1

From natural selection to the history of nature

For their key role in the revival of heliocentrism, we may certainly excuse Copernicus, Kepler and Galileo for being less successful at explaining *why* planets move. It was only 150 years after Copernicus anonymously circulated his pamphlet *Commentariolus* in 1510 that Isaac Newton's *Principia Mathematica* offered a theory of planetary motion and established the first modern paradigm for celestial mechanics. Darwin's theory of evolution represented a scientific revolution of the same importance and magnitude, but its historical development bears a crucial difference from heliocentrism. Despite a number of forerunners in Britain and France and the work of his contemporary Alfred Wallace, most of us associate both the substantiation of biological evolution as a fact, and the theory of natural selection as evolution's first and only theoretical paradigm, with Charles Darwin and the *Origin of Species* (1859). First, Darwin made species intelligible by setting them in motion as Copernicus had done with the planets. Biological evolution, or the movement of species in time, implied that the ultimate cause of species features is their history. And second, Darwinism found its fundamental law of motion in the principle of natural selection, the consequence of variation and competition within populations. For his contributions to the facts and theory of evolution, Darwin can be seen both as the Copernicus and the Newton of biology (Padian, 2001).

Although the triumph of Darwinism is denied by few, scientific revolutions also demonstrate that even the most successful and influential theories may be wrong or incomplete. The Newtonian paradigm, for example, started to crumble when some of its predictions at the very large (cosmological) and the very small (atomic) physical scales were refuted by experiments and theory; although still an essential component of physics, the Newtonian world picture is no longer accepted as a satisfactory account of the nature of space, time and physical causality (Penrose, 1989). For us, it is important to notice that the major debate involving Darwinism since its origin is also a matter of scale, namely the relationship between microevolution and macroevolution. This is more than coincidence. While natural selection is widely accepted as the microscopic cause of change in populations, when it comes to accounting for the big picture of life in all its forms, living and extinct, it has become commonplace to question the power of selection against a background of virtually infinite and unique events accumulating for billions of years in every line of descent, and then to appeal to contingency and history as the true cause of evolution at larger scales (Gould, 2002). In summary, while Darwinism introduced natural selection as a fundamental law of biological change, it established at the same time that nature had a history; for this reason, the main challenge of evolutionary theory since the *Origin of Species* has been to discover the missing links between the *law* of nature and the *history* of nature.

NATURE AFTER NATURAL LAW

The theory of evolution by natural selection was a late product of the Modern Scientific Revolution and its emphasis on efficient causes and natural law (Cohen, 1987). The rise of natural law meant that science should rely on purely mechanistic explanations, and the study of life should be no exception. The biologist, even when familiar with the appearance of design in living organisms, should 'disregard for a time, as in physical philosophy, the immediate *purposes* of the adaptations which he witnesses; and must consider these adaptations as themselves but the *results* or *ends* of the general laws for which he should search' (Carpenter, 1839: 461). The principle of natural selection provided biologists with a classic example of general law, but its formulation exhibited a distinguishing feature. Newtonian dynamics and the field theories of electromagnetism and general gravitation were originally formulated as sets of deterministic laws (Penrose, 1989). As such, their experimental predictions were meant to be exact (apart from measurement error): given similar conditions, all objects were expected to free-fall predictably, all negatively charged bodies were expected to respond equally to an electric field, and so on.

In contrast, Darwin saw natural selection from the beginning as a statistical or probabilistic principle (Hempel, 1966). The aim of statistical laws is to extract information from systems or populations when the behaviour of their components is either unknown or unpredictable (Menand, 2002). A good example of the logic of statistical laws is the analysis of radiocarbon decay into nitrogen (Bowman, 1990). A radiocarbon (^{14}C) is a carbon atom with six protons and eight neutrons, instead of the six neutrons found in the more common and stable ^{12}C isotope. In the Earth's atmosphere, it is produced when a thermal neutron derived from cosmic rays collides with a ^{14}N atom (the most common nitrogen isotope, with seven protons and seven neutrons) and converts one of its protons into an extra neutron. The resulting ^{14}C atom is unstable and eventually decays back into ^{14}N by emitting a β-particle. In living organisms, a constant ratio of ^{14}C to ^{12}C is maintained on the one hand by the absorption of new radiocarbons into organic matter through photosynthesis (and their subsequent transfer to organisms that feed on photosynthetic plants and algae), and on the other hand by their loss through decay. But after death there is no further absorption of radiocarbons, and the decay of the ^{14}C trapped into the dead organism causes their ratio to ^{12}C to decrease gradually with time.

The work of Willard Libby and collaborators (Arnold and Libby, 1949) showed that such decrease followed an exponential curve and could be used as a method for dating organic samples, provided only that the parameters or calibration of the curve were obtained from fossil or archaeological material of known dates. Among other things, Libby's exponential model showed that 50% of the radiocarbon atoms trapped in a dead organism are expected to decay into nitrogen after 5568 ± 30 years. In other words, if we start with a sample of one million radiocarbon atoms, half a million are expected to disappear after 5568 years, plus or minus an error of 30 years. The error exists because the calibration method itself is based on a finite sample of ^{14}C atoms: larger initial samples of radiocarbon atoms would reduce the error. Thus, provided the amount of organic material in the specimen (i.e. the trapped radiocarbon sample) is large enough, dating of fossils can be very precise.

However, precision completely vanishes at the level of individual atoms (Gell-Mann, 1994). The reason is that radiocarbon decay is spontaneous and unpredictable: whether a particular ^{14}C atom decays instantly after its formation in the atmosphere, decays after a thousand years, or never decays, depends on contingent factors – namely, the outcome of an unrehearsed and never-ceasing dance of particles in the atomic nucleus (Libby, 1955). For this reason, although we know almost with certainty that 50% of radiocarbon atoms in a large sample will decay in about 5568 years, all we can say about an *individual* radiocarbon atom is that the probability that it will decay in less than 5568 years is 50%; we cannot determine when the decay will happen, or whether it will happen at all.

The point is that the principle of natural selection relies on the same distinction between population-level predictability and individual-level uncertainty. Natural selection is, after all, a 'statistical bias in the relative rates of survival of alternatives' (Williams, 1966: 22). Based on the radiocarbon example, natural selection may be described as the decay or rise of allele frequencies across generations, of phenotypic traits towards certain character states, or of organisms towards higher levels of fitness. Provided certain conditions are met, it is even possible to predict fairly accurately the direction and speed of change in measurable traits or gene frequencies in populations (Charlesworth *et al.*, 1982).

However, as far as individual fate is concerned, the struggle for survival is to a large extent a matter of chance. This is because an individual carrying the genes and phenotype corresponding to a fitness peak may (and very often does) simply fail to survive or reproduce for a variety of contingent factors, such as an accident, a disease, or a competent predator. In other words, under the hypothesis of natural selection the link between individual variation and individual fitness was never meant to be deterministic. We can certainly describe or even sometimes predict changes in proportions of alleles or phenotypes in populations; and based on that, we may ascribe individuals a probability of reproductive success. But as in the case of radiocarbon decay, the difference between individual and population scale is all there is between chance and order.

REDUCTIONISM AND EMERGENCE

A possible interpretation of the differences between deterministic and statistical law is that they imply a distinction between exact and historical sciences. If natural selection or the law of evolution itself rests upon statistical ground, how can we avoid the conclusion that evolutionary history is to an even higher degree the domain of contingency and unpredictability? The view that historical sciences deal with subjects too complex to be grasped in terms of exact causality dates back at least to Laplace (1774):

> Chance has no reality in itself; it is only a term fit to designate our ignorance concerning the manner in which the different parts of a

> phenomenon are arranged among themselves and in relation to the rest of Nature.

According to Laplace, chance is not an objective property of things. Parts are always well ‘arranged among themselves’ as described by deterministic laws; the problem is that our brains are limited devices, able to calculate the outcomes of a few interactions, but not an exact event equation whenever the event consists of too many parts or interactions – that is, whenever the event is too complex. When too many parts or interactions are involved, Laplace argued that it was necessary to appeal to a statistical approach and to the use of averages and error estimates. In summary, chance and uncertainty are not objective properties of entities, but effects of scale transitions from simple to complex; in our mind, those transitions are manifested as the transition from determinism to probability.

Despite its intuitive appeal, reasons for rejecting the Laplacean view have accumulated for more than a century. The advance of reductionism, the belief that ‘all the complex … things and processes that we observe in the world can be explained in terms of universal principles that govern their common ultimate constituents’ (Nagel, 1998), has provided compelling evidence that the very foundations of reality, or its ‘simplest’ elements, lie on uncertainty and probability (Feynman, 1965). The most fundamental laws of atomic physics and quantum mechanics, such as Heisenberg’s principle of uncertainty, Pauli exclusion principle and Schrödinger’s wave function, are probabilistic laws; even when applied to a single electron, they imply respectively that simultaneous measurements of variables such as speed and position will necessarily exhibit greater-than-zero variance and confidence intervals, that there is only a probability that it will lie at a certain region around the atomic nucleus, and that there is a certain probability that it is at a given quantum state (Penrose, 1989).

But if uncertainty is the fundamental rule of reality, why does the world around us appear so orderly? The current answer, which completely inverts the logic proposed by Laplace, is complexity itself: while simple parts or ultimate constituents of matter such as atoms or atomic particles are intrinsically uncertain, predictability often emerges at the larger macroscopic scale as a very special kind of statistical effect; this effect we call physical ‘order’.

> Only in the co-operation of an enormously large number of atoms do statistical laws begin to operate and control the behaviour of these assemblées with an accuracy increasing as the number of atoms

> involved increases. It is in that way that the events acquire truly orderly features. All the physical and chemical laws that are known to play an important part in the life of organisms are of this statistical kind; any other kind of lawfulness and orderliness that one might think of is being perpetually disturbed and made inoperative by the unceasing heat motion of the atoms (Schrödinger, 1944: 10).

Thus, it is only *because* of the large number of atoms and their complex interactions in ordinary macroscopic objects such as a table that we grasp an appearance of order. Well before the expression became fashionable, Schrödinger was describing our surrounding reality as an 'emergent property' (Bunge, 2003) of a complex system of innumerable microscopic parts. He summarised his views in the classic 'Principle of Order from Disorder', or the origin of macroscopic predictability from microscopic uncertainty. Any other kind of lawfulness claimed by science – in particular the deterministic lawfulness implied by classical physics – was illusory, or just the name we give to our ignorance of the probabilistic roots of reality.

As importantly, for Schrödinger the Principle of Order from Disorder was valid for physical laws too. An electromagnetic field, originally described by Maxwell through deterministic laws, is in a sense just like a table. It is produced when a source (for example an atomic nucleus) emits so many photons that both the probability of a passing electron absorbing them and the number of photons absorbed depend only on the position of the electron in space and time (Feynman, 1985: 122). For this reason, the predictable and seemingly deterministic response of any electron to a magnetic field is in fact the orderly appearance of a system involving countless probabilistic interactions between many photons and the passing electron. If in another example we replaced the photon source with the Sun, and the passing electrons with two glasses of water sitting on a beach, we see that although each interaction between photon and water molecule is unique and probabilistic, the liquid from each of the two glasses would take very similar times to evaporate.

In summary, the dichotomy between chance and determinism proposed by Laplace was radically inverted by contemporary science. Rather than exact expressions of pure and deterministic physical laws, simple interactions and particles of reality are intrinsically uncertain and probabilistic; and rather than the cause of ignorance, complexity or the statistical composition of entities and their interactions is the very reason for predictability and order that we observe around us.

Remarkably, Darwin seems to be closer to Schrödinger than to Laplace; in fact, his definition of natural selection in the *Origin* did not sound too far from the contemporary views on complexity and order:

> Then, considering the infinite complexity of the relations of all organic beings to each other and to their conditions of existence, causing an infinite diversity in structure, constitution, and habits, to be advantageous to them, I think it would be a most extraordinary fact if no variation ever had occurred useful to each being's own welfare … But if variations useful to any organic being do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life … This principle of preservation, I have called, for the sake of brevity, Natural Selection (Darwin, 1859: 127).

For Darwin natural selection was a pattern or principle of preservation observed in populations and across generations, resulting from virtually infinite interactions among organic beings; but at the level of isolated individuals, the best we can do is to ascribe a better or worse chance of being preserved. Natural selection is another example of natural law, and of order emerging as a scaling effect against a background of individual events inextricably tangled with chance.

HISTORY AND INHERITANCE

If the ultimate components of reality are intrinsically unpredictable, what is the appeal of the 'simple' in the natural sciences? In other words, why is reductionism so influential? The answer is that reductionism was never meant to be a tool against chance; it is a tool against *history*. An enlightening example is given by the work on antiparticles by the physicists John Wheeler and Richard Feynman. According to Dirac's antiparticle theory, every existing electron implied the existence of a positive sister particle (a positron) somewhere in the universe (Wheeler and Ford, 1998: 117). Wheeler then proposed a very original explanation for why electrons and positrons existed in pairs: all observed electrons and positrons would be a single existing particle, travelling back and forth in time and drawing a continuous 'world line' or trajectory of its existence. This single electron could be seen many times simultaneously, as its swaying world line intersects the present multiple times and creates the perception of many existing electrons. When the world line crosses towards the future, an electron is observed, and a travel back to the past would correspond to a positron; hence their similar mass, but inverse charge. Finally, since there is a single continuous line moving back and forth, it follows that between two travels to the future there must necessarily occur a travel to the past; hence the existence of one positron for every electron.

Although Wheeler's idea of world lines was incorporated into quantum mechanics, explaining antiparticle theory as a cosmic *déjà vu* was considered too extreme. Wheeler was aware of flaws in the argument, but what matters for us was his justification for reducing all electrons and positrons to a single particle:

> We realized that such a way of thinking made sense because of the extreme simplicity of the electron. You can tell by looking into a person's face something about what that person has been through. You cannot tell anything about an electron's history by looking at it. Every electron is exactly like every other electron, unscarred by its past, not blessed with a memory – of either human or the computer variety (Wheeler and Ford, 1998: 348).

Since electrons do not have memory and do not bear a record of their past accidents and scars, they lack individuality. Exactly for this reason, they are the ideal starting point for the study of more complex macroscopic entities that do display physical memory and uniqueness. Reductionism thus serves two main purposes. The first is the understanding of things composite or complex, such as tables or electromagnetic fields. The second purpose is the conquest of *history*. While statistical composition generates order from disorder as claimed by Schrödinger, it may also generate history or a record of past events when time is a relevant variable (i.e. when systems are not in equilibrium). This category of problem has come to the fore only recently in the physical sciences as the study of dynamical systems (Nicolis and Nicolis, 2007) and shows that the concept of history can be extended well beyond its original use as a feature of human societies.

Galaxies have persistently eluded cosmologists for their gigantic scale and morphological variability accumulated over billions of years. Spiral galaxies are currently the prevailing form in the universe, but the reasons for their larger number were not obvious (Smolin, 1997). Some sophisticated thermodynamic (*n*-body/gas dynamical) models have been proposed but were not particularly successful; surprisingly, until recently the best simulations of spiral arm dynamics were derived from evolutionary biology, more specifically from simulations of host–parasite interactions, with the forming stars in a spiral arm sweeping through the galaxy playing the role of infecting parasites, and the dark matter that becomes the material source of the forming stars playing the role of hosts.

In a review suggestively named ‘The morphological evolution of galaxies’, Abraham and van den Bergh (2001) pointed to further similarities between cosmology and organic evolution. They recognise that cosmology also needs to cope with taxonomic problems (the standard ‘tuning fork’ classification scheme proposed by Hubble in 1926 fails to describe most of the distant galaxies), phylogenetic problems (spiral galaxies seem to be primitive, while elliptical galaxies are probably derived from collision of spirals; however, elliptical galaxies may revert back to spirals due to gas accretion) and especially evolutionary problems:

> At this stage, all that is clear is that the morphology of spiral galaxies is evolving rapidly and systematically, even at quite low redshifts [i.e. at shorter distances from the Earth]. Familiar types of galaxies, such as barred and grand-design spirals, appear to be relatively recent additions to the extragalactic zoo. The nature of the many morphologically peculiar galaxies at high redshift remains a complete mystery. These objects might be mergers, protogalaxies, new classes of evolved systems, or a combination of all three (Abraham and van den Bergh, 2001).

The use of various biological terms and the analogies with organic evolution is possible because galaxies are historical objects whose number, size and configuration depend not only on first principles, but on original conditions and accidental factors occurring over long periods of time. In other words, physical sciences also aim at the eventual conquest of complex historical entities, and this blurs the boundaries between ‘exact’ and ‘historical’ sciences. This also means that the reliance of evolutionary biology on the statistical law of natural selection does not tell it apart from the other natural sciences: first because natural laws are statistical, or attempts to grasp patterns against a background of contingency; and second because even the physical sciences aim at accounting for the history of nature.

Undeniably there is a fundamental difference between physics and evolutionary biology, but contrary to the common belief it does not lie in a distinction between exact and historical sciences: the distinction is in the type of *memory* behind physical and organic history. Both galaxies and cells must bear some principle of order that opposes their decay towards randomly moving atoms and molecules; but the principles operating in living and non-living entities are different. As seen above, for Schrödinger the Principle of Order from Disorder operating in non-living matter is the result of statistical averaging, and 'is never the consequence of a well-ordered configuration of atoms'. In contrast, Schrödinger postulated a second Principle (of 'Order from Order'), unrelated to statistical averaging and derived from organisation at the level of atoms and molecules, perhaps in the form of 'aperiodic crystals' within chromosomes.

Less than a decade later, the structure of DNA, the aperiodic crystal behind living organisation, was determined. It is now known that in all living organisms macroscopic order, or phenotype, is symbolically stored in a microscopic sequence of nucleotides or genotype. Since the possible configurations of DNA are virtually unlimited but resources are not, selection arises as the result of competition between alternatives (Eigen, 1971). As discussed below, unlimited inheritance, the material and functional distinction between genotype and phenotype, and competition in the form of natural selection explain why there is more history and complexity to a single cell than to a galaxy.

DARWINIAN PROGRESS: ORDER AT THE MACROEVOLUTIONARY SCALE

Organic history is the record of the scars, accidents and contingent factors crossing the path of evolutionary lineages; but, as seen, contingency must coexist with the operation of the universal law of natural selection. Darwin believed that the logic of inheritable variation and competition would apply not only to individuals within species, but also to the origin, divergence, extinction and succession of species themselves. The title *On the Origin of Species by Means of Natural Selection* is in fact a statement of Darwin's controversial theory (which later became known as 'reinforcement') that speciation between sympatric (physically overlapping) populations can evolve as an adaptation, rather than as a consequence of other adaptations or of accidental factors (Coyne and Orr, 2004). Darwin did not deny that speciation occurs mostly due to allopatry (physical isolation) and therefore by accident, but emphasised that species resulting from allopatry derived their existence from the absence of competition rather than selection of the fittest. Owing their existence to luck, they were destined to play a secondary role in the history of life.

> I conclude that, although small isolated areas probably have been in some respects highly favourable for the production of new species,