution of populations. This is connected with contingency issues, as mentioned earlier.

Finally, Lenormand *et al.* (2009) explored the way these sources of stochasticity influence evolution. The typology produced acknowledges four potential influences:

- (i) stochasticity often leads to deleterious mutation and maladaptation;
- (ii) when a neutral mutation occurs, its stochastic fate can lead to evolutionary freedom and increased diversity;
- (iii) more controversially, stochasticity can allow a population to jump from one adaptive peak to another; and, finally,
- (iv) stochasticity may itself generate new selection pressures when it affects an organism's robustness or ability to evolve.

3. The ordered world of molecular biology: how to deal with randomness

The long association between randomness and biology seemed to reach a paradoxical conclusion at the dawn of molecular biology – the sub-field of biology dedicated to unravelling the molecular basis of biological activity (Morange 1994). Molecular biology has had a prominent role in the life sciences since its very beginning during the 1950s and 1960s, when at least three universal features of the living world became established:

- The first was the structure of the DNA molecule, which was published in 1953 by James Watson and Francis Crick (Watson and Crick 1953) with Rosalind Franklin's decisive input (Klug 1974). The linear 'double helix' structure displayed molecular features that were used to explain what were at the time a number of elusive characteristics of genetics: the storage of information, its mutability and its equal transmission across somatic cell divisions. In this way, Watson and Crick (1953) provided an outstanding demonstration of the material basis of heredity, linked with evolutionary processes.
- The second was that the above-mentioned 'central dogma of molecular biology' was able to explain the concrete relationship between genetic information and proteins (Crick 1958). In addition, this unidirectional flow of information from genes to proteins provided confirmation at the molecular scale of the impossibility of the inheritance of acquired characters, strengthening Darwinian over Lamarckian dynamics as valid explanations of evolutionary processes.
- The third was that the genetic code, the 'rosetta stone' linking nucleotide sequences of DNA to amino-acid chains of proteins, was deciphered at the same time by teams led by Nirenberg, Khorana and Holley, and it turned out that, despite some partial exceptions, it was a simple and shared feature of all living beings.

These three pillars of molecular biology contributed to the emergence of a unified view of the living world. Furthermore, molecular biology soon evolved from an analytical discipline to an applied one: the discovery of restriction enzymes, which can copy and paste DNA, and further techniques for DNA synthesis, sequencing and amplification,

progressively allowed advantage to be taken of this general feature, and entry to an era of modification of organisms.

In this framework, chance, or even probabilistic events, was no longer considered a fruitful source of explanations. Molecular biology also emerged at the dawn of computer science, and adopted much of its lexicon (Longo and Tendero 2007). As mentioned earlier, terms such as ‘information’, ‘code’ and ‘regulation’ contributed not only to the description of biological phenomena, but more importantly to shaping the discipline itself. Major examples of this influence include the fact that enzymatic complexes processing DNA and RNA were soon compared to Turing machines, and Norbert Wiener’s cybernetic laws were applied at an early stage to describing retrocontrol loops in gene networks. Biology entered an era for which precision was the rule, and randomness became relegated to a distant component of the evolutionary process, far from the biomolecular scale. Strikingly, two independent articles from 1961 were the first to evoke the concept of a genetic ‘program’. One, Mayr (1961), addressed evolutionary issues, and linked the issue of ‘cause and effect in biology’ with the existence of such a program, while the other was the seminal paper Jacob and Monod (1961) by François Jacob and Jacques Monod, which laid the foundations of genetic regulation studies, making it one of the most important milestones of molecular biology. In the conclusion of Jacob and Monod (1961), the authors state that ‘the genome contains not only a series of blue-prints, but a co-ordinated program of protein synthesis and the means of controlling its execution’. It is easy to see similarities between this statement and the one we quoted at the start of the present paper, illustrating the robustness of this view during the last fifty years: as François Jacob wrote enthusiastically some years later, ‘in short, everything urges one to compare the logics of heredity to that of a computer. Rarely has a model suggested by a particular epoch proved to be more faithful’ (Jacob 1970).

It is important to stress the influence of the ‘epoch’, because it may be imagined that this deterministic road was not the only one that molecular biologists could have taken in its early days. Perhaps the best example of a missed opportunity to infuse molecular biology with probabilistic laws can be found in Erwin Schrödinger’s book *What is life?*, which was first published in 1944 and is known to have inspired the first molecular biologists, first and foremost, Watson and Crick. In this visionary essay, the quantum physicist foresaw that an aperiodic crystal located in chromosomes would be an adequate structure for storing genetic information, a prediction that turned out to be true when Watson and Crick’s paper was published nine years later. But more importantly, Schrödinger investigated whether or not the regular features of living entities, such as cells, could be explained by the statistical reliability of the law of big numbers. In counting atoms present in the chromosome volume, he stated that they do not fulfill such conditions, so probabilistic dynamics cannot explain *per se* biological order, and he concluded by advocating a search for the additional features that would explain this new specific order. It is puzzling here that the emerging discipline resisted one of the founding fathers of quantum physics in his attempt to reinforce its link with randomness (Schrödinger 1944).

Instead, molecular biologists took the opposite, deterministic path, which significantly affected their research goals. In this new logic, they focused on describing phenotypes by means of cascades of self-regulated molecular reactions, triggered by precise gene

expression. Evolutionary processes were then understood, in this framework, as acting on optimisations of such networks, resulting in them performing their tasks accurately, as well as buffering variations in external conditions. It was expected, at least in theory, that once each such reaction was decrypted, a step-by-step integration would be achievable, so that a global synthesis would eventually provide a general view of organisms as complex networks of molecular reactions.

For decades, molecular biology relied on such assumptions, which in turn modified the vision of the default state of living structures. Genes and genetic programs were assigned a role so central that, even today, it is a common view that cells with the same genome (for example, clonal cells), living in a homogeneous environment, will, all things being equal, behave in the same way. And if they do not, additional causes are often looked for, such as epigenetic or cryptic environmental variations: both of these have in common the fact that they are external and supposed to disrupt the basic genetic order. But the founding fathers of molecular biology had no technical ways to prove, let alone any actual reasons to hypothesise, this postulate of homogeneity. The available techniques only allowed them to harvest RNA or proteins from large populations of cells. In using rules of three to infer what was going on in each cell, they explicitly ruled out any variability, not to mention randomness, to explain intra- and inter-cellular dynamics.

To put it bluntly, even though it is constitutive of genetics, randomness has long been rejected in gene expression studies resulting from molecular biology. Again, things might have been different if the ‘epoch’, and the success of computer science, had not been so influential. Ironically, Alan Turing was using a probabilistic theory of information where signals are intrinsically linked with noise and perturbation, the cybernetician Norbert Wiener was a pioneer of stochastic calculus, and the very first molecular biologists, an informal group around Max Delbrück, named themselves the ‘Phage group’ after the bacteriophage virus they were studying, which was known for its ability to switch between two states (lytic and lysogenic) in a probabilistic way. However, none of these probabilistic or stochastic ideas diverted molecular biology from the direction it was set on.

The result was a sort of schizophrenic biology: acknowledging random events when generating combinatorial diversity subjected to selective pressures, but roughly excluding them from the molecular explanation level. As far as simple cascades of gene regulation were concerned, this framework proved its efficiency. Indeed, single mutations on genes or gene promoters may trigger modifications in pathways and phenotypes that can be modelled, and then confirmed *in vivo*. These principles are the basis of genetic engineering, as well as gene-therapy. For decades, reductionist assumptions that living systems would be understood by:

- (i) describing elementary parts, and then
- (ii) integrating them in a coherent way,

remained the organising principle of the scientific community. Research groups focused on small networks and gene-by-gene studies, but the more complex the organism and/or pathway considered, the harder it became to apply such basic principles to understanding it. From an epistemological point of view, molecular biology was running the risk of no longer being refutable: when a pathway under consideration failed to explain the

expected phenotype corresponding to it, it was always possible to postulate that a yet to be discovered gene or cofactor, or an upstream indirect regulation step, would account for the difference. In other words, there is the potential that theories relying on programs are impossible to prove wrong, hence undermining them as scientific explanations, even if some of them may accurately describe reality.

Alternative theories then started to surface, such as various versions of the so-called 'self-organisation' theory. For example, a theory of self-organisation by noise has been developed since 1972 by Henri Atlan. This states that apparent sources of background noise, such as redundancy in genetic sequences, actually increase the quantity of information in a system. This increase in complexity would, in turn, induce it to self-organise. But even if they open new doors, such fruitful propositions do not in fact contradict the intrinsic deterministic nature of the system since they continue to rely on the concept of information, and as discussed elsewhere, they ultimately stick to the notion of noise as an external perturbation of the system (Kupiec 2008).

4. Randomness strikes back

Despite what we have said above, some probabilistic theories began to be put forward, and it is noteworthy that not only did they refer to chance, but they did so in a Darwinian way. In immunology, the reference theory of clonal selection (Jerne 1955; Burnet 1957; Tonegawa 1976) states that when an antigen is detected by the immune system, it identifies a pre-existing small clone of immune cells displaying the adapted antibody from amongst the many others that do not fulfill this condition, and then triggers its proliferation. Importantly, this theory relies on an implicit upstream step in which many different types of cells are generated by massive random genetic recombinations so that small clones co-exist by default, with each displaying a different antibody type. This theory displaced so-called 'instructive' theories, according to which the antibody was moulded upon contact with the antigen, an idea that challenged the central dogma of molecular biology and its unidirectional flow of information from genes to proteins. By analogy with Darwinian dynamics, the random generation of diversity allows further selection, and there is no need to refer to any accurate processing of information.

The same 'selectionist' logic prevailed for another major theory related to the development of neural networks, *viz.* the 'selective stabilisation of synapses' (Changeux *et al.* 1973; Changeux and Danchin 1976), which was later developed by Edelman (Edelman 1987; Edelman and Mountcastle 1978), who even explicitly used the term 'neural Darwinism'. This theory states that at some critical points of brain development, a greater number of synaptic contacts are formed than is needed, and some of these are later stabilised on the basis of network activity, while others are destroyed. Here again, the development of such a complex organ as the brain was found to be better explained by random exploratory cell behaviours followed by subsequent selection than by any sophisticated program. Several other examples of similar mechanisms embedding a 'cellular selection' component were progressively documented as a constitutive dimension of multicellularity (see Michaelson (1993) for a review).

However, apart from these system-specific theories, the stochasticity of cellular events *per se* remained confined to a marginal role for a long time. So when extensive observations of this stochasticity began to be made, it was explained as unavoidable and/or residual background noise of biological interactions. During the 1990s, techniques for observing single-cell gene expression became available, and these allowed in-depth analysis for the first time. However, the existence of technical limitations up to that time is not sufficient to explain why molecular biologists in general were so reluctant to face up to randomness. For example, flow cytometry, which is a technique of cell (or chromosome) counting and subsequent multiparametric analysis, had for decades provided 2D plots featuring clouds of dots showing, albeit roughly, that even clonal cells could have different sets of coordinates, and consequently different characteristics. This could have led to a thorough investigation of the basis of such heterogeneity, and might have revealed its stochastic component. Instead, these observations were just used to determine the average behaviour of cell populations, and to label it ‘normal’, and consider any variation around it as behavioural ‘noise’ in each given cell. Thus, this technique, which had been available since the late 1950s, could have fueled probabilistic hypotheses, and the fact that it did not underlines the influence of the ‘genetic program’ metaphor, which was so strong that it even confined counter-examples, and experimental evidence of potentially indeterministic behaviours, to the margins of its own explanatory system.

However, in parallel developments, a true stochastic molecular and cellular biology progressively took shape in various model organisms – in particular:

- In 1964, the seminal paper Till *et al.* (1964) addressed the issue of hematopoietic differentiation using stochastic models.
- Spudich and Koshland (1976) reported on experiments on what they called ‘non-genetic individuality’ in homogeneously grown populations of bacteria.
- Bennett (1983) proposed an ‘on-off stochastic model’ for cell differentiation.
- Kupiec had already started to elaborate a ‘probabilist theory for cell differentiation’ (Kupiec 1981; Kupiec 1983), which was, as far as we are aware, the first theoretical attempt to take into account the stochasticity of gene expression as a central feature of cell behaviour, while escaping the instructionist paradigm.

At first sight, it may seem unlikely that these few references could seriously challenge a widely accepted framework, but two crucial supporting factors came into play:

- (1) This emerging true probabilistic framework soon turned out to be supported by a large amount of experimental data.
- (2) It had the potential to provide much simpler explanations than the pre-existing ones.

The following sections discuss these factors in more detail.

5. Experimental evidence for stochastic gene expression

In the early 1990s, new cytological imaging techniques, followed by single-cell genetic amplification methods, progressively gave access to gene expression at the individual cell level. Thus, there was a rapid expansion in single-cell studies, and these provided new insights into cellular regulation mechanisms. Ross *et al.* (1994) studied the randomness of

gene expression in eukaryotic systems, which they linked to its low frequency. Wijgerde *et al.* (1995) published important results addressing the dynamics of chromatin interactions *in vivo*, and provided evidence of forms of topological competition on DNA that were logically leading to the idea of stochastic variations from cell to cell. Stochastic cell behaviours, resulting from pulsative transcription, were also observed in mouse muscle cells in Newlands *et al.* (1998). By the end of the decade, the stochasticity of cell behaviour and gene expression had been substantiated by many experimental observations, in bacterial as well as in various eukaryotic systems, so that Thattai and van Oudenaarden (2001) could say that cells were ‘intrinsically noisy biochemical reactors’. The groundbreaking paper Elowitz *et al.* (2002) went a step further in demonstrating and even monitoring the stochasticity of gene expression in a bacterial context. Not only did they confirm it using fluorescent reporter genes, but this demonstration also came with two important observations:

- (1) Intercellular stochastic variations were proved to be more important when the expression of the reporter genes was low (that is, highly repressed).
- (2) They confirmed *in vivo* theoretical models of two-component stochasticity, namely extrinsic and intrinsic ‘noise’ (Elowitz *et al.* 2002).

Here, noise is defined as the standard deviation of gene expression divided by its mean. Extrinsic noise refers to fluctuation of amounts and locations of the molecules involved in gene expression, while intrinsic noise is due to the random microscopic events that lead to unpredictable outcomes and orderings of biochemical reactions, which would exist even without extrinsic noise. Intrinsic noise explains differences in gene expression between two reporter genes in a given cell, whereas extrinsic noise accounts for the variation of the expression of a given reporter gene between two cells. These two components may display different autocorrelation times.

Since the publication of Elowitz *et al.* (2002), the existence of this stochasticity has been uncontroversial, and is now considered to be a general feature of simple as well as complex organisms (Ferguson *et al.* 2012; Nijhout 2006; Raj and van Oudenaarden 2008). It has been described extensively in various contexts and species, in either normal (Raser and O’Shea 2005) or pathological processes (Rojo *et al.* 2011), and its heritability has also been addressed (Ansel *et al.* 2008). It has been proved to be transmissible in a cascade of gene expression (Hooshangi and Weiss 2006), and even, to a certain extent, from one cell generation to another (Kaufmann *et al.* 2007).

6. Why is gene expression stochastic?

Stochasticity in gene expression has various mechanistic causes. First, as mentioned earlier, gene expression is a multistep process (Kaern *et al.* 2005). Each step is associated with a certain amount of reliability, and the global process leads to variability in the final outcome, namely the amount of protein. Retroaction loops may buffer some of these variations as a result of natural selection, but evolutionary pressures may also maintain differential ‘efficiencies’ between steps. Typically, the more upstream a noisy step is, the more likely the final outcome is to be unpredictable.

Another source of stochasticity is the low availability of many regulatory proteins. It must be remembered that a large majority of a cell's proteins are present in very low copy numbers. It is estimated that 80% of the proteins have fewer than one hundred copies per cell, and among these, there are many regulatory proteins (Guptasarma 1995). This means that the law of big numbers is often irrelevant, so DNA sites can 'compete' for interactions with given proteins, and, inevitably, differences are expected to occur between cells, even clonal ones.

Further causes of potential increases in the unpredictability of gene expression at the cell level are:

- (i) Macromolecules are densely crowded, which may affect their properties, and implies that the physical interactions between molecules are not straightforward, and may come with delays from one cell to another (Ellis 2001). Here, genomes fundamentally differ from computer software or hardware, where the time between two instructions is independent of any spatial considerations.
- (ii) Chromosome territories, that is, the three-dimensional structure of chromosomes within the nucleus at gene expression steps of the eukaryotic cell cycle, affect gene expression, and this topological configuration changes probabilistically from cell to cell (Parada *et al.* 2003; Cremer *et al.* 2006). It is noteworthy that global features of this topology have been deciphered and have shown some heritability across cell generations.
- (iii) Even though it is still controversial, growing evidence supports the idea that 'pervasive transcription' (gene expression in 'non-coding' genome regions) may actually exist and is widespread (Clark *et al.* 2011). This would result in a huge increase in the cellular gene expression network's complexity, and in potentially massive perturbations of the conventional regulatory mechanisms, which even brings into question their relevance.

All of these may be classified as epigenetic, since they can introduce new sources of variability without invoking genetic variation. Epigenetics has become an intense and promising field of research in recent years: for example, studies of gene methylation and genome imprinting have revealed new levels of complexity, and provided new keys to understanding inheritance mechanisms, as well as cellular physiology. Therefore, it is reasonable to think that epigenetics could hold a clue for explaining stochastic dynamics, but although it is beyond doubt that even if it partially accounts for it, it does not follow that epigenetics challenges *per se* the foundations of biological determinism. Indeed, epigenetic studies often develop so many *ad hoc* hypotheses, not to mention 'epigenetic codes' that basically follow the instructive paradigm, that their reductionist aggregation, rather than a true synthesis, may lead molecular biology even further along in its alarming tendency to a patchwork incompleteness.

Finally, we can consider a provisional set of features at the protein level leading to the intrinsic unpredictability of gene expression. The proteins in a cell are both the outcome of individual gene expression pathways and intermediate factors along such pathways, whether it is by internal retrocontrol or external influence, leading to the notion of a gene expression network. Proteins are the building blocks of cells and have various functions including catalysis, transport, reception, structure and many more. The precision with

which these functions are carried out is due to the adaptation of the protein spatial structure, called stereospecificity, which is assumed to be coded in the genome. As a result, proteins are expected to interact with spatially compatible structures. But recent developments in biochemistry show that not all proteins, if any, are structured in this way. For example, some may switch conformations with pH variation, while others may even be intrinsically disordered (Uversky 2011). It remains to be demonstrated in general whether such individual variability can have a biological function, but if evolutionary pressures can select for precise protein actions when needed, there is *a priori* no reason to exclude the possibility that they can also select loosely structured proteins, provided they are functional. In addition, it has been stressed that the theoretical connectivity of protein networks is very important: even if a stereospecific relation can theoretically exist between two proteins, the usefulness of this concept is seriously impaired by the huge number of different connections a small network of proteins can establish (Noble 2006). Combined with the small number of many protein types, and cellular crowding, this feature leads, by definition, to intrinsic variations due to local sampling effects, which cannot be accounted for by any genetic program.

7. Stochastic gene expression and reproducible pathways: from noise to function

In the face of so much evidence that cellular processes are inherently stochastic, various interpretations are possible. Unsurprisingly, many, if not most, of the biologists who study it regard it as unavoidable background noise around the correct expected value for gene expression. In doing this, they simply acknowledge that cells are not ideal computers, and that they may tolerate slight variations in their responses to environmental changes. This view does not challenge the traditional connection between evolutionary and molecular biology, where selective pressures are supposed to shape increasingly precise programs to adapt to such variations (Lehner 2008). Following this logic, cells are expected to display frequent ‘appropriate’ responses, and occasional ‘deviant’ ones, as revealed by such expressions as ‘illegitimate transcription’ (Chelly *et al.* 1989) or ‘transcriptional infidelity’ (Gordon *et al.* 2009). Some authors go even further in this direction by assuming that, counterintuitively, stochastic gene expression can even be controlled to display synchronised reactions to fluctuating environments (Zhou *et al.* 2005; Springer and Paulsson 2006) in a way that is reminiscent of self-organisation by noise.

An intermediate position is to recognise that the unpredictability of gene expression might be functional, but still the result of deterministic genetic systems. Proponents of such explanations deny, or at least play down, the status of random variation as noise, but rather analyse it as a useful generator of cellular phenotypic diversity (Chang *et al.* 2008; Fiering *et al.* 2000; Mettetal and van Oudenaarden 2007; Pipel 2011; Stockholm *et al.* 2007) that is compatible with regulation (Ferguson *et al.* 2012). Indeed, simple deterministic systems may generate unpredictable outcomes at the individual level, but reproducible patterns at the population level. Bistable equilibria are a good example of such systems at both the theoretical and biological levels (McAdams and Arkin 1997; McAdams and Arkin 1999). In such networks, a protein A can enhance the expression of two downstream genes b

and c, producing the proteins B and C, respectively. In addition, each of these proteins inhibits the expression of the other (B inhibits the expression of c, and conversely). The outcome of such a simple system is obvious: it leads either to the expression of protein B or to the expression of protein C in each unit, that is, cell, with a resulting generation of diversity. Assuming, for example, that the affinity of A is equal for b and c promoters, each cell will have a 50% chance of expressing only B and a 50% chance of expressing only C. Consequently, in the cell population reacting to A, half of the cells will produce C and the other half B. Of course, this diversity generation process can become more complex, with more than two target genes, or with differential affinities leading to different proportions.

This theoretical network can actually be observed in living systems such as in the compound eye of the drosophila fly, where colour vision relies on a comparison between two colour-sensitive photoreceptors cells, R7 and R8, present in each of the 800 ommatidia, which are the elementary 8-cell units of the eye. These photoreceptor cells discriminate either short or long wavelength, depending on what type of rhodopsine protein they express. It has been shown that in wild-type individuals, 30% of the ommatidia express a rhodopsine combination that specialises them in the detection of short wavelengths, with the remaining 70% expressing another combination that makes them more responsive to long wavelengths. Interestingly, this proportion is not achieved through any instructional programming of each cell, but rather by a stochastic commitment resulting from a bistable loop with an upstream protein having the corresponding differential affinities for the rhodopsine promoters (Mikeladze-Dvali *et al.* 2005; Wernet *et al.* 2006).

Finally, some authors have even proposed a new way of bringing stochasticity and evolution closer together through gene expression studies. (Polev 2012) suggests that organisms could take advantage of stochastic gene expression, which would allow more flexibility in the selection processes for gene sequences. Specifically, random gene expression would allow the real-time testing of the emergence of new combinations of genes and promoters (regulatory sequences of a given gene that typically facilitate gene expression), and the enhancement of the emergence and diversification of gene expression patterns and functions.

Some authors have proposed going even further in challenging the standard deterministic vision. As far as we are aware, Kupiec was, thirty years ago, the first to elaborate a new theoretical framework based on the assumption that the unpredictable stochastic behaviour of cells is not noise but a biological parameter that opens the way to a chance/selection dynamics based on the variability of gene expression (Kupiec 1981; Kupiec 1983). In this framework, the default state of the cell is variation, which is later stabilised by downstream interactions. This challenges the idea that cells are stable entities needing a biochemical 'signal' to trigger differentiation processes. In this way, this framework overcomes the epistemological tautology arising from explaining how differences between cells appear by assuming pre-existing differences, namely, cells emitting different kinds of signal, or pre-existing gradients that presuppose upstream differences (Kupiec 1986; Kupiec 1996; Kupiec 1997). In addition, it also challenges the 'program' metaphor, since from this viewpoint, reproducible patterns of gene expression are not achieved through precise regulations at the single-cell level resulting from an integrated

information flux, but rather by stabilisation of cellular states on the basis of their adaptation to their local environment.

The first versions of this theoretical proposal suggested that surface contact interactions may act as such stabilising factors, while more recent ones, substantiated by computer simulations, hypothesise that trophic relationships between cell types may contribute to the constitution of structures (for example, bilayer structures) (Laforge *et al.* 2005). This process is called ‘ontophylogenesis’: the apparent randomness of the system depends on the level of internal physical constraints, themselves resulting from evolution and phylogenesis, and this ‘internalised environment’ affects the progress of probabilistic ontogenetic events (Kupiec 2008). Thus, phylogenesis and ontogenesis are here two sides of the same coin. In this way, the framework shows that it is possible to conceive of a gene expression dynamics without referring to any genetic program. Genomes here are no longer software or data, but a toolbox of genes that cells use in probabilistic ways, taking into account the numerous sources of randomness that have been described earlier. This chance/stabilisation process recalls Darwinian dynamics, so this theoretical proposal has sometimes been termed ‘endodarwinian’ (Heams 2004). They are also partially reminiscent of Wilhelm Roux’s visionary proposal in the late nineteenth century, through which he attempted to apply Darwinian laws to ‘parts’ within an organism, albeit it was done ambiguously (Roux 1881).

Undoubtedly, this framework based on randomness deserves a thorough examination. In particular:

- In challenging the deterministic concept of a program, it fulfills the Occam’s razor principle by avoiding such a potentially unnecessary assumption. This epistemological view with regard to a core notion of molecular biology alone suggests the need for a careful and critical scrutiny of this framework.
- It is compatible with the experimental evidence for stochastic gene expression, as well as with the reproducibility of gene expression patterns at the population level. The latter is indeed an important issue: biological objects, such as individuals from the same species, do have similarities (such as morphological and developmental similarities) related to their genetic proximity, for which any general framework must give convincing explanations. At first sight, stochastic behaviours may appear to be irrelevant where similarities are concerned, but this may not be true in the biological context. Above all, a probabilistic process is obviously reproducible at the population level. If the functional efficiency of a given cell type is determined at the tissue or organ level (for example, with many secretion pathways), there is no real need for strictly homogeneous single-cell behaviour: individual variations around the average value of a cellular phenotype do not prevent these average values existing.
- These dynamics rely not only on randomness, but on chance/selection processes, in which the selection step is crucial to the shaping of the biological order observed at the macroscopic level.

An example of this is given by a recent experimental study on murine brains. This demonstrated that basic stochastic gene expression allows the constitution of different

pools of hippocampal neurons, which enhance the survival of certain fractions of them after a traumatic brain injury, which may be seen here as the selection step, and without which the rationale of the overall dynamic does not make sense (Rojo *et al.* 2011).

In situations where neither the law of large numbers nor chance/selection mechanisms can entirely account for reproducibility, such as in the first steps of embryo development, it should be remembered that links can still be drawn between determinist and probabilist dynamics. Indeed, it could be that probabilism is the global and actual framework of biological mechanisms, but this does not prevent ultra-constrained steps from taking place in an apparently deterministic fashion. The amount of stochasticity-driven variability could thus be understood as the result of evolutionary pressures: for example, the variation in gene copy number, genomic topology, the proportion of disordered proteins and similar internal constraints that can typically be monitored by natural selection, could shift the amount of intrinsic variability towards a maximum (when diversity is functionally valuable), or a minimum (when reliability is crucial, even at small cell number levels), or proportionally in any intermediate situation. Therefore, it could be that there is no true opposition between the two systems of explanation (Buiatti and Buiatti 2008; Velasco 2012), in particular, because probabilistic processes would encompass deterministic ones (Heams 2004).

Hence, it can be seen that this increasingly used framework has many convincing features, not least amongst which is the fact that it is refutable and testable by investigating if the correlations between levels of constraints and the restriction of stochasticity can be made at a general scale, though this would require some commonly accepted methods to index such values.

Another important open question is the accurate definition of what triggers the selection or stabilisation step(s). In Darwinian evolution, adaptative selection occurs at the organism level as a result of a competition for resources (food, space. . .). At the cellular level, determining whether resources are limited is not a simple issue, and is likely to depend on many local conditions, and on whether resources are defined in various ways at different scales.

Provided substantive answers can be given to these questions, this framework could have fundamental applied and even medical uses, as exemplified by recent results that take it as a starting point for cancer studies (Capp 2012).

This is certainly both a complex issue and a large research program, but increasing knowledge of stochastic gene expression may make it at least as relevant as the no less complex problem of understanding organisms by hunting down some elusive global synthetic diagram showing all their fine-tuned cellular reactions – particularly, if they do not exist.

8. Some consequences for emerging fields of biology

In the light of this discussion, molecular biology seems to be at a crossroads. However, it is striking how fruitful it still remains, in particular, because even studies of stochastic

gene expression are grounded on pre-existing clues that the subject has already produced, and is continuing to produce. But this optimistic description cannot hide the fact that potentially insuperable contradictions may grow if the relationships between molecular biology and randomness are not investigated. Among many fields where this is a critical issue is the rapidly expanding field of ‘synthetic biology’, which is currently in many ways the paradigmatic example of deterministic biology. This composite discipline has set several targets, some of which are spectacular, including:

- starting simple life forms ‘from scratch’, that is, the creation of life; and
- massively re-engineering bacteria or yeasts (and ultimately multicellular organisms) by rewiring their so-called genetic program to produce new functions, such as producing exotic compounds for various applications such as pharmaceuticals, phytochemicals or decontaminants.

The approach to both these targets can be questioned in the light of randomness.

Convincing proofs of principle have been given for the re-engineering target, and there have even been some spectacular results (Szczepara *et al.* 2003; Ro *et al.* 2006). However, many other objectives still remain unreachable, so it may be useful to explore the reasons for this. If we rely on deterministic assumptions, it may just be a question of time and complexity: the higher the expectations, the harder it will be to achieve them. But if randomness rules cellular behaviours in large part, it may be that looking at cells as small logical devices is just pointless. Interestingly, some of the scientists producing cutting-edge research on the stochasticity of gene expression are also involved in synthetic biology (Cox *et al.* 2010), and major research teams into synthetic biology now investigate the influence of noise (Murphy *et al.* 2010). Even if this is only one way to account for randomness, this may be a promising development for future integrative studies: time will tell if an enriched synthetic biology results.

Similarly, trying to understand what might be the simplest life form by determining its theoretical minimal gene set may prove to be a biased approach, which reduces a genome merely to information (Koonin 2003; Delaye Moya 2010). Indeed, with the exception of some endosymbionts (Tamames *et al.* 2007), minimal autonomous life forms have far more genes than the estimated minimal gene sets. Here again, a classical explanation might be that all phyla, including the simplest ones, have evolved toward a certain form of complexity, but it may also be that cells as we know them actually need more than minimal gene sets, including a certain amount of redundancy or degeneracy in terms of ‘information’, just so that probabilistic dynamics can take place. It would then be misleading to equate ‘simplest’ to ‘minimal’ when talking about life forms (and this observation would certainly raise new issues for how transitions from non-living to living structures could by-pass the ‘minimal’ step).

These two examples show that the debate over the status of randomness is certainly not just a definition issue, but has important and direct consequences for critical research strategies.

9. Conclusion: other examples of randomness in biology and the roads ahead

We have demonstrated in this paper that randomness and biology have fruitful but conflicting connections, which result in large part from the intrinsic partitioning of biology into many disciplines, which has certainly been legitimate in the setting up of the foundations of the science.

Other connections may also be made in addition to those already discussed. In particular, the emerging field of quantum biology is producing some intriguing results (Arndt *et al.* 2009; Ball 2011). Amongst these, recent work seems to confirm an old hypothesis by which photons randomly hitting photosynthetic pigment molecules in leaves are channelled towards the reaction centres through quantum mechanical effects (Engel *et al.* 2007). Quantum-assisted magnetic sensing has also been suggested as a possible mechanism explaining how migrating birds orient themselves (Ritz *et al.* 2004; Maeda *et al.* 2008). Even though it is not yet fully confirmed and accepted, such a mechanism may be widespread, and may be found in organisms as varied as insects and plants. There is also currently a lively ongoing debate between proponents and opponents of a vibrational theory of olfaction, which asserts that this function could be achieved through the vibrational energy of odorant molecules allowing quantum electron tunnelling leading to signal transduction (Franco *et al.* 2011; Hettinger 2011).

Obviously, these provisional and somewhat controversial results are still fragmentary. Quantum biologists themselves acknowledge that it is unclear whether such effects are limited to a few mechanisms or are widespread (Ball 2011), but they open the door to a possible new field of research, which presents new bridges between the atomic and macroscopic scales, and may be helpful in heading towards a more integrated biology. Randomness itself may play a part in achieving this goal: one of the striking characteristics of randomness in biology is that when it is associated with selection, it provides a fruitful explanatory system encompassing many levels of the living world, so extended Darwinian principles may be a way to unify biology even more than previously expected. Hence, it is likely that there will be a need for evolutionary biologists to apply their models within molecular biology and determine their ranges of validity so that biological order gets the parsimonious explanations it deserves. After all, why should Darwinian-like principles be excluded from molecular biology given that they have sometimes been used in exotic fields unrelated to the life sciences? Indeed, they have even been used in quantum physics: a 'Quantum Darwinism' theory was recently proposed in which the classical world is derived from a quantum world through the 'selection' of quantum states (Zurek 2009). Hence, randomness and biology still have a lot to exchange, and in both directions.

References

- Ansel, J. *et al.* (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genetics* 4 e1000049.
- Arias, A.M. and Hayward, P. (2006) Filtering transcriptional noise during development: concepts and mechanisms. *Nature Reviews Genetics* 7 34–44.

- Arndt, M., Juffmann T. and Vedral, V. (2009) Quantum physics meets biology. *HFSP Journal* **3** 386–400.
- Atlan, H. (1972) *L'organisation biologique et la théorie de l'information*, Seuil, Paris.
- Atlan, H. and Koppel, M. (1990) The cellular computer DNA: program or data. *Bulletin of Mathematical Biology* **52** 335–348.
- Badyaev, A. V. (2005) Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. *Proceedings of the Royal Society B: Biological Sciences* **272** 877–886.
- Ball, P. (2011) Physics of life: The dawn of quantum biology. *Nature* **474** 272–274.
- Bennett, D. C. (1983) Differentiation in mouse melanoma cells: initial reversibility and an on-off stochastic model. *Cell* **34** 445–453.
- Bjedov, I. *et al.* (2003) Stress-induced mutagenesis in bacteria. *Science* **300** 1404–1409.
- Brisson, D. (2003) The directed mutation controversy in an evolutionary context. *Critical Reviews in Microbiology* **29** 25–35.
- Buiatti, M. and Buiatti, M. (2008) Chance vs. necessity in living systems, a false antinomy. *Biology Forum* **101** 29–66.
- Buiatti, M. and Longo, G. (2011) Randomness and Multi-level Interactions in Biology. arXiv: 1104.1110v1.
- Burnet, F. M. (1957) A modification of Jerne's theory of antibody production using the concept of clonal selection. *Australian Journal of Science* **20** 67–69.
- Capp, J. P. (2012) Stochastic gene expression stabilization as a new therapeutic strategy for cancer. *Bioessays* **34** (3) 170–3.
- Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. and Huang, S. (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453** 544–547.
- Changeux, J. P. and Danchin, A. (1976) Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks. *Nature* **264** 705–712.
- Changeux, J. P., Courrege, P. and Danchin, A. (1973) A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proceedings of the National Academy of Sciences of the United States of America* **70** 2974–2978.
- Chelly, J., Concordet, J. P., Kaplan, J. C. and Kahn, A. (1989) Illegitimate transcription: transcription of any gene in any cell type. *Proceedings of the National Academy of Sciences of the United States of America* **86** 2617–2621.
- Clark, M. B. *et al.* (2011) The reality of pervasive transcription. *PLoS Biology* **9** e1000625.
- Cox 3rd, R. S., Dunlop, M. J. and Elowitz, M. B. (2010) A synthetic three-color scaffold for monitoring genetic regulation and noise. *Journal of Biological Engineering* **4** 10.
- Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I. and Fakan, S. (2006) Chromosome territories a functional nuclear landscape. *Current Opinion in Cell Biology* **18** 307–316.
- Crick, F. H. C. (1958) On Protein Synthesis. *Symposia of the Society for Experimental Biology* **12** 138–163.
- Darwin, C. (1859) *On the Origins of Species*, John Murray.
- Darwin, C. and Wallace, A. R. (1858) On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London Zoology* **3** 46–50.
- Davidson, E. H. *et al.* (2002) A genomic regulatory network for development. *Science* **295** 1669–1678.
- Delaye, L. and Moya, A. (2010) Evolution of reduced prokaryotic genomes and the minimal cell concept: variations on a theme. *Bioessays* **32** 281–287.
- Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*, Basic Books, New York.

- Edelman, G.M. and Mountcastle, V.B. (1978) *The Mindful Brain: Cortical Organization and the Group Selective Theory of Higher Brain Function*, MIT Press.
- Eldredge, N. and Gould S.J. (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T.J.M. (ed.) *Models in Paleobiology*, Freeman, Cooper and Company 82–115.
- Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* **26** 597–604.
- Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science* **297** 1183–1186.
- Engel, G.S. *et al.* (2007) Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* **446** 782–786.
- Ferguson, M.L. *et al.* (2012) Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. *Proceedings of the National Academy of Sciences of the United States of America* **109** 155–160.
- Fiering, S., Whitelaw, E. and Martin, D. I. (2000) To be or not to be active: the stochastic nature of enhancer action. *Bioessays* **22** 381–387.
- Franco, M.I., Turin, L., Mershin, A. and Skoulakis, E.M. (2011) Molecular vibration-sensing component in *Drosophila melanogaster* olfaction. *Proceedings of the National Academy of Sciences of the United States of America* **108** 3797–3802.
- Gordon, A. J. E., Halliday, J. A., Blankschien, M. D., Burns, P. A., Yatagai F. and Herman, C. (2009) Transcriptional Infidelity Promotes Heritable Phenotypic Change in a Bistable Gene Network. *PLoS Biology* **7**: e1000044.
- Gould, S.J. and Lewontin, R. C. (1979) The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences* **205** 581–598.
- Guptasarma, P. (1995) Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* **17** 987–997.
- Hears, T. (2004) *Approche endodarwinienne de la variabilité intercellulaire de l'expression génétique*, Ph.D. Thesis, Institut national agronomique Paris-Grignon, Paris, France. (Available at <http://pastel.archives-ouvertes.fr/pastel-00001189>.)
- Hettinger, T.P. (2011) Olfaction is a chemical sense, not a spectral sense. *Proceedings of the National Academy of Sciences of the United States of America* **108** E349.
- Hooshangi, S. and Weiss, R. (2006) The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos* **16** 026108.
- Jacob, F. (1970) *La logique du vivant, une histoire de l'hérédité*. Gallimard. (English translation: *The Logic of Life, a history of heredity*, Pantheon Books, New York, 1974.)
- Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3** 318–356.
- Jerne, N. (1955) The natural selection theory of antibody formation. *Proceedings of the National Academy of Sciences of the United States of America* **41** 849–857.
- Kaern, M., Elston, T. C., Blake, W. J. and Collins, J. J. (2005) Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics* **6** 451–464.
- Kaufmann, B.B., Yang, Q., Mettetal, J.T. and van Oudenaarden A. (2007) Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biology* **5** e239.
- Klug, A. (1974) Rosalind Fanklin and the double helix. *Nature* **248** 787–788.
- Koonin, E.V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiology* **1** 127–136.
- Kupiec, J.J. (1981) Théorie probabiliste de la différenciation cellulaire. *XIIème rencontre de MÉRIBEL* 161–163.

- Kupiec, J.J. (1983) A probabilist theory for cell differentiation, embryonic mortality and DNA C-value paradox. *Speculations in Science and Technology* **6** 471–478.
- Kupiec, J.J. (1986) A probabilist theory for cell differentiation: the extension of Darwinian principles to embryogenesis. *Speculations in Science and Technology* **9** 19–22.
- Kupiec, J.J. (1996) A chance-selection model for cell differentiation. *Cell Death and Differentiation* **3** 385–390.
- Kupiec, J.J. (1997) A Darwinian theory for the origin of cellular differentiation. *Molecular Genetics and Genomics* **255** 201–208.
- Kupiec, J.J. (2008) *L'Origine des Individus*, Fayard, Paris. (English translation: *The Origin of Individuals*, World Scientific, 2009.)
- Laforge, B., Guez, D., Martinez, M. and Kupiec, J.J. (2005) Modeling embryogenesis and cancer: an approach based on an equilibrium between the autostabilization of stochastic gene expression and the interdependence of cells for proliferation. *Progress in Biophysics and Molecular Biology* **89** 93–120.
- Lehner, B. (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular Systems Biology* **4** 170.
- Lenormand, T., Roze, D. and Rousset F. (2009) Stochasticity in evolution. *Trends in Ecology and Evolution* **24** 157–165.
- Longo, G. and Tendero, P.E. (2007) The Differential Method and the Causal Incompleteness of Programming Theory in Molecular Biology. *Foundations of Science* **12** 337–366.
- Maeda, K. *et al.* (2008) Chemical compass model of avian magnetoreception. *Nature* **453** 387–390.
- Madlung, A. and Comai, L. (2004) The Effect of Stress on Genome regulation and Structure. *Annals of Botany* **94** 481–495.
- Martincorena, I., Seshasayee, A.S.N. and Luscombe, N.M. (2012) Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485** 95–98.
- Mayr, E. (1961) Cause and effect in biology. *Science* **134** 1501–1506.
- McAdams, H.H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94** 814–819.
- McAdams, H.H. and Arkin, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics* **15** 65–69.
- Merlin, F. (2009) *Chance and the sources of biological variation: a critical analysis of a multiple notion*, Ph.D. thesis, Institut d'Histoire et Philosophie des Sciences, Université Paris-1, Paris, France.
- Mettetal, J.T. and van Oudenaarden, A. (2007) Necessary noise. *Science* **317** 463–464.
- Michaelson, J. (1993) Cellular selection in the genesis of multicellular organization. *Laboratory Investigation* **69** 136–151.
- Mikeladze-Dvali, T. *et al.* (2005) The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors. *Cell* **122** 775–787.
- Morange, M. (1994) *Histoire de la biologie moléculaire*, La Découverte, Paris. (English translation: *A History of molecular biology*, Harvard University Press, 2000.)
- Murphy, K.F., Adams, R.M., Wang, X., Balázsi, G. and Collins J.J. (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Research* **38** 2712–2726.
- Newlands, S. *et al.* (1998) Transcription occurs in pulses in muscle fibers. *Genes and Development* **12** 2748–2758.
- Nijhout, H.F. (2006) Stochastic gene expression: dominance, thresholds and boundaries. In: Veitia R.A., (ed.) *The Biology of Genetic Dominance*, Landes Biosciences 61–75.
- Noble, D. (2006) *The Music Of Life*, Oxford University Press.

- Parada, L. A., Roix, J. J. and Misteli, T. (2003) An uncertainty principle in chromosome positioning. *Trends in Cell Biology* **13** 393–396.
- Pilpel, Y. (2011) Noise in biological systems: pros, cons, and mechanisms of control. *Methods in Molecular Biology* **759** 407–425.
- Polev, D. (2012) Transcriptional noise as a driver of gene evolution. *Journal of Theoretical Biology* **293** 27–33.
- Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135** 216–226.
- Raser, J. M. and O’Shea E. K. (2005) Noise in gene expression: origins, consequences, and control. *Science* **309** 2010–2013.
- Ritz, T., Thalau, P., Phillips, J. B., Wiltschko, R. and Wiltschko, W. (2004) Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature* **429** 177–180.
- Ro, D. K. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440** 940–943.
- Rojo, D. R. *et al.* (2011) Influence of stochastic gene expression on the cell survival rheostat after traumatic brain injury. *PLoS One* **6** e23111.
- Rosenberg, A. (2001) Discussion note: Indeterminism, probability, and randomness in evolutionary theory. *Philosophy of Science* **68** (4) 536–544.
- Ross, I. L., Browne, C. M. and Hume, D. A. (1994) Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunology and Cell Biology* **72** 177–185.
- Roux, W. (1881) *Der Kampf der Teile im Organismus. Ein Beitrag zur Vervollständigung der mechanischen Zweckmäßigkeitstheorie*, Engelmann, Leipzig.
- Ruvinsky, A. (2009) *Genetics and Randomness*, CRC Press.
- Schrödinger, E. (1944) *What is Life? The Physical Aspect of the Living Cell*, Cambridge University Press.
- Springer, M. and Paulsson, J. (2006) Biological physics: harmonies from noise. *Nature* **439** 27–28.
- Spudich, J. L. and Koshland Jr., D. E. (1976) Non-genetic individuality: chance in the single cell. *Nature* **262** 467–471.
- Stockholm, D. *et al.* (2007) The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS One* **2** e394.
- Szczebara, F. M. *et al.* (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature Biotechnology* **21** 143–149.
- Tamames, J., Gil, R., Latorre, A., Peretó, J., Silva, F. J. and Moya, A. (2007) The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evolutionary Biology* **7** 181.
- Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* **98** 8614–8619.
- Till, J. E., Mc Culloch, E. A. and Siminovitch, L. (1964) A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming. *Proceedings of the National Academy of Sciences of the United States of America* **51** 29–36.
- Tonegawa, S. (1976) Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America* **73** 203–207.
- Uversky, V. N. (2011) Intrinsically disordered proteins from A to Z. *International Journal of Biochemistry and Cell Biology* **43** 1090–1103.
- Velasco, J. D. (2012) Objective and subjective probability in gene expression. *Progress in Biophysics and Molecular Biology* **110** (1) 5–10.

- Watson J. D. and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171** 737–738.
- Werndl, C. (2012) Probability, Indeterminism and Biological Processes. In: Dieks, D., Wenceslao, J. G., Hartmann, S., Stoeltzner, M. and Weber M. (eds.) *Probabilities, Laws and Structures*, The Philosophy of Science in a European Perspective, Volume 3, Springer-Verlag 263–277.
- Wernet, M. F., Mazzoni, E. O., Celik, A., Duncan, D. M., Duncan, I. and Desplan, C. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440** (7081) 174–180.
- Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature* **377** 209–213.
- Zhou, T., Chen, L. and Aihara, K. (2005) Molecular communication through stochastic synchronization induced by extracellular fluctuations. *Physical Review Letters* **95** 178103.
- Zurek, W. (2009) Quantum Darwinism. *Nature Physics* **5** 181–188.