biology, diversity and probability

Christian Lanctôt _ illustration Gergely Kiss

The scientific mind maintains an ambiguous relationship with diversity. On the one hand, diversity excites curiosity, one of the driving forces of scientific inquiry. Would genetics ever have developed if all peas had the same texture or all flies the same eye color? What would neuropsychology study if the wiring of our brains were always identical? Would life even exist if DNA polymers were composed of one instead of four nucleotides? Without observation of unexplained variations, that is, observation of natural diversity, scientists could neither formulate new theories nor confirm existing ones. On the other hand, diversity makes the scientist's life terribly complicated. A monotonous world is easily explained; one in which no two snowflakes are exactly alike poses a much greater challenge. The goal of this essay is to examine how life sciences tackle this challenge and I will limit myself here to the question of whether probabilistic explanations can account for diversity in biology, as they do in physics [1].

The study of diversity may be limited to cataloguing instances of variations; encyclopaedias have been filled in this manner. Although fascinating in its own right, it is doubtful whether this activity could ever have attracted generations of curious minds as science has done for the past centuries. To put it another way, it is doubtful whether "descriptive science" – nowadays a derogatory expression – can ever entirely satisfy the inquiring mind. The reason for this is that curiosity is aroused as much by the cause of diversity as it is by its manifestations, and since man is part of this diversity, understanding it may help him to find his place in the grand scheme of things. The fascination with biological diversity is also rooted in the fact that it is intimately linked to the passage of time. I believe that its inability to control time, and hence diversity, has conditioned mankind's reaction to it. This point will become clearer after a brief historical introduction.

Confronted with diversity, the reflex of the scientific mind is to define, to group, to rank, to classify, in other words to erect artificial boundaries between events, structures or facts so that *similar* phenomena can be treated as one. The ancient Greek philosophers, distant precursors of modern scientists, were great fans of such classification schemes. It even led one of their most illustrious members to assert that classes – he called them Ideas – had an existence of their own and that tangible objects were but imperfect copies of these absolute Ideas [2]. Plato's philosophy emphasized the importance of knowing the "essence" of such Ideas, in other words the criteria used for classifying tangible things. Knowledge of these criteria was required in order to correctly deduce to which class the observations made by our senses belonged. The problem of course was to identify these criteria, as opposed to defining them *a priori*, a matter which was never satisfactorily resolved by the Greeks. In fact, as they tried to rationalize diversity, deductive reasoning degenerated into endless, pointless arguments over what actually constituted diversity. This situation may not be so surprising if one considers that classes were being defined by minds that were sometimes as diverse as the objects they attempted to classify. In some sense, deductive reasoning about diversity was defeated by the very diversity of minds practicing it.

It was inevitable that a reaction should develop against what had become a sterile approach. It did so in the form of Christianity, which solved the problem by stating that there exists only one all-encompassing class called *God*. The Christian doctrine developed as the Roman world was becoming increasingly diverse through military conquest and immigration. At some turning point in the history of every civilization, diversity is no longer considered an intellectual stimulant but rather a source of confusion and what is perceived as too much diversity is invariably equated with disorder. To restore conceptual, moral, and not least social order, the human intellect substitutes an all-encompassing "Theory of Everything" for nature's diversity. Despite its best efforts however, reason never succeeds in completely subjugating the senses and the quest for unity inevitably gives way sooner or later to a renewed appreciation for diversity. The strength and duration of a given "Theory of Everything" therefore depend on its ability to more or less accommodate diversity. In this regard, Christian philosophers offered one of the most original and effective solutions to the problem of diversity: unity is not to be found in the creatures but rather in the process of creation itself. If all things are created by a unique Supreme Being, then whether they are built alike or not becomes irrelevant. According to the Christian doctrine then, diversity is rationalized by commonality of origin, not by commonality of design or purpose and its cause is assigned to an omnipotent God. Interestingly the application of such an all-encompassing theory by the Church had its limitations and as a matter of fact not every variation was "allowed." On the contrary, as the grasp of Christianity tightened, society grew ever more intolerant of diversity. Public trials of deformed cows and mute roosters during the Middle Ages for instance showed the extent to which exceptions were rooted out by the system. But what are exceptions, if not variations that cannot be accounted for by the "Theory of Everything" [3]? What are exceptions, if not the source of diversity itself?

The religious explanation for nature's diversity could only be accepted as long as knowledge was ultimately based on divine revelation, and not on factual observation. It was therefore gradually rejected as modern science and empiricism developed. Yet diversity remained, and the "Fathers of Science" had to provide another philosophical net to contain it. They did so by resurrecting another Greek idea – ironically the exact opposite of the one the Greeks had actually favored: unity is to be found in the parts rather than in the whole itself, i.e. at some basic level all things are made of the same constituents. According to the scientific doctrine then, diversity is rationalized by supposing different arrangements of basic constituents and the generation of novel arrangements, that is the cause of diversity, is ultimately assigned to an omnipotent *Chance*. Here again, the system is rather intolerant of exceptions since the various relationships between the basic constituents are expressed by fixed, universal and immutable mathematical formulas, i.e. by physical laws. The particular genius of scientific determinism, however, is that these laws can be invoked not only to explain but also to *predict* future diversity. Thus, science re-introduced the notions of *change* and *transformation* into human thinking. This proved to be a decisive advantage over revealed religion, which, by assuming a single act of creation, also had to assume that nature was somehow frozen in time, a concept that had become irreconcilable both with the observation of natural changes and the growing will to effect political changes.

The scientific "Theory of Everything," i.e. reductionism, can only be satisfactory as long as the whole is considered the sum of its parts, a proposition that continues to be intensely debated. What is certain is that for almost three centuries advances in reductionism supported a deterministic view of nature, a particular way to account for diversity based on universal physical laws. That is, until the energy spectrum that is emitted by matter had to be explained. In order to do so, Planck had to introduce the notion of a quantum of energy, i.e. a fixed minimal energy value by which two atomic states could differ: in other words, to introduce discontinuity in the fabric of nature. Since the discrete behavior of matter implies that only a subset of possible configurations is allowed, quantum physics succeeded in placing *de facto* limits on nature's diversity. It was soon realized, however, that imposing such a constraint carried a heavy price. Indeed, quantum theory states that one cannot know the exact configuration of matter, only the *probability* that it will adopt a given configuration. To broaden this conclusion: one cannot know the class to which an observation belongs, only the probability that it falls within a given class. In essence, diversity is rationalized as a probability distribution.

If I introduced briefly some of the major and most popular milestones in the history and philosophy of modern science, it was to show that the rationale for diversity, whether dialectic, religious or scientific, is actually based on some form of negation of this diversity. Notions such as "Ideas", "God", "Scientific laws" and "Theory of Everything" are different attempts to circumscribe the incredible diversity that our senses detect. This "rationalization" is perhaps not surprising considering the difficulty for the human mind to embrace the true extent of nature's diversity. It is even less so when considering, as already mentioned, that diversity is a product of something that man is unable to control: time. Hence, the negation of diversity can be understood as a negation of time, the passage of which is much more haunting to transient beings than any amount of diversity. In this regard, the quantum physicists who established a link between diversity and probability proved to be particularly creative. When compared to religion and scientific determinism, probabilistic explanations are unique in that, by definition, *they allow for exceptions*, i.e. a probabilistic measure in nature never equals 1. It follows then that probabilistic explanations allow for time to create novelty and increase diversity. This is a very important point. Given a sufficient number of trials, in other words given a sufficient amount of time, new molecular configurations will occur even if their probability of existing is extremely low. Hence, unlike its predecessors, probabilistic thinking succeeds in integrating diversity and the passage of time. This feature should make this intellectual approach particularly useful in biology, the scientific field where diversity is arguably the most apparent and time the most important. But does it? Can probabilistic notions be of any help in explaining biological diversity? Can the varied forms and behaviors of living organisms be described in terms of probability distributions? This essay attempts to answer these questions. Before doing so however, it might be useful to distinguish between probability and statistics. Statistics are mathematical tools applied to a large number of observations made without a priori knowledge. For example, if colored balls are put in a bag, then drawing out a sufficiently large number of balls allows one to statistically model the content of the bag. No knowledge about what was put in the bag is needed. By contrast, the probability of a given event can be accurately estimated only if the various alternatives are known beforehand. Indeed, it may well be that coins were put in the bag along with colored balls, but the probability of taking one out cannot be determined in any way unless this fact is known a priori or until one has been drawn experimentally. The throwing of a dice further illustrates this point. The probability of obtaining any given face is 1/6 provided the dice is not biased. This particular information is required to assess the odds, as any gambler should be well aware! When in doubt, the gambler can always throw the dice a couple of hundred times to make sure that it is not rigged. He can perform a statistical analysis in order to estimate a probability. The notion of probability is therefore intimately linked to some sort of *a priori* knowledge, either transmitted (being told by a friend that the dice is not rigged) or acquired through experimentation (throwing the dice many times), the latter always being the most reliable form. I am not arguing *ab absurdo* that every possible condition has to be known before the probability of an event can be estimated, but I do contend that *some* statistical analysis of a system is required before the merits of a probabilistic explanation can be assessed. This is where the inherent complexity of biological systems creates difficulties. Complex biological systems are composed of a large number of interdependent parts, yet observations are typically made on one or a few parts *treated as isolated components*. The problem here is not reductionism, but rather the fact that biologists do not yet agree on what constitutes the proper level of observation to carry out a statistical analysis. How can evolution be modeled as a probability distribution if no agreement can be found about whether selection acts at the level of the species, the group, the individual or the gene? Statistical analysis in biology is further complicated by the fact that the number of observations is usually very low because of experimental limitations. I think that biologists instinctively realize the incompleteness of statistical analyses in their field, and that this is the reason why probabilistic thinking has not taken hold of the biologist's mind as it has the physicist's. This resistance is obvious in the field of genomics. Gene expression profiling experiments generate an impressive amount of data; to a biologist at least. Statistical analysis of the results is routinely carried out. Yet the global gene expression of individual cells is not viewed as a probability distribution but rather as the result of a fixed genetic program. The idea that any population of cells could in reality express a Gaussian distribution of genetic loci is anathema to the dogma of molecular biology. In fact, genetic determinism is so entrenched that molecular and cell biologists often prefer to speak of stochastic events instead of probable ones. The difference is important: stochastic events are the result of random, *non calculable* fluctuations whereas probability is a *calculable* mathematical function. I think that molecular and cellular biologists are right not to commit themselves to probabilistic explanations. What is often missing in biological equations is not only the value of the variables but their identity and significance as well.

There is one prominent issue in biology which can be phrased in terms of probability: the origin of life. Whether the probability that life were to appear on our planet is close to 1, as the Structuralists maintain, or close to 0, as most Darwinists maintain, both sides would agree that the question can be stated in terms of probability. It is therefore interesting to inquire into what makes the problem of the origin of life amenable to probabilistic thinking. The answer may lie in the parallels between this particular biological question and the physical sciences. 1) The transition from inert to living matter, like the quantum leap, is viewed as a single, irreducible event. Discontinuity is accepted *a priori* and the only diversity that is allowed is the inert/living dichotomy. 2) As in quantum mechanics, the observer perturbs the analysis since he defines *a priori* what he considers to be living [4]. For instance, viruses may or may not qualify as living entities depending on whether one considers the ability to adapt or the ability to self-perpetuate as the essential attribute of life. 3) Perhaps most significantly the problem of diversity is deferred, if not obviated, in the minds of reductionists by assuming the appearance of a single form of life in time and space. Notwithstanding Occam and his razor, is it not possible to postulate that the primordial environment favoured the *simultaneous* formation of a variety of self-replicating macromolecules (nucleic acids, micelles, prions, etc.)?

Unlike what happens with the origin-of-life problem, biologists cannot ignore diversity when turning their attention to the progeny of the first living organism, "endless forms most beautiful and most wonderful." On the contrary, they must squarely face nature's diversity: the tree of life may well have only one root but it clearly has a multitude of branches. Can their sprouting be rationalized through probability? Does the appearance of diversity follow probabilistic rules? To answer, I must return to the issue of calculability, one that is central to probabilistic thinking. The probability that the molecules of oxygen will cluster in one corner of a room is infinitesimal, yet *it can be calculated*. Are we ready to make similar claims concerning ontogenetic and phylogenetic events? Can we calculate the probability that *Pan troglodytes* were to evolve into *Homo erectus*? Is there an infinitesimal probability that a human egg will ever develop into a shrimp? The obvious answers to these questions should make it clear that at present probabilistic notions are of little help in solving the problem of biological diversity.

As I have alluded to, physicists were the ones who introduced probability in the realm of science. It may be instructive to review the circumstances that led them to do so. Indeed, a number of historical conditions had to be met before the new paradigm could even be considered. First, reductionism had to provide a framework in which all parts of a system, as well as the *nature* of their mutual interactions, are known [5]. As we have seen, this framework could only exist after quantum theory had placed *de facto* limits on nature's diversity. Second, and historically this most important step preceded the first one, it was essential to assume that the macroscopic world is in fact a statistical sampling of elementary constituents. Only then can an observation (e.g. temperature increase upon compression of a gas) amount to the result of a statistical analysis (e.g. movements and collisions within a cloud of atoms). Only then can natural phenomena correspond to the probability of a given behavior. Only then can nature's diversity be represented as a probability distribution. Needless to say, biology has not yet taken this path. Even the most ardent reductionist biologist would not dare to claim knowledge of all the constituents of a biological system, not to mention their mutual interactions. Furthermore, unlike quantum physicists, biologists have not yet imposed constraints on nature's diversity : potentialities remain wide open, neither determined nor probable. Finally, biologists do not interpret the observations they make on a macroscopic level as a direct statistical analysis of a very large number of events occurring at a lower level. If they did, the discovery of genes associated with Mendelian diseases would not be heralded as the identification of *causative* agents but as that of probability determinants.

⌂

Considering the points that were raised in this essay, I venture to conclude that probabilistic thinking is at present inadequate to explain most biological observations. Until biological reductionism reaches a point that allows true statistical analysis of living systems, probabilistic explanations only serve to mask our ignorance [6]. As a closing remark, I should mention that probability did not enter the scientific discourse alone ; its famous companion was uncertainty. Together, these concepts have brought the scientific mind to new levels of abstraction and opened up broad avenues of inquiry. Will biological thought ever reach this level of abstraction ? Probably.

Notes

[1] In particular in quantum mechanics where the action of measuring a physical quantity representing a system or part of a system is subjected to an intrinsic uncertainty on the outcome.
[2] A parallel can be drawn between this faulty copying process and modern evolutionary theory, which stipulates that diversity (i.e. mutants) arises because of DNA replication errors.
[3] How can God, a perfect Being, create imperfect ones ?
[4] In reality, the observer has nothing to "observe". The empirical data on the origin of life is no longer available : it has evolved for the past 3 billion years. The question of the appearance of life on Earth is an historical one. The fact that any answer can only be a probable one makes it all the more interesting.
[5] The easiest way to know all parts of a system is to assume that it is composed of a very limited number of parts, the unassailable premise of reductionism.
[6] Like developmental constraints, gene redundancy, junk DNA, etc.