Cambridge University Press
978-0-521-83568-8 - Deep Learning: How the Mind Overrides Experience
Stellan Ohlsson
Excerpt
More information

PART ONE

INTRODUCTION

1

The Need to Override Experience

Upon those who step into the same stream ever different waters flow.
Heraclitus[1]

Nothing so like as eggs, yet no one, on account of this appearing similarity, expects the same taste and relish in all of them.
David Hume[2]

Life is change. Natural forces continuously knead our material environment as if it were so much dough in the cosmic bakery, and few features of our social environment remain constant for even one generation. How do human beings live in ubiquitous change? How do we cope with, and adapt to changes in our environment? How do we initiate change and create novelty? Due to late 20th-century advances in the natural sciences, these questions are more difficult to answer than they once seemed.

Few changes are so beguiling to children, poets and whomever else has the good sense to stop and look as a change of season. In four-season climes, the visual transformation of the landscape as the dominant color moves from green to red and on to white is stunning. The spectacle of an early autumn snowstorm must have overwhelmed the first band of hunter-gatherers to push north into a temperate climate zone and convinced them that their world was coming to an end.

Yet, children, poets and hunter-gatherers are wrong; striking as it is, seasonal change is no change. Winter displaces summer, true enough; but after winter, summer returns, and all is as it was. The change is cyclic, hence stable; hence not a change. The world remains constant; it is merely going through motions. According to this view, change is an illusion because the fabric of reality is a weave of eternal laws of nature.

It follows that creatures who can remember, reason, imagine and plan can respond to the superficial changes by accumulating experiences, extracting the underlying regularities, projecting them forward in time and acting accordingly. To survive, the hunter-gatherers only needed to see through the transient snowstorm to the repeating sequence of seasons, predict the return of summer and stockpile food. This picture of the interlock between world and mind has dominated Western intellectual traditions since the beginning of written scholarship.

As it turns out, this picture is a mirage. We live in a world in which change is no illusion. Adaptation to this world requires cognitive capabilities over and beyond those implied by the traditional view.

A CLOCK SO TURBULENT

Since the scientific revolution of the 17th century, the natural sciences have scored astonishing successes by viewing nature as an unchanging machine, a kaleidoscope that generates shifting appearances.[3] Searching behind the complex appearances of the night sky, astronomers found a simple geometric system of spheres traveling around the sun in elliptical orbits that are fixed by eternal laws and universal constants.[4] The changes in the night sky from dusk to dawn, from day to day and from month to month are clearly visible and yet no changes at all, merely the way the stable planetary system looks from the limited perspective of an earthbound observer.

On the surface of the Earth, pendulums, projectiles and pulleys turned out to be understandable as instances of a single category of mechanical motion.[5] All such motion is shaped by the constant force of gravity. Over time, physicists came to realize that an observer has a choice of reference frame and that an object that appears to be in motion within one frame appears to be at rest in another. Hence, motion – a change of place – is not a genuine change, but the way an object appears from certain points of view. Reality is captured in the mathematical equations that specify the translation from one frame of reference to another, and those equations are invariant.

Astronomy and physics are not the only success stories in the search for unchanging regularities. Looking behind the multitude of material substances, each with its own color, melting point, smell, taste, texture, weight and so on, chemists also found a simple system consisting of a short list of building blocks, the atomic elements, and a handful of mechanisms – the co-valent bond, Van der Waals forces – for combining them into larger structures, molecules, that determine the observable properties of material substances.[6] A chemical reaction might transform its reactants into very different substances and yet there is no fundamental change, only a rearrangement of the eternal building blocks.

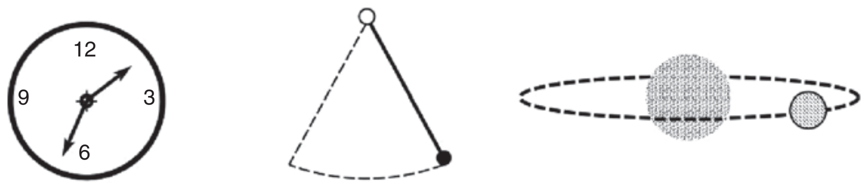

FIGURE 1.1. For clockwork systems, past behavior can be extrapolated into the future.

In these and other scientific breakthroughs, Nature appears as a clock, a machine that endlessly repeats the same movements. Figure 1.1 illustrates the analogy. The movements drive superficial events like the successive displacements of a clock's hands or the successive phases of the moon, but the underlying machinery does not change. Quoting science pioneers Robert Boyle, René Descartes and Johannes Kepler, historian Steven Shapin writes: "Of all the mechanical constructions whose characteristics might serve as a model for the natural world, it was the *clock* more than any other that appealed to many early modern natural philosophers."[7]

Scientists found that they could describe clockwork nature with linear differential equations, a mathematical tool for deriving the future state of a system, given two pieces of information: the current state of the system and the equations that describe how each property of the system is related to its other properties. Time plays an important role in such equations, but it is symmetrical between past and future. The physicist's equations can be used to calculate the position of the moon a hundred years ago as easily as to predict its position a hundred years hence. Because nothing truly changes in a clockwork world, past and future are mirror images.

The success of this approach to nature compelled scientists to regard natural systems that can be understood from within the clockwork mind-set as prototypical models of nature. Systems with a transparent link between the changing appearances and the underlying mechanism became scientific showcases, dooming generations of schoolchildren to the study of pendulums, projectiles and batteries hooked up to lightbulbs. The clockwork model was so successful that little importance was attached to the fact that not every natural system conforms. A system that did not behave like a clock was assumed to represent either a temporary state of ignorance about how to analyze it or an unimportant or peripheral part of nature.

In the last two decades of the 20th century, scientists surprised everyone, including themselves, by formulating a fundamentally different view of material reality. Consider a mountain river. The waters along its banks are forever shifting. The appearance of eddies and whirlpools cannot be predicted with the linear differential equations of clockwork physics. A hydrologist might try to describe the action of the river waters in terms of the movements of small, imagined cubes of water, but such a description does not predict when and where eddies will form. It does not help to appeal to probability. Descriptions in terms of statistical aggregates only work when the underlying system is random, and river water fails to conform. The scientist labors over his description, but the river rages on, spraying water in his face; its turbulence is irreducible.[8]

Mountain rivers exemplify a class of natural systems that scientists have come to label chaotic, complex, dynamic and nonlinear; for brevity I will use the single term *complex*.[9] Examples of complex systems include anthills, earthquakes, epidemics, forest fires, storms and volcanoes. As a group, complex systems are characterized by properties that do not fit, or can only be made to fit with great difficulty, into the clockwork model.

Instead of cycling through a fixed orbit forever, complex systems come into being, run their course and cease. That is, they have a *history*. A tree sprouts, grows, matures, topples and rots; a storm gathers, releases its energy and abates. The historical character of a system is difficult to establish when the system is so large that it cannot be encompassed in a single glance and when it lasts considerably longer than a human life, but scientists have found that climate zones, continents and ecosystems are historical entities. According to the Big Bang theory favored by some cosmologists, the universe itself might have a history.[10] It began as an incomprehensibly dense kernel that exploded and expanded rapidly, creating both space and materia in the process. The universe is still expanding, but it might one day reverse direction and contract, eventually ending in a Big Crunch.

In historical systems, past and future are not mirror images. Changes are *irreversible*, not repeating or cyclic. A river, once it reaches a plain, meanders.[11] It slowly and patiently alters its course, becoming more serpentine over time. A meandering river might be altered by a variety of processes, but it will not spontaneously straighten itself again. The ever more pronounced bends do not represent a phase in the life of the river that alternates with a phase of ever shallower bends. Likewise, a volcano rumbles, erupts and becomes dormant again. The resulting changes are not mere appearances. An eruption can fragment a mountain and replace it with a crater, an event that might have consequences for the biology, geology and weather of the surrounding area. Those consequences can be further modified by subsequent geological processes but not reversed; once the mountain has exploded, it does not come back together again. Unlike the cyclic processes of clockworks, such changes are directional with respect to time; they progress or unfold from past to future.[12]

In complex systems, changes are not illusory, no mere surface appearances driven by a constant causal machinery hidden behind the appearances. Instead, change is *thoroughgoing*. The turbulence of a river is not *caused* by the behavior of the water; it *is* the behavior of the water. Wind gusts are not *indicators* of a storm in the sense in which the hands of clock are indicators of the clock's internal state; they *are* the storm. There is no stable reality underlying the ceaseless movements; there are only the movements themselves.

If the changes appear regular, we have no guarantee that those regularities are themselves stable over time. Even in the strongholds of the clockwork mind-set, astronomy and mechanics, scientists discuss whether the laws of nature are the same everywhere in the universe and at all times.[13] Is the gravitational constant – the celebrated $g = 9.81\ m/s^2$ that plays a central role in Newtonian physics – one of the eternal constants of the universe or a variable that slowly drifts from value to value as the universe expands? If changes in constants and laws are themselves regular, we have no guarantee that those second-order regularities are stable. Reality might be turbulent all the way down.

Complex systems have to be understood in terms of multiple *system levels*. At each level, system components exhibit characteristic properties and interact in characteristic ways to determine the properties of the next higher system level. The prototypical examples are the particle-atom-molecule-substance sequence of material science and the cell-organ-organism-species sequence of biology. Each system level is associated with a characteristic *scale* of complexity along size, time or some other dimension. The interactions among the components and processes at system level N propagate upward to shape the components, processes and system properties at level $N+1$. The propagation process can operate in different modes and exhibit different properties, depending on the characteristics of the system.

Some systems consist of components of qualitatively different types, interacting in qualitatively different ways. Typically, there are only a few, perhaps only a single component of each type. The human body and a car engine are examples. There are only two kidneys, one heart and one stomach, and the kidneys interact with the heart in a different way than with the stomach. Similarly, there are only a few cylinders in a car engine, and they interact differently with the fuel injector than with the differential. In systems of this type, the fine-grained details of one level seldom matter at higher levels. It does not matter *how* the heart pumps blood. Some individuals have lived for some time with artificial hearts that operate very differently from an organic one. But it matters greatly *that* the heart pumps blood. Similarly, if we replace the combustion engines in our cars with electrical engines, the internal structure and functioning of the cars will change radically, but there will be little effect on the traffic system. We will still have traffic jams, rush hours, speeding tickets, parking shortages, teenagers wanting to reserve the family car for Saturday night and so on. It does not matter from the point of view of the traffic system *how* an engine makes the car go, but it matters greatly *that* it makes the car go. In general, only gross properties of system level *N* punch through to become causes at higher system levels. I refer to this flavor of scaling as *direct impact*. It appears primarily when the operation of the higher-level system is dependent on a single, unique component at a lower level.

Systems with many components often exhibit *cascading causation*: Properties at system level *N* cause phenomena at level *N+1*, which in turn propagate upward to determine system characteristics at yet higher levels. The cascade can be *dampened*, so that the effect is smaller and smaller at each successive level. The accidental death of a single fox will have some effect on the local prey population, but at the level of the entire ecosystem, Mother Nature takes the fox's demise in her stride. The important cases are those in which the causal cascade is *amplified*, so that the consequences grow in magnitude from level to level. Even a minor change in, for example, the average global temperature can trigger processes of climate change: melting of polar caps, alterations in the flow of the ocean currents and so on. Amplified propagation of minor perturbations is popularly called a "butterfly effect," but is technically labeled *sensitivity to initial conditions*. Amplified cascading causation makes systems *massively contingent* on the exact properties and interactions of their components, one source of unpredictability.

Cascading causation can create patterns at a higher system level that are *emergent* – that is, impossible, at least in practice, to predict from a description of the lower system level. The meandering of a river can once again serve as example.[14] What makes a straight river develop ever more loopy bends? A river carries sediments – silt, sand and gravel – and the rate of sedimentation depends on the speed of the river. Where the river turns around a bend, the waters along the inner bank slow down and the sediments sink and become deposited on the bottom, extending that bank. At the same time, the rapid passage of the waters along the outer bank excavates that bank. As a result of these simultaneous processes, one riverbank will become thicker and grow toward the middle of the river, while the opposite bank is being scooped out and hence recedes in the same direction. The combined effect of these two processes is to

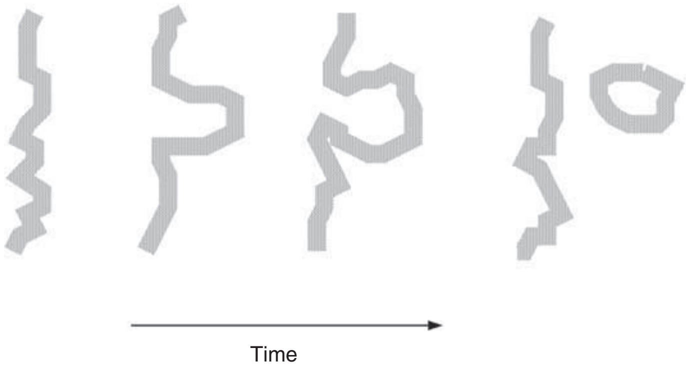

FIGURE 1.2. Successive views of a meandering river and the creation of a circular lake.

move the river sideways. Repeat this process at any point along the river where the two banks are such that the waters flow at different speeds, and the result is the snake-like shape we see in an aerial view of a meandering river. The river might spin off a small circular lake where a bend becomes so pronounced as to reconnect with itself; see Figure 1.2. Standing on the bank where a mountain river meets a coastal plain, a wanderer can observe these component processes without any other instruments than a pair of eyes hooked up to a brain. Nevertheless, the concept of *a small, circular lake with an island in the middle* does not fall out of statistical manipulations of immense numbers of water molecules. It is an emergent feature of rivers.

Direct impact and cascading causation contrasts with a third flavor of scaling called *self-organization*.[15] This concept applies to a system that consists of a large number of similar components and the components interact according to local rules that are the same for each pairwise interaction. Self-organization also produces emergent characteristics. Under certain rules, the components fall together into structures that are stable and exhibit properties that differ from the properties of their components. An anthill is the prototypical example.[16] The intricate social organization of the hill does not follow any master plan but is created by the interactions among the individual ants. The latter interact pairwise in accordance with relatively simple rules which, when followed by thousands of ants, generate complex patterns that we, the observers, recognize as *defending*, *foraging*, *nesting*, *tending the young* and so on, none of which exist at the level of the individual ant.

A fourth flavor of scaling is *level-invariance*. Some patterns are independent of the material constitution of the relevant system, so they recur at multiple system

Cambridge University Press
978-0-521-83568-8 - Deep Learning: How the Mind Overrides Experience
Stellan Ohlsson
Excerpt
More information

levels.[17] Sometimes this is referred to as *self-similarity*; the system looks like itself at each level of scale. On a map, a small tributary to a larger river looks the same as the river: a gradually widening band of water winding its way through the landscape. It is difficult to tell how big a waterway we are looking at without consulting the scale on the map. Famously, a map of the coast of Britain looks much the same regardless of its scale.[18] More abstract examples have been proposed by both natural and social scientists. Evolutionary biologists debate whether natural selection scales across levels.[19] Organisms, species and perhaps even taxa might be units of selection. Scaling in the other direction, some biologists argue that individual genes are subject to natural selection. In economy, the interaction between supply and demand applies to a village souk as well as to the global economy, or so economists claim.[20] In this scaling flavor, the units at system level *N* exhibit some property P such that when multiple units are combined into a larger-scale unit, that unit also exhibits P. Two generations ago, Arthur Koestler anticipated the centrality of level-invariance in contemporary systems theory by proposing that in most hierarchical systems the laws of behavior are the same at each level in the hierarchy.[21]

Most material systems interact with their environments and their trajectories are significantly influenced by events outside their own boundaries. Economists have coined the convenient term *externalities* to refer to events that are not themselves economical in character but that nevertheless have significant economic consequences (droughts, technical inventions, wars, etc.) and the concept is useful outside economics. A famous example of an externality is the meteor that might have slammed into the Earth some 65 million years ago, spelling the doom of the dinosaurs and perhaps thereby giving mammals a chance.[22] The emphasis on sensitivity to externalities in complex systems research is in stark contrast to the strategy of clockwork science to identify material systems that are so decoupled from their environments that their state variables can be expressed as mathematical functions of each other. Table 1.1 summarizes the key properties of complex systems.

The implication of historicity, irreversible, thoroughgoing change, propagation across multiple system levels, emergence and sensitivity to externalities is, in the words of Nobel laureate Ilya Prigogene, that “the laws of physics, as formulated in the traditional way, describe an idealized, stable world that is quite different from the unstable, evolving world in which we live.”[23] This conclusion extends to the core paradigms of clockwork science. Science writer Ivars Peterson summarizes the developments in astronomy: “Long held up as a model of perfection and the symbol of a predictable mechanical universe, the solar system no longer conforms to the image of a precision machine. Chaos and uncertainty have stealthily

TABLE 1.1. *Key properties of complex systems.*

Property	Description
Historicity	No cyclic behavior; unidirectional unfolding from past to future. The past and the future are not mirror images.
Irreversibility	Changes are not reversible. Effects can be undone by further changes, but the system cannot return to a previous state.
Thoroughgoing change	The laws of change are themselves changing. There are no eternal change constants or laws, no fixed building blocks.
Multiple levels	A system must be described in terms of multiple levels of analysis. A property or a change at level *N* may or may not project upward, and determine system properties or changes at level *N+1*.
Multiple modes of projection	Events at level *N* can be related to events at level *N+1* through direct impact, cascading causation or self-organization.
Emergence	The consequences of projections onto higher system levels are not always predictable.
Externalities	Systems are not decoupled from their environments, so a system trajectory can be radically influenced by events that follow other laws and principles than the system itself.

invaded the clockwork."[24] Natural systems are, by and large, unpredictable. Although clockwork science proudly designated successful predictions as the arbiters of scientific controversies, predictions about natural systems outside the laboratory are in fact rare. This insight is not new. In the early 20th century, the philosopher Charles Sanders Peirce wrote, "There is no greater nor more frequent mistake in practical logic than to suppose that things which resemble one another strongly in some respects are any the more likely for that to be alike in others."[25] What is new is that scientists now realize that unpredictability is not the exception but the typical case.

The lack of predictability is not due to a lack of regularities. But the regularities exhibited by complex systems are of a different kind from those that support the predictions of clockwork science. Earthquakes are, unfortunately, not predictable; that is, there is no known technique of deriving a conclusion of the form *there will be an earthquake of magnitude M at time t on such and such a day, with epicenter located at geographic coordinates x and y.*[26] Nevertheless, earthquakes exhibit regularities. For example, their frequency and size are inversely related: There are many small earthquakes but few large ones. This relationship follows a simple and elegant mathematical form. It is a regularity, not in the individual earthquakes, but in their statistical distribution and so provides no basis for predicting the occurrence, location, size or unfolding of