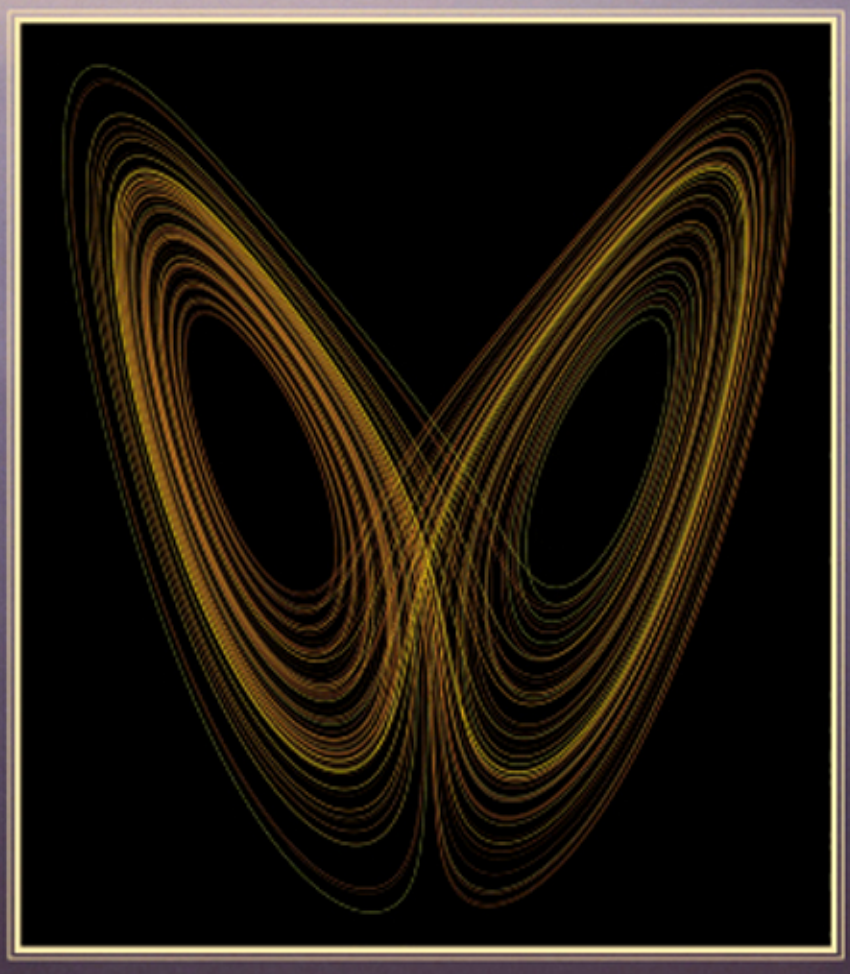
Handbook of
THE PHILOSOPHY OF SCIENCE

General Editors: DOV M. GABBAY, PAUL THAGARD, AND JOHN WOODS

PHILOSOPHY *of* COMPLEX SYSTEMS



Edited by Cliff Hooker

THE COMPLEXITY OF CELL-BIOLOGICAL SYSTEMS

Olaf Wolkenhauer and Allan Muir

PREFACE

The structure of the essay is as follows:

Introduction: The complexity of cell-biological systems has an "inherent" basis, related to the nature of cells (large number and variety of components, non-linear, spatio-temporal interactions, constant modification of the components) and arises also from the means we have for studying cells (technological and methodological limitations).

Section 1: Ontological and epistemological questions are intertwined: to study the nature of living systems, we require modeling and abstraction, which in turns requires assumptions and choices that will influence/constrain what we can know about cells. We study cells, organs and organisms at a particular, chosen *level*, are forced to select subsystems/parts of a larger whole, and pick a limited range of technologies to generate observations. The study of cell biological systems requires a pragmatic form of reductionism and the interpretation of data and models is subsequently linked to a *context*. The framework to develop (mathematical) models is systems theory.

Systems theory is the study of organization *per se*. While investigations into the *structural (material) organization* of molecules and cells have dominated molecular and cell biology to this day, with the emergence of systems biology there is a shift of focus towards an understanding of the *functional organization* of cells and cell populations, i.e., the processes ("laws" and "mechanisms") that determine the cell's or organ's behavior. The necessity to select a level and subsystem, leads inevitably to a conceptual close in the theory of dynamical systems: by classifying observables into dependent, independent and invariant ones (parameters) we draw a boundary between an interior and exterior.

The main challenge is then to assemble a coherent whole from an (partial) understanding of its parts. We argue that this is only possible through an iterative process of modeling, design of experiments, further modeling and so on, in which hypotheses about the whole guide interim models.

Section 2: Applying systems theory to molecular and cell biology, we seek an understanding of structural and functional organization of the subcellular and macroscale level. The cell's functional organization at subcellular level can be

Handbook of the Philosophy of Science. Volume 10: Philosophy of Complex Systems.
Volume editor: Cliff Hooker. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

grouped into three classes of processes: gene expression, metabolism, and cell signaling. This classification involves a range of technologies for each class, leading to an operational division.

Section 3: The preservation of genomic properties through evolution motivates the notion of "model organisms". Unfortunately, even the simplest model organism is very complex and we give examples of the practical considerations involved in understanding so called "pathways". Not only are pathways, cells, organs or organisms complex, the structures of knowing are similarly complex. The concept of a "gene" and the notion of a "pathway" are examples of tools for understanding that develop. Our discussion highlights the importance of discussing how we try to make sense of observations in molecular and cell biology.

Section 4: With hundreds or thousands of components that need to be considered as actors in a pathway/network/subsystem, regardless of how sophisticated the technologies are, our cognitive skills and mathematical tools seem very limited to two- dimensional visualizations and only a handful of system variables. We criticize the suggestion that methods of "artificial intelligence" could avoid thinking of the experimentalist — data do *not* speak for themselves. Mathematical modeling remains an art that (fortunately) cannot be automated.

Section 5: In systems theory objects and relations between objects have identical ontological status. We emphasized above the focus on the cell's behavior (functionality) as a consequence of spatio-temporal interactions of molecules. At any level of a organism, its subsystems are interacting objects whose relationships and properties are largely determined by their function in the whole. While we can study a liver cell in isolation to investigate its stimulus-response behavior, we will only understand the cell's function fully by considering the cell and its environment as an undivided whole. The whole-part relationship emerges as a major stumbling block in dealing with the complexity of cell biological systems.

Section 6: The cells key functions include growth, proliferation, differentiation and apoptosis. Mathematical modeling of any of these processes seeks simplifications to reduce their behavior to its essence, to extract a principle that serves as an explanation. We argue that mathematical modeling is the art of making appropriate assumptions, balancing necessary reductions/approximations due to experimental/methodological limitations with abstractions serving explanatory purposes. This is true for any level (atoms, molecules, cells, orgas, and organisms) and since at any level there is another level above or below, every model is rendered macroscopic or phenomenological. While physics-style mechanial models of interacting mass points are not meaningful in systems biology, the attribute "phenomenological" does not imply arbitrariness in the construction of these models and their explanatory power — to paraphrase G. E. Box: all models are wrong, some are useful.

Section 7: In previous sections we highlighted the fact that living systems can be investigated at different levels, but also processes at subcellular and macroscale can be understood in terms of organizational levels (e.g. gene expression, metabolic networks and signal transduction pathways). The concept of a *domain of auton-*

omy for different levels, suggests a systems-theoretic framework to identify levels of bounded autonomy as subsystems that can be studied in relative isolation, while preserving a chance to understand the larger whole from knowledge about domains of autonomy. A rudimentary body of theory exists and we believe further research into such theoretical concepts and their application in systems biology could lead to practical tools, taking us a small step further in an attempt to resolve the tight whole-part relationship discussed in previous sections.

Section 8: While previous sections focused on dealing with complexity, the necessary reduction to subsystems will introduce uncertainty. The isolated view of subsystems, the necessity of ignoring observables, the inability to keep external variables constant in an experimental set-up, motivate stochastic model formalisms to capture uncertainty in form of *stochasticity*. While this form of stochasticity emerges from epistemological considerations, evolution is an example of purposeful randomness (required to generate alternatives/variations). We briefly discuss the semantics of deterministic vs. stochastic models.

Section 9 concludes our discussion with a summary of key points and an outlook on the field of systems biology. The tight whole-part relationship and the fact that ontological aspects of molecular and cell-biological systems are intertwined with epistemological questions lets us conclude that philosophers of science could actively contribute to the developments of the life sciences. There is a long history of dynamical and mathematical systems theory during which concepts of self-organization, emergence, feedback or system identification have been developed. Studying the difference between physical and biological systems, between living and non-living systems and studying the means by which we have investigated such systems, could improve our chances of managing the complexity of cell-biological systems. In the words of Ludwig Wittgenstein: "The fact that we can describe the motions of the world using Newtonian mechanics tells us nothing about the world. The fact that we do, does tell us something about the world."

INTRODUCTION

Cells are basic building blocks of living systems. Whether one considers a single cell, a colony of bacterial cells or populations of mammalian cells that form tissue, organs and whole organisms, the attribute "complex" is appropriate for any of these systems. An initial, intuitive analysis identifies for the complexity of living systems the following sources:

- Cells are composed of a very large number and variety of components interacting in space and time.
- Cell-biological systems are difficult to observe.
- The dynamic functioning of cells is of a nonlinear nature.
- Living systems are subject to continuous change.

The biologist Ernst Mayr [2004] argued that it is owing to their complexity, that biological systems have the capacity to reproduce, replicate, grow, adapt and evolve: new biological properties can emerge from others. The process that exemplifies the dynamic nature of cell-biological systems is the cell cycle. The *cell cycle* is the series of events leading to the cell's replication. Initially the cell grows, accumulating nutrients and duplicating its components needed for "mitosis", the phase during which the cell is duplicating its DNA. Finally, the cell splits itself into two distinct "daughter" cells. The cell-division cycle is a vital process that underlies the development of an organism and the maintenance of tissue. The cell cycle is a process subject to tight control, which includes the detection and repair of genetic damage, and provision of various checks to prevent uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle [Morgan, 2006].

A striking feature of living systems (organisms) is that their parts interact, modify and create themselves so as to realize an autonomous self-fabricated, self-organized whole. In living systems nothing remains constant; everything is in a perpetual state of transformation; everything comes from other things and gives rise to other things. This is true for the macromolecules making up a cell, as well as for the cells that form organs, which in turn make up whole organisms. A feature of living systems is that "the whole is in the parts", that is, each cell contains a copy of the genetic information for the whole organism. Related is the notion of *emergence*, which asserts that the whole is also more than the (logical) sum of its parts: The system as a whole displays a behavior that could not be predicted from studying its subsystems. Taken together, this provides the basis for a tight whole-part relationship, which is observed at each level of the system — molecules, organelles, cells, organs, organisms. As we shall discuss in detail below, the complexity of cell-biological systems, forces us to study subsystems/parts, raising the question of what we can then learn about the system as a whole?

The term "self" in "self-organization" suggests a form of closure of the system to outside influences, a kind of autonomy. Physically a cell is an open system which relies on a constant exchange of energy, matter and information with its environment. The closure of a living system is therefore not with respect to material causes but with respect to *efficient causation*:[1] In a living system each part/process is at once cause and effect, a means and an end — the cell is a self-organizing biochemical system that fabricates itself [Wolkenhauer, *et al.*, 2007]. Living systems are thus self-referential, every part owes its existence/explanation to the organization of the remaining parts. Through cyclic self-organizing and self-maintaining processes, parts generate the whole as does the whole define the context in which its parts function. The principle of autonomous self-fabrication is a, if not *the*, fundamental property that distinguishes living from non-living

[1]Following Aristotles classification of causes, the material cause of a phenomenon is related to the matter of what something is made of, efficient causation refers to the source of change or process that produces something.

systems. "Autopoiesis"[2] is a terminological framework to discuss the nature of living systems.[3] The ideas are due to Humberto R. Maturana and Francisco J. Varela [1987]. The boundary of an autopoietic system (e.g. the membrane in the case of cells), that is the space in which its components exist, is actively produced by the network of processes that define the system. While acknowledging the importance of self-organization/self-fabrication in living systems, we shall not discuss this theme in greater detail. Instead we refer to an early discussion of these concepts by Immanuel Kant [1892]:[4]

> "In such a natural product as this every part is thought of owing its presence to the agency of all the remaining parts, and also existing for the sake of the others and of the whole, that is an instrument, or organ. But this is not enough — for it might be an instrument of art. ... On the contrary the part must be an organ producing the other parts — each, consequently, reciprocally producing the others. No instrument of art can answer to this description, but only the instrument of that nature from whose resources the materials of every instrument are drawn — even the materials for instruments of art. Only under these conditions and upon these terms can such a product be an organized and self-organizing being, and, as such, be called a natural end."

We define the *structural organization* of a system as the configuration that relates material components of the system. Structural changes of the cell do not necessarily imply changes to the *functional organization*, that is, the processes that determine the cell's dynamic behavior, it's functioning. For example, the three-dimensional structure of a molecule indicates binding partners but does not define its "function". The function of a molecule is understood as its role in processes: changes in the population/concentration of a protein within a region of the cell underpin a process related to a cell function, i.e., growth, proliferation, differentiation (the process by which cells transform into specialized cell types) or apoptosis ("programmed" cell death). The behavior of such dynamical systems, realized through networks of biochemical reactions, would subsequently be analyzed with respect to the system's stability, responsiveness, robustness and sensitivity with respect to change in parameters. The function of a protein in such a network may thus be a contribution that leads to the network functioning as a "switch", "amplifier", "oscillator" or "filter". The theory of dynamical systems provides the conceptual framework to investigate the system's behavior as a result of feedback mechanisms [Tyson, *et al.*, 2003; Novák, *et al.*, 2008].

Studying biological cells we require technologies to observe their behavior and methodologies to construct models. Due to the complexity of cells, existing tech-

[2]The term *autopoiesis* is constructed from *autos*=self and *poiesis*=generation or production.

[3]We note that the emergence of life itself and its chemical origins are important questions relevant in this context. We refer to Luisi [2006] for a recent discussion of this issue.

[4]The quote of Kant ([1892], Section *Critique of Telelogical Judgement*, §65) can also be found in Ernst Cassirer [1981, p. 336].

nologies to generate quantitative measurements are restricted to a particular context in which they can be interpreted. Similarly, we lack suitable methodologies to analyze large scale nonlinear dynamical systems. There is no comprehensive approach to study the functioning of living systems at all levels (molecules, organelles, cells, organs, organisms). Therefore, by considering the multitude of technologies by which we generate measurements and the range of methodologies we employ towards an explanation of observed phenomena, organisms appear to be characterized by an unlimited set of qualities. The notion of a *qualitative infinity of nature* is due to the physicist David Bohm [1957], who noted that at any level at which we study a material system, the fundamental qualities and properties defining the modes of being of the system are limited in number. On the other hand, for every level there always appears to be another lower (respectively higher) level of description; such that the richness of properties and qualities apparently never approaches exhaustion.

Having settled on a particular level at which one investigates a living system (say the "cell level", studying spatio-temporal changes in protein concentrations), the large number of components forces us to decompose a larger whole into smaller, tractable parts for which we construct models. The process, by which a model is established, relies on assumptions and approximations. These may be for mathematical convenience but there is also the fact that in modeling we put a high value on simplicity: the aim of modeling is a reduction of complexity, an abstraction to extract essential properties of a system in a compact form, so as to formulate *generic principles*,[5] laws or mechanisms.

Taken together, the decision for a particular level, the limited range of technologies, the decomposition into subsystems, and the necessary simplifications in modeling a system, we find that the question of how a system functions and the process by which we describe the system are intertwined: what we can know about the system depends on the availability and choice of technologies and methodologies with which we probe the system. Adopting the words of the physicist Werner Heisenberg: "What we can observe is not nature itself, but nature exposed to our method of questioning".[6]

1 SOME ELEMENTS OF SYSTEMS THEORY

As a consequence of the complexity of biological systems, full understanding of cells and their function(ing) cannot be assured. Instead, simplified hypotheses must be formulated and tested by experiments. This requires a conceptual frame-

[5]The word "law" suggests universality of the principles discovered. A difference between physics, where one seeks "natural laws" and biology is that in biology everything seems to make sense only in a context (defined by the chosen organism, cell type, experimental set up, level etc.). Any system is composed of subsystems and is located in a super-system, leaving modeling to be an endless and always provisional exercise.

[6]In Werner Heisenberg: *Physics and Philosophy: The Revolution in Modern Science* [1958] Lectures delivered at University of St. Andrews, Scotland, Winter 1955-56.

work appropriate for making precise and empirically testable predictions. Such a framework is provided by (dynamical) systems theory. A central theme of this text is thus the role of modeling, as a means to formulate and test hypotheses. A mathematical *model* is a representation, a simplified version of the part of the biological system studied, one in which exact calculations and deductions are possible. An obvious priority in modeling is assurance that the model's behavior (established through numerical simulation or formal analysis) corresponds closely to the empirical behavior of the biological system, that the formal system (read "mathematical model") in some way resembles the behavior of the natural system. In addition to replication/reproduction of certain observed qualities or behavior, simplicity and mathematical tractability can be important criteria in developing a model.

A natural system (e.g. a cell or organism) is our interpretation of observable facts in the light of a formal system that we ourselves invent/construct. Understanding a complex system requires abstraction, reducing one type of reality to another. Mathematical modeling facilitates understanding through abstraction. If we are to describe the mechanisms/principles/laws by which the components of a system interact (and thereby realize the (sub)system functions), then the purpose of the model is to distill something complex to a simpler, essential aspect. Modeling does therefore imply for most cases a reduction of complexity; a model is then understood as an excerpt or selection from the biological system under consideration.

Abstraction can reveal truths, but never the complete truth. The growing understanding it affords will push us to the boundary of its validity, where it will eventually mislead if it is not continually re-examined. No abstraction is fully forced upon us, however evidently it appears to suggest itself. It always involves choice, those choices resulting from our standpoint — the position from which, and the purposes for which, we view the system — the level at which we are equipped to examine the system and how extensive are our powers so to do. Diverse approaches, by researchers with different expertise and interests, result in different abstractions which, taken together and made coherent with each other to form a more comprehensive abstraction, can enrich our understanding of the totality.

Fundamental philosophical issues present themselves quite concretely within a scientific discipline: ontology describes basic assumptions about the specific object of study; epistemology concerns itself with the tools we have for obtaining knowledge of that object. Maybe ontology is merely self-deluding epistemology. However the basic materialist presumption of science is that epistemology follows ontology: there is 'what is so' and then there is the extent to which we *know* 'what is so'. The separation, or better the "inseparability", between what the things are "in themselves" and how we represent observed phenomena, will haunt us throughout the essay. The mingling of ontological and epistemological problems tells us how important it is to reflect upon the modeling process itself, the diversity of approaches by which we make sense of observations and limitations these may

pose on what we can know about the natural systems under consideration. As shall hopefully become clear towards the end of this essay, scientists in the life sciences (biotechnology, biomedicine, genomics, molecular and cell biology, systems biology) should or could benefit from interactions with philosophers of science.

A basic ontological assumption of all life sciences is that biological entities are self-regulating through closed causal loops — through *feedback*, that is. The appropriate language for discussing such matters is that developed in *systems theory* [Klir, 1991]. Systems theory is the study of organization *per se*, a general system being understood as a set of interrelated objects, organization being the form of interdependence of objects. For some authors systems theory, and as a consequence systems biology, is essentially the study of organization through *mathematical* analysis.[7] In this essay we shall treat systems theory as a branch of mathematics. An exposition of a rather general mathematical setting was given by Mesarovic and Takahara [1970; Mesarovic, *et al.*, 1975]. A more specialized framework is the theory of dynamical systems [Katok, 1995; Wiggins, 2003]. For a discussion of self-organization in physico-chemical (nonlinear, nonequilibrium) systems through nonlinear systems theory we refer to the book by Gregorie Ilya Nicolis and Ilya Gregorie Prigogine [1989]. Despite its long history, the theory of nonlinear dynamical systems continues to provide various challenges for practical applications in systems biology. In contrast, the study of linear dynamical systems has found numerous applications in the physical and engineering sciences [Kalman, *et al.*, 1969; Padulo, *et al.*, 1974].

A basic assumption in systems theory is that the natural system under consideration can exist in a set of distinct *states*. The experimentalist is probing the behavior of the system through stimulating it with *inputs* to produce observable responses, or *outputs*. It is presumed that the stimuli employed have some similarity with the conditions which the system experiences in its usual context. Using these *observables* the experimentalist may or may not be able to determine which state the system is in; in principle there may be an infinite number of observables necessary for an exact representation.[8] The degree to which he can be certain about the state depends on his choice of observables selected for consideration in the model, for which measurements are possible. This involves the resolution of the measurement devices and the design of experiments, providing suitable stimuli from which the system might be identified. So an observable in a mathematical model is formulated as a mapping that relates states of the system with numbers; the collection of observables which have been chosen is then a vector function of the state.

The *identification* of a system consists not just of finding the state but in determining how that state changes over time, as a function of the history of its

[7] With the (re-)emergence of systems biology as an active and supported research field there is a natural tendency of researchers to adopt definitions of systems biology that suit their interest. We here support the view that systems biology is a new paradigm, complementary but also distinct to activities in the fields of genomics and bioinformatics [Wolkenhauer and Mesarovic, 2005; Wolkenhauer, 2007].

[8] See also the ,,infinite number of qualtities of living systems" discussed by David Bohm [1957].

previous states and the inputs to which it has been subjected. This is reflected in similar functional relations between the observables over time which may be classified into

- Observables whose values remain fixed for every state, referred to as *parameters* of the model — of course, these features which are "fixed" for the given experimental context may well vary if that context is altered.[9]
- Observables that are determined by other observables and inputs to the system.

The definition of inputs, outputs and parameters, draws a boundary, separating the system from its environment. We shall discuss this conceptual closure further below. The choices of observables and the decision about their type also define the context in which the model is valid.

From these general and abstract considerations to the application of systems theory in molecular and cell biology one has to accept further assumptions. Most important for practical applications is the assumption that neither the relationship between observables, nor the parameter values change with time. The next step for a formal analysis and simulation of a dynamical system is that the set of abstract states is replaced by a *state space* with suitable mathematical properties. The mathematical model of the system subsequently encodes relationships between state variables, for which difference or differential equations are most commonly chosen. More radically, the notion of a state in such mathematical modeling can be considered as a secondary concept, being just the functional relation between stimulus and response. This is a purely external account, a severe ontological denial which refuses independent status to states of the system which now crucially depend on the observer (modeler) and his choices of experimental methods. The state of the system (model) is then an encoding of the past behavior of the system, sufficiently informative to form (together with knowledge of the current input value) the basis for the prediction of the output of the system.

Given that numerous assumptions, approximations and simplifications that are necessary for the use of mathematical modeling and computer simulations in practical applications, the phrase "the unreasonable effectiveness of mathematics in the natural sciences"[10] has been coined. Ludwig von Bertalanffy [1969] whose work laid the foundations for theoretical biology and systems biology wrote: "Considering the inconceivable complexity of processes even in a simple cell, it is little short of a miracle that the simplest possible model — namely, a linear equation between

[9]Experimentalists refer to variables such as temperature, pH etc., which they can manipulate in the experiment as ‚parameters'. In modelling biochemical reaction networks, environmental variables like temperature and pH are frequently assumed constant leading to constant values of rate coefficients. In a general mathematical model these rate coefficients are refered to as ‚parameters'.

[10]See, for example, Eugene Wigner's "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," in *Communications in Pure and Applied Mathematics*, Vol. 13, No. I (February 1960).

two variables - actually applies in quite a general number of cases." In the light of recent suggestions for computer modeling projects that should lead to virtual cells, virtual organs and even an entire virtual human,[11] one should not forget the primary role of mathematical modeling in the understanding of complex systems: a reduction of complexity, that is, a simplification through abstraction. Abstraction, while unpopular in its mathematical form amongst experimentalists, serves a practical purpose — the reduction of complex relationships to their essence.

The modeling of cell biological systems with differential equations and the treatment of cells or subcellular processes as physical systems has been criticized, notably by Robert Rosen [1985; 1991; 2000] — see also George Kampis [2003]. Rosen [2000] defines "A system is simple if all its models are simulable. A system that is not simple, and that accordingly must have a nonsimulable model, is complex." Part of Rosen's work is dedicated to show that living systems are complex in the sense that they are not Turing computable. Turing-computability encompasses the class of recursive functions and the formalism of state-based Newtonian physics is just such a recursive formalism. What this entails is that state-based Newtonian physics applies within the realm of Turing-computability, only adequate for modeling simple systems; and conversely, are inadequate for modeling complex systems [Rosen, 1991].

A system theoretic framework formalizes the idea that everything exists only in relation to something else. Causation understood as the principle of explanation of change is thus treated as a relation, not between things, but between changes of states of the system under consideration. Life is considered a relation among molecules/cells and not a property of any molecule/cell. David Bohm [1957] writes: "In nature nothing remains constant. Everything is in a perpetual state of transformation, motion and change. [..] everything comes from other things and gives rise to other things. [A]s we study processes taking place under a wide range of conditions, we discover that inside of all of the complexity of change and transformation there are relationships that remain effectively constant. [..] The necessary relationships between objects, events, conditions, or other things at a given time and those at later times are then termed causal laws."

A central idea of systems theory is that the study of any system requires consideration both of its interior, the subsystems of which it is composed, and of its exterior, the context in which it normally operates as a component of a larger system. The wholeness of a system, its self-maintaining capability against external changes within tolerable ranges, is achieved through interior causal adjustments of its component parts. Thus the relation between part and whole manifests itself dualistically, through seeing the object of study as a whole composed of parts, and seeing it as being itself part of some larger whole and so on. Logically, such a view

[11] See the news brief ,,Systems biologists hatch plan for virtual human", *Nature* Vol. 451, 879, 20 February 2008. The so called ,,Tokyo Declaration" states that ,,Recent advances in Systems Biology indicate that the time is now ripe to initiate a grand challenge project to create over the next thirty years a comprehensive, molecules-based, multi-scale, computational model of the human ('the virtual human'), capable of simulating and predicting, with a reasonable degree of accuracy, the consequences of most of the perturbations that are relevant to healthcare."

implies a tower of systems seemingly unending in both directions; we will return in Section 7 to some implications of this multi-leveledness for biological systems.[12]

In order not to grapple with the whole of this potentially infinite reality at once, some conceptual closure will be an inevitable feature of any model. Modeling, therefore, will always impoverish reality. It can proceed in two ways: either collapsing the exterior into a simple characterization, when attention is focused on the internal mechanisms operating between its component parts; or black-boxing the interior, when the intervention of the exterior is the principal concern. In any model both of these simplifications will be present, allowing us to focus on the dynamics of selected variables with the rest of the totality, the elements excluded from the dynamics, appearing as phenomenologically described contingencies. Enlarging the system to account for the sources of these contingencies incorporates them into a larger model, but again there will be unexplained elements arising from the inside or outside of this larger system. A consistent adoption of the systems approach will alert us to the actual openness of all systems, as opposed to the necessarily closed models we employ [Muir, 1982].

There is an awkward confusion of terminology in this area. We will be referring throughout by the word "reduction" to the necessary selection which occurs when forming any model. This is to be distinguished from the philosophical stance of "reductionism" which insists that higher level variables be expressible only in terms of lower level ones, without the importation of emergent factors — factors which cannot be grounded in properties of the components. This is usually contrasted with "holism" which is dismissed as resting upon such emergence; we would suggest this term be used merely to remind us that each part of a system is constrained by the context of the whole. We will also allow ourselves use of the term "emergent", but only to express the possibility that new behaviors may appear when the scope of a model is extended to take fuller account of the exterior.

The reductionism/holism controversy becomes a non-issue whenever adequate conditions permitting closure of a system (model) are formulated. Indeed, the interrelation of interior and exterior of a system is reflected in the need to consider, in any model, how reductionist and holistic descriptions are related. When confronted with the problem of giving a reductionist explanation of a system's behavior, one is inevitably guided by one's knowledge of what one is trying to explain; one cannot expect, merely by sufficiently understanding the parts, to be able to assemble them to a coherent whole without, as in doing a jig-saw puzzle, some guidance from the overall picture.

[12]Due to the space limitations for an essay we will only be able to sketch personal perspectives derived from the literature and our own work. This will unfortunately neglect a vast number of relevant publications. With regard to theories of living systems we point towards James Grier Miller's magnum opus "Living Systems" [1978] and to the theory of Memory Evolutive Systems, rooted in category theory, beautifully presented in Jean-Paul Ehresmann and Andree C. Vanbremeersch book "Memory Evolutive Systems: Hierarchy, Emergence, Cognition" [2007].

2 THE CELL AND ITS COMPLEXITY

Research programmes in biology focussing on cells generally make an implicit assumption that this will allow us to draw conclusions about higher levels of the organisation (tissue, organs etc.). Our outline of systems theory suggests the formulation of two central questions for cell biology:[13]

> *How do the components within cells interact, so as to bring about the cell's structure and realize its functioning? (The cell's interior aspect)*
>
> *How do cells interact, so as to develop and maintain higher levels of structural and functional organization? (The cell's exterior aspect)*

In the following we assume that the functioning of the cell can be roughly divided into three major classes of processes [Alberts, *et al.*, 2008]: *Metabolism* describes those processes that construct and maintain the cell; processes that realize cell growth and the duplication of the genome before cell division. Metabolism is usually related to the energy household of the cell, divided into *catabolism*, yielding energy, and *anabolism* to describe processes that use this energy to construct the components of the cell (including proteins, organelles etc.) [Fell, 1997; Cornish-Bowden, 2004]. *Cell signaling* subsumes processes of inter- and intra-cell communication and the coordination of cell function [Kholodenko, 2006]. While cell signaling is realized through the generation, modification, degradation and translocation of molecules, the primary focus of research in this field is the explanation of signal transduction, information transfer, cellular decision making and "higher-level" coordination of basic cellular processes. *Gene expression and regulation* is here defined as the process by which information, encoded in the DNA, is transcribed and translated into a *gene product* (e.g. a protein).

A system's complexity arises from three factors: the quantity, the variety and the interconnectivity of its constituent elements. An essay by Warren Weaver [Weaver, 1948] is widely considered a founding text for thinking about complexity. Weaver distinguished between *disorganized complexity* and *organized complexity.* Disorganized complexity results merely from the quantitative aspect — having a very large number of similar parts. The interactions of the parts are perceived as "largely random" suggesting methods from statistical mechanics and probability theory to understand properties of the system as a whole. Gas molecules floating around in a container are a classical example of such disorganized complexity, where one is not interested in, or able to, trace/describe the trajectory of each material component using Newton's law of motion. Models of such systems are defined in terms of distributions, and predictions are usually expressed in terms of their mean values or standard deviations. (We return to questions of randomness in Section 8). Problems of *organized complexity* on the other hand are related to systems with properties that arise from non-similarity of a variety of parts from which it is composed and in which, moreover, the organization of the interacting

[13] cf. Ricard Solé and Brian Goodwin [2000].

parts cannot be derived from a study of the parts in isolation. Moreover, this internal complexity can be matched by a corresponding external complexity of the supersystem of which it may be a part.

While live cell imaging, tracing organelles or molecules in the cell, gives the impression of disorganized complexity, it is obviously a case of organized complexity we are dealing with in systems biology. Studying the functioning of cells, the experimentalists face various practical hurdles, discussed in the following section.

3 EXPERIMENTAL METHODOLOGY

The analysis of a system includes the design of experiments to generate data, the construction of models and the possibility for predictions about the system from a model.

Complexity, of itself, throws up immense challenges to experimental methodology: selection of some overall function of a system as a suitable object for study and identification of just a few relevant features from the immense variety of components present. In biological systems experimental difficulties arise not just from the three aspects of complexity but also from problems of size — technical issues of visibility, monitoring and control.

For example, *Escherichia coli* is one of many species of bacteria living in gut flora of mammals and which measures only two micrometer in length and a cell volume of 10^{-15} Litre, containing an estimated number of 2,600,000 proteins, generated from about 4252 protein coding genes.[14] In biology "Seeing is understanding", but making cellular processes measurable (visible) is an obvious technological challenge.

A *model organism* is a species (e.g. yeast, bacterial systems, worms, fish or flies) that is extensively studied to understand particular biological phenomena, with the expectation that discoveries made in the model organism will provide insights into the workings of other organisms. In particular, model organisms are widely used to explore potential causes and treatments for human disease when human experimentation would be unfeasible or unethical. This strategy is made possible by the common descent of all living organisms and the conservation of metabolic and developmental pathways and genetic information over the course of evolution. Model organisms are often chosen on the basis that they are amenable to experimental manipulation. This usually will include characteristics such as short life-cycle, techniques for genetic manipulation and non-specialist living requirements. The complexity of human cells leads then to a situation in which *E. coli*, a bacterial cell living in the lower intestines, is used as a model to study intracellular processes occurring in mammalian cells.

The threefold division of cellular processes into metabolism, signaling and gene expression associates with each a range of specialized technologies for generat-

[14] See `http://redpoll.pharmacy.ualberta.ca/CCDB/` for the *E.coli CyberCell* Database (Information accessed November 2007).

ing experimental data. The nature of the data can differ considerably, making their integration a challenge. At the methodological level, where one is trying to model and simulate these processes a range of approaches are used, depending on what type of pathway or network one is looking at. The biochemical reactions of metabolism are organized into *metabolic pathways*, while reaction networks to do with signaling are organized into *signal transduction pathways*. Genes are sometimes regarded as nodes in a *gene regulatory network*, with inputs being proteins such as transcription factors, and outputs being the level of gene expression.

The study of metabolism, cell signaling and gene expression requires a range of technologies, often leading to an operational division of researchers into "Omics" disciplines, including "metabolomics", "proteomics" and "transcriptomics". While there is an obvious relation between metabolism, signaling and gene expression, the complexity of the cell, specifically the technological difficulties of measuring these processes has forced researchers to specialize with obvious consequences for the overall endeavor – we can't see the wood for the trees. Understanding neurodegenerative diseases or cancer requires the integration of knowledge and models of metabolism, signaling and gene expression.

To model inter and intracellular processes one requires quantitative spatio-temporal data for a relatively large number of components. At present these are not available, forcing us to handle uncertainty and "reduce" complexity. For practical purposes to do with technological limitations, but also with the time and money required to conduct the experiments, a subset of components is chosen. This leads to the pragmatic notion of *pathways* or *networks* as a selected subsystem of biochemical reactions (relevant to some cell function). For example, the Mitogen-activated protein (MAP) kinase signaling pathway is a system that responds to extracellular stimuli (mitogens) and is linked to various cellular activities, such as gene expression, mitosis, differentiation, and cell survival/apoptosis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[15] includes about 40 proteins in a graphical representation of one variant of this particular pathway. For most experiments one will, at the present time, only be able to focus on less than 10 proteins — which ones?

One criterion for the identification/separation of subsystems is based on different time scales. For metabolic systems where processes are assumed to be at steady state on the time scale of changes in metabolite pools, one can treat sets of enzymes as modules with a defined input and output. This approach is however not applicable to cell signaling systems where the transient behavior of the fast system (transfer of the signal) has consequences for the behavior of the slow system (subsequent gene expression), which then feeds back to the faster system after a delay. Examples include systems where the signal may be encoded in oscillations (e.g. Ca signaling) or transient movements between cellular compartments (e.g. NF-κB). Epidermal growth factor stimulation of the MAPK signaling cascade causes transients on time scales of minutes and relaxes to a new quasi-steady state within an hour but subsequent consequences can take much longer to emerge;

[15] See `http://www.genome.jp/kegg/` (Information accessed November 2007).

entry of cells into the cell cycle and commitment to cell division requires several hours sustained signaling, whilst receptor internalization and recycling and gene expression alter the concentrations of the components in the signaling system also on time scales of hours. To this day, most pathways are studied in isolation, while there are always many pathways that are relevant to any one cell function. The recent concept of "cross talk", more than anything else, is a proof of failure of our initial attempts to define a subsystem.

The question of how to identify subsystems (modules, pathways etc.) as functional subunits which possess, to some degree, bounded autonomy, and how one could subsequently integrate the knowledge and models achieved into a larger whole, are the two most important challenges for systems-theoretic research. A key problem is that in a complex system the whole is more than the sum of its isolated parts. In other words the interaction of subsystems can lead to emergent behavior irreducible to the system's constituent parts considered separately. Emergent behaviors occur through *interconnectivity* – intricate causal relations across different scales and feedback.

The difficulties listed above force the experimentalist/biologist to collaborate with technologists and modelers. While nowadays there is no doubt that interdisciplinary collaborations are necessary for advances in the life sciences, it is fair to say that most scientists would prefer to rely only on their own skills. While there may be some true "hybrids" with an equally good training in biology and mathematical modeling, the complexity of cell-biological systems requires specialization (in wet-lab experimentation, the development of technologies, data analysis, mathematical modeling and simulation). The need to combine different expertise, possibly across departments, universities, countries and cultures is a complex undertaking. People might study a particular MAPK cascade because their supervisor did the same, or because of what piece of equipment they have access to, what cell lines are available, etc. — the choice is made for historical and social reasons rather than scientific ones.

Finding the forest among the trees is a major challenge for today's life sciences as no individual, no single research group or institute can provide all of the necessary expertise, technologies and experimental systems. Studying living systems requires an interdisciplinary approach. The complexity of cells makes it necessary that highly specialized experts from different disciplines communicate and this often across long physical distances, countries and cultures. So we can add to our earlier list a further group of problems for the experimentalist, which arise from the social context, *interdisciplinarity:* specialisation, education, distances between institutes.

If epistemology is to follow ontology in the field of systems biology, the structures of knowing should mimic the structures of existence. So the structures of knowing will similarly be complex systems — inter-disciplinary teams exchanging information on what is known within the specialized spheres of each participant. This raises an intriguing philosophical question of the sense in which a group of people can be said to "understand" something. Must there always be an individ-

ual mind whose role is to be the overall understander — perhaps a convenor or chairperson who coordinates the separate understandings into a coherent whole, without necessarily knowing fully the details of any part. And what, anyway, is *understanding*: we will encounter later a related question of whether a computer program can deliver understanding by due processing of complex data.

Communication between participants in a collective scientific enterprise will be mirrored in complex knowledge structures, primarily now on the internet, reflecting the separate disciplines and how they interconnect. So the interrelated objects of a system may be material or informational. This leads us to a distinction between *natural systems* and *formal systems* [Rosen, 1985]. A natural system is a selected portion of the external world which we investigate through experiments and model through physico-chemical dynamics. A formal system is based on symbols, syntax and transformational rules of symbol manipulation. The *modeling relation* [Rosen, 1991] describes the process by which we establish congruence between the two systems; allowing us to study the formal system as a *model* of the natural system. Genomics has a peculiar status as a *natural information system*, where the dynamic modeling at the molecular level gives rise to rules taking the form of informational transformations.

4 DATA HANDLING

Studying complex systems one is forced into practical forms of reductionism. Our brains have evolved to cope with a relatively small number of pieces of information, dealing with processes taking place in the time scales of everyday events and for which linear/proportional relations may apply. Many nonlinearities, very fast or slow dynamics, as well as delayed responses are beyond our intuitive common sense; they surprise us.

Biological cells offer all of these sources of difficulty. However, there are now technologies available to measure the expression levels of thousands of genes simultaneously. The resultant gene expression data is initially stored in the form of a data matrix; the number of rows n, usually denoting genes, takes values up to 30,000, the number of columns m would denote either a few time points at which measurements were taken or a couple of different conditions. How does one recognize pattern in thousands of gene expression profiles? Plotting the raw data simultaneously and unordered would fill any computer screen or sheet of paper with thousands of lines, which almost certainly would not allow a detection of pattern by eyesight. Any analysis of high-dimensional data therefore requires a dramatic reduction/projection into lower dimensions allowing visualization and pre-grouping, in order that its structure may be grasped, in some intuitive sense. This is a necessary prerequisite for forming hypotheses leading to models. While 3D visualizations of data are possible on computer screens, this approach is largely for illustrative purposes. In practice, an analysis of experimental data is almost certainly conducted through plots in the plane fitting a sheet of paper, a computer screen or white/black-board.

A cynic might say that studying high-dimensional multivariate data does not only imply a form of reductionism into the two dimensions of computer screens and standard-sized sheets of paper, but scientists are also forced to reduce their understanding of complex systems into a 20 minute presentation and into about eight pages of a publication. Regardless of how complex a system is, the communication of research results will, in practice, almost always be limited to a short oral or written presentation that fits the constraints set by journals, conferences but also not exceeding attention span of an interested audience. So called "holistic approaches", so desirable in the life sciences, are at present wishful thinking.

We are led to enquire what tools we possess to aid this reduction. The usual suspects are mathematical modeling, statistics and computer programming. The first of these forms a proper basis for choosing or developing techniques from the other two, which is why in systems biology mathematical modeling has become an essential component: mathematical modeling is the refinement of common sense into the realm of complex systems. Most particularly for the present discussion, it possesses the appropriate language for describing and analyzing high dimensions.

The statistics we use to reduce the raw data to a comprehendible form require an underpinning mathematical analysis of their purpose. For example, we need to take care when interpreting data via the familiar two-dimensional reduction afforded by forming covariances between pairs of variables, since part of the essence of complexity is the interpenetration of many variables acting together. So some essential information will be lost by the very act of proceeding in this way. We are probably obliged to employ such methods, whenever more appropriate techniques are lacking, but we should remain aware of the desirability of their justification in any new circumstances.

Once suitable methods for reduction have been settled on, we have the facilities offered by computer programming to deliver the resultant reduction. Once again, though, it might be dangerous to merely lift tools which have proved efficacious down from the shelf without at least attempting to understand their underpinning justification. The availability of high-throughput and whole-genome technologies, which generate gene expression profiles for thousands of genes, has tempted some researchers to speak of a "holistic" perspective on cellular function. Douglas Kell [2003] contrasts two strategies for understanding cell-biological systems with "omics-data": "The reductionist view would have it that if we can break the system into its component parts and understand them and their interactions in vitro, then we can reconstruct the system physically or intellectually. This might be seen as a 'bottom-up' approach. The holistic (or top-down) approach takes the opposite view, that the complexities and interactions in the intact system mean that we must study the system as a whole." During the heyday of genomics and bioinformatics it became common practice to collect data without preconceptions, doing "some" (any) experiment and then search the data for pattern in what is called "data-driven discovery". It then seemed that no theory, hypothesis or model was required for these "hypothesis-free fishing expeditions".

With all respect to the area of "artificial intelligence", the reason why high-

throughput genomics and bioinformatics have been able to do without a more focused approach is that in the early days of genomics it has been rather easy to discover "something". To continue the metaphor of a fishing expedition, fishing in the North Sea or Atlantic during the 1960's did not require much of a hypothesis about how to find fish; putting out your net almost anywhere, would catch some fish. Nowadays, you need to plan your fishing expeditions more carefully, supported by a good understanding of the fish's behavior. It is naive to believe that "any" experiment would do. A living system, observed at steady state or equilibrium, is either dead or does not reveal the information necessary to understand its behavior. We can only reconstruct/model the mechanisms or interactions that generate the observed behavior of a dynamic system if we systematically perturb/stimulate it and then observe its response. Even the simplest stress response or knock-out experiment implies a hypothesis — if only about the fact that the induced changes alter the behavior.

Some members of the computer science and bioinformatics community, providing boats and tools for these hypothesis-free fishing expeditions, argue that rather than beginning with a hypothesis, "artificial intelligence" approaches would generate hypotheses, extract them from data by letting "data speak for themselves". These approaches would then be considered inductive, inferring a general law or principle from the observation of instances. The philosopher Karl Popper famously argued that induction is a myth. John F. Allen [2001a; 2001b] responded to these genomic and bioinformatics fantasies, and this led to a series of articles documenting this interesting debate [Gillies, 2001; Kelley, Scott, 2001; Smalheiser, 2002]. Allen argues that knowledge cannot arise *de novo* from computer-assisted analysis of biological data. What he comments on is the proposition that analysis of data can enlarge human understanding in the absence of any hypothesis or preconceived idea. "It makes little sense to insist on collecting genomic and structural data before you, or someone else, has posited an underlying mechanism. Without having an underlying mechanism — in essence an explanatory, tentative hypothesis — you have no basis on which to decide which data to collect. Data do not, and cannot, 'speak for themselves'." [Allen, 2001c].

Studying complex systems one has to be able to adapt or refine the goal or hypothesis as you go along. As Allen [2001b] writes: "Computers are necessary to analyze large data sets, but they are not sufficient. [..] Creativity consists of a willingness to consider the relevance of observations that have no apparent connection with the problem as it is viewed conventionally." The computer program can only act via the given algorithm and cannot venture beyond the ominous "background knowledge" which has been fed into the program.

Apart from practical considerations we face methodological challenges. For instance, uncertainty can be seen as a consequence of complexity ("complex systems are difficult to understand") but apart from epistemological aspects of dealing with uncertainty we shall later also consider ontological aspects of randomness in nature.

At present, the questions asked in systems biology are largely determined by

the technologies used to generate data, by the (model) organism chosen, by the choice of a particular cell type and cell line, antibodies available etc. Take for example the question of how the Ras/Raf/MEK/ERK (MAPK) signaling pathway works. While most researchers would readily agree that this is a reasonable and important question, there are several problems with this approach. To begin with this pathway is not a module of bounded autonomy as discussed below. The question should not be how this particular chosen pathway works but how cells grow, differentiate, proliferate or die as this is in fact the process relevant to diseases (cancer research being the primary motivation for studies on MAPK pathways). One should therefore first identify a question or hypothesis about the functioning/behavior of the cell, then identify a suitable model organism, cell type, cell line, to decide upon a subset of components, a pathway and only then identify the technologies adequate to generate the data required.

Assuming there are functional modules and levels of bounded autonomy in cellular systems, how do we best identify their boundaries and constituent components in experiments? Given a selected subsystem, how do we then unravel feedback mechanisms giving rise to the observed dynamical behavior and how do we integrate knowledge and models of subsystems to understand the interconnectivity of organizational levels and explain emergent behavior? Clearly, the study of complex systems is not just about the nature of things (an ontological problem) but this research also raises questions about the way in which we generate knowledge.

5 THE CELL AS A SYSTEM

Let us return to our two central questions of systems biology:

1. *How do the components within cells interact, so as to bring about the cell's structure and realize its functioning?*

2. *How do cells interact, so as to develop and maintain higher levels of structural and functional organization?*

Studying living systems one is bound to confess that the more we learn about them the less we are prepared to generalize. The complexity of cellular systems, the difficulties in studying them in experiments, has led to high levels of specialization within disciplines, hindering the generalization of results. Living systems appear so complex, so diversified that no general statement can safely be made about them. While systems biology has emerged from the need to put the pieces of the puzzle together, it does not offer a holistic salvation in the sense that universal laws ("physics style") can be derived. Throughout the essay we have emphasized the irreducible wholeness of living systems but also the impossibility of a truly holistic approach in which we can study/observe a cell as a whole.

An important aspect of a systems-theoretic approach is that objects and relations between objects have identical ontological status: Life is a relation between molecules/cells and not a property of any molecule or cell. Paraphrasing Henri

Poincaré[16] we might say that a cell is built up of molecules, as a house is with stones but a soup of molecules is no more a cell than a heap of stones is a house. Organisms, their cells, genes, and proteins are complex collections of interacting objects whose relationships and properties are largely determined by their function in the whole. In living systems everything exists only in relation to something else. The cell or any subsystem of it, together with the associated environment, has to be understood as an undividable whole. This almost obvious fact is constantly ignored by the reductionism that is forced upon us by the complexity of cells. To use the language of physics, the cell is a many-body system in which non-local interactions between the constituent molecules exert influence on the locally analyzed components. In other words, the inter-relationships of the parts (sub-wholes) within a system depend crucially on the state of the whole, in a way that is not expressible in terms of the properties of the parts alone. The irreducible wholeness of living systems suggests principle limitations to what we can know about them.[17] We are forced into reduced representations, necessarily leaving things out. The uncertainty arising from reduced or approximate representations could be captured with stochastic models but this would not solve the problem of how to distinguish between an intrinsic feature of the natural system and methodological considerations.

Living systems are dynamic systems, they are constantly changing, almost every part of an organism being exchanged throughout its lifetime. In systems theory this is reflected in the interpretation of causality as the principle of explanation of change: causal entailment is not considered to be a relationship between things (genes, proteins, etc.) but a relationship between changes of states of things. Not only do cells dynamically respond to immediate external changes and stimuli, they are also subject to evolution, not only at a time scale that covers generations of the organism but also in the range of hours, days and weeks. We can distinguish between two dynamic principles of key importance in studying cell function: a system's ability to maintain its current state against external perturbations (e.g. homeostasis) leading to some form of *robustness* and the system's *responsiveness* to environmental cues, to adapt its state or even modify its biophysical make-up. The basis for all forms of regulation, control, adaptation and coordination is the notion of *feedback*. Feedback loops provide the system with information about its current state and possible divergence from a desirable state/trajectory. Based on the current values of system or state variables a change is induced to move the system into a "desirable state" or follow an "intended trajectory". For example, in development stem cells should grow, proliferate and differentiate but this implicitly assumes the existence of an objective. In complex dynamic systems the change of state is influenced not only by inputs to the system but also by an overall goal or

[16] "A collection of facts is no more a science than a heap of stones is a house" Henri Poincaré (Science and Hypothesis, 1908) or, also attributed to Poincaré: "The aim of science is not things in themselves but the relations between things; outside these relations there is no reality knowable."

[17] See George Kampis [2003] for a critique of state-space based (differential equation) models as descriptions of living systems.

objective (i.e. the distance between a current and desirable/reference state): Living systems are *anticipatory* [Rosen, 1985]. The existence of feedback mechanisms also highlights the importance of systematic perturbation studies as only then we will be able to unravel the structure of dynamic networks from stimulus-response data. A system that is self-organized or robust to external perturbations does not reveal its internal functional organization in simple observations but requires an experimental manipulation.

6 SYSTEMS BIOLOGY OF THE CELL

Molecular and cell biology to this day has been preoccupied with the identification and molecular characterization of cellular components, leaving little time to conceptualize biological information and to develop "theories". The recent interest in systems biology is associated with the hope that it will be possible to manage the complexity of cellular systems, leading to postulated generic *principles* that govern those cellular processes that underlie the development and (mal)functioning of cells, cell populations, tissues, organs and organisms.

The (re-)emergence[18] of systems biology over recent years signals a shift of focus from the study of the *structural organization* of cells towards an understanding of the *functional organization* of cells. By structural organization we refer to the physical structure and material basis of cells, including macromolecules (e.g. DNA,[19] enzymes,[20] and receptors[21]), organelles[22] as well as the outer cell wall (and inner membrane in eukaryotes). The functional organization of the cell refers to processes that determine the cell's activity (its dynamic behavior). The word "function" refers to a role defined by the context of a system or process. For example, the role of stem cells can be the regeneration of tissue. This provides the context for *cell differentiation* (a specialization of stem cells, turning them into a specialized cell type). Cell differentiation in turn is the context for various networks in which proteins interact in order to realize this function (or an aspect of it). The most important *cell functions* studied in systems biology include cell growth, cell proliferation, cell differentiation and cell death (apoptosis).

Not only are we forced to select a subset of proteins, respectively a subsystem, even if we could quantify larger number of components, the analytical tools for the

[18]The need for a research field of systems biology was first formulated by Mesarovic [1968].

[19]Deoxyribonucleic acid (DNA) is a macromolecule that encodes the genetic information used in the development and functioning of all known living organisms (virus being a special case). The entirety of hereditary information of an organism is also referred to as the *genome*.

[20]Enzymes are proteins that catalyze (accelerate) biochemical reactions. The vast majority of processes in a biological cell require enzymes to facilitate the modification or transformation of molecules. Inhibitors are molecules that decrease enzyme activity; activators are molecules that increase activity.

[21]In cell signalling a common mechanism for information transfer is the binding of signalling molecules (ligands) to receptor proteins on the outer cell membrane. The binding leads to a biochemical modification which transfers the information through a series of intracellular processes into the nucleus where the information can lead to changes in gene expression.

[22]An organelle is a specialized subunit within the cell.

analysis of such large, nonlinear models are missing. Proteins are modified (e.g. activated), each of these states adding to the number of variables in a mathematical model. A system with 10 components can subsequently lead to 20 or more system variables. The theory of nonlinear dynamic systems, the methodologies and tools available to identify models (their structure and parameter values) from experimental data, to investigate their behavior analytically or through numerical simulations remains to this day limited. We are once more forced to simplify for practical considerations (e.g. through linearization). The reduction of complexity through abstraction and modeling does however not only serve practical purposes. Studying complex systems we seek simplifications to reduce complex processes to an essential aspect of their functional organization, to extract a principle that serves as an explanation. Studying complex systems we are seeking general principles underlying the observations we make in experiments. Mathematical modeling is then the *art* of making "appropriate" assumptions, balancing necessary reductions due to methodological and experimental limitations with abstractions serving explanatory purposes.

The construction of dynamical models is informed by experimental data. In an ideal situation, experimental time course datasets can be used to identify the structure of a network (and hence of the equation that form the model) and parameter values can be directly estimated from time series. At present, there is a lack of technologies that allow us to quantify temporal changes of gene activity and changes in protein concentrations with sufficient accuracy/reproducibility, for a sufficient number of time points and for a larger number of molecules (and their activation states). There are on the other hand technologies that can detect thousands of proteins simultaneously (e.g. 2D gels) or indicate the activity of genes for whole genomes (e.g. microarray or gene chips). Such "Omics" data, coming from high-throughput and whole genome technologies, have been analyzed in the area of bioinformatics using methods from multivariate statistics, "machine learning" or "data mining". Their qualitative character has, so far, prevented the identification of models from dynamical systems theory.

The study of complex systems is difficult, pushing state-of-the-art technologies to their limits and demanding new methodologies to interpret data through modeling. We can distinguish between two complementary general aims for modeling: reproducing complexity for computer simulations and reducing complexity in models that encode general principles. In the first case we try to establish a detailed replica computer representation of a complex system. Typical examples are large-scale mechanical/physical models of engineering systems (say airplanes). Computer simulations would then allow the study of the system's behavior, predicting the behavior of the system under unobserved conditions. For as long as the system is mechanical, subject to Newtonian physics the parameter values for the computer model can be derived from "first-principles" (considering mechanical properties of the constituent components). The second principle aim of modeling is to simplify, reduce complexity to some general principle through abstraction. For cell biological systems we cannot develop microscopic models in which molecules

are treated as mass-points, instead one models changes in molecular concentrations in a macroscopic sense. Since parameter values cannot be derived from "first (physical) principles", one could estimate them from time course data. As discussed above, state-of-the art technologies cannot — at present — deliver suitable datasets, nor is system identification simple for nonlinear spatio-temporal systems. Even if these macroscopic models may be phenomenological, this does not mean that the structure of the equations is arbitrary as in black-box modeling. The structure of the mathematical model encodes in this case a hypothesized principle. We are going to focus our discussion on the second type of models, which cannot be derived from first principles. Whatever the goal of modeling, it will be important to distinguish between the complexity of the natural system under consideration and the complexity of the effort by which we gather information and gain knowledge about the complex system.

7 THE MULTILEVELEDNESS OF CELL-BIOLOGICAL SYSTEMS

The structural (physical/material) organization of cells is the outcome of an elaborate self-organizing process, involving gene expression, regulation, signalling and metabolism. This structural organization of the cell then serves as an environment for the cell's functional organization, leading to growth, differentiation, proliferation and cell death (apoptosis). The cell is itself a component of a larger system with higher levels of structural and functional organization. For the human organism these can be summarized as follows.

At the spatial level of the entire organism the human body grows, reproduces and dies in time scales that can be years. The human body is made up of organs, which help to maintain, renew, repair or regenerate the organism and adapt it to its environment. Organs realize their functions over hours and weeks. Organs are made up of cells, which go through the cell cycle, grow, divide, specialise and die. These processes take place over minutes and hours, while intracellular biochemical reactions take place in seconds.

Multileveledness is a key organizing principle in complex systems where the responsibility for proper functioning of an overall system is shared by the subsystems that constitute the different levels. A fundamental property that is determined by interlevel relations is that the levels have the latitude to focus on their 'allocated' tasks, and which implies that each level must posses a *domain of autonomy* [Mesarovic, *et al.*, 1970]. If two levels each possess a domain of autonomy it means that each level has some range of behavior which is autonomous in the sense that the two levels do not affect each other through changes in these ranges; changes within the domain of autonomy of one level is perceived as "background" by the other level and vice versa. The influence of one level is treated as a bona fide signal by the other level whenever the receiving level is outside of its domain of autonomy.

Making a distinction between the interaction (signaling) and the interdependence of levels is useful in this respect. Although the levels, belonging to the same

system, are interdependent in many ways, they are non-interacting (non-signaling) within their respective domains of autonomy. Identification of the domains of autonomy of pathways is therefore a major challenge for systems biology. Domains of normal behavior are delineated by tolerances. The system can become pathological either when the a function on a level strays outside of the domain of autonomy or when a tolerance on a level changes due to internal or external influences. Bounded autonomy provides *cross-level harmonization* and illustrates that nature's design is not optimization of the behavior over time but rather optimization of a system's organized complexity. A major challenge for systems biology is to develop methodologies and experimental designs that allow the identification of functional modules, the separation of subsystems and levels of bounded autonomy.[23]

The very existence of different levels for experiment and modeling suggests an ontological basis for the complexity reduction which is necessary for our understanding. Our epistemological efforts arise not merely from arbitrary selection of features of a system. The fact that we observe coherence of the parts into higher level structures, of which we can speak with a suitably-tailored language, suggests the ubiquity of automatic stabilization. The procedures we adopt arise from our attention naturally being drawn to objectively real coherent structures which exhibit bounded autonomy.

The kinds of stability encountered in the major part of current systems theory is entirely inappropriate to handle such phenomena. The most elaborated results of this theory have been developed, mainly in engineering contexts, in the framework of linear systems. In such systems the possible forms of stable states are very limited and will not allow the kind of complexity reduction we are seeking, a dramatic decrease in the number of dynamic variables which need to be considered.

Our attention should therefore be directed towards *non-linear* systems theory. Recall that the simplest kind of dynamical system — a finite-dimensional, deterministic, memoryless system — is described mathematically by a collection of first-order differential or difference equations. The collection of dependent variables, say n in number, are quantities characterizing the system's state which develops through time. Our principal concern is with complex systems in which n is large, but for which we can regard as significant only a few functional combinations of these. This may be achievable in the following way.

Suppose each orbit of interest to have a simple attractor — an attractor for an orbit being its asymptotic limit as time tends to infinity. If the time-scale for the dynamical events is short when compared with that employed by an observer of the system, the state will appear to that observer to jump rapidly on to the attractor and stay there. We need only assume that all the usual initial conditions have orbits going to the same attractor and the observer will see an apparent self-organization of the system to a coherent behavior. If the attractor, considered as a subspace of the state space of dynamic variables has low dimension, we can regard the motion within the attractor as a simple description of the system's trajectory in terms of just a few aggregate variables which play the role of coordinates within

[23]See Mesarovic *et al.* [2004] for a discussion of these ideas in systems biology.

that subspace. Even allowing for perturbations away from the attractor, these variables can approximately describe the actual position of the system's state when it moves within a sufficiently close neighborhood of the attractor. One might then hope that the perturbations around the stable values, remaining small in magnitude, may be handled by suitable statistical techniques.

We initially stated the intention to avoid ontological questions about the role of randomness in living systems, instead focusing on epistemological questions, i.e. uncertainty arising from reduced and approximate descriptions. As long as our models capture a level of the functional organization of cells at which randomness does not matter, we are fine. However, models should after all be a representation of what the things are "in themselves", which means that we cannot always ignore the role of randomness. For nonlinear dynamical systems randomness is most potent at bifurcation points and in systems sensitive to initial conditions. How can we then distinguish between intrinsic (possibly purposeful) randomness and a signal that is formed from a temporal average of a molecular concentration? What this problem suggests is the importance of stating the context in which a model is valid.

8 DEALING WITH UNCERTAINTY: RANDOMNESS, STOCHASTICITY

Observing a cell-biological system, irregularities and the absence of an obvious pattern/trend in data induce uncertainty in the analysis of the system. The first question is then whether this *randomness* is an inherent, possibly purposeful aspect of the system or whether it is a consequence of limitations in observing the system (the choice of subsystem looked at, components that are ignored or limitations to measurement technologies)? In either case, one may consider a *stochastic model* to describe the system in terms of probabilities [Ullah, *et al.*, 2007; 2008].

Note that our discussion will be limited to the level of cells, where we investigate the function(ing) of cells in terms of changes in the abundance of molecules within cells and consequences this may have for populations of interrelated cells [Raj, *et al.*, 2008]. The discussion of randomness in physics, specifically statistical mechanics, may thus be avoided in our present context. While thermal and perhaps quantum fluctuations may in fact influence events at the cellular level and above, instead of modeling them in detail we may, without losing essential cellular and higher order modeling power, represent their consequences by irreducible stochasticities. The cell is here considered an open, non-equilibrium system, with a constant flux of material and information in and out of the cell. At the level of single molecules, the irregular motion of atoms and molecular bonds within the system may well be relevant but will here be referred to as effects of the *microscopic level*. This includes thermal fluctuations and Brownian motion. Looking at changes in the concentration of molecules, following a clear trend that may well be described in terms of differential equations, such models may be referred to as *macroscopic*.

Where necessary, a stochastic model can be formulated comprising both the

deterministic laws and the fluctuations about them. Such models are sometimes referred to as *mesoscopic models* [van Kampen, 1992]. Considering a system of interacting mass points, fluctuations in non-equilibrium systems do not arise from a probability distribution of the initial micro-state, but are continuously generated by the equations of motion of the molecules. While mesoscopic stochastic models are attractive theoretical concepts, in a practical context where such a (nonlinear) model and its parameter values would have to be extracted from experimental data, we face various problems (which are in part a reason for the wide use of ordinary differential equations).

We can illustrate the notions of microscopic, mesoscopic and macroscopic in the context of cell biology by considering gene expression, the process by which information of the genome is first transcribed into RNA before being translated into proteins. These two stages involve two levels, the transcription of a gene being microscopic compared to fluctuations in the concentration of the protein for which the gene encodes the information. While for the initiation of transcription, say through the binding of transcription factors, a stochastic model may be appropriate; changes in the concentrations of the proteins involved in the function of a single (e.g. cell cycle) may on the other hand be described macroscopically by ordinary differential equations. Taken together, the whole model is mesoscopic.

In many situations random fluctuations are sufficiently small to be ignored, allowing macroscopic equations to predict the behavior of a system with great accuracy. Cells however are "open systems", where the environment may force them into a stationary non-equilibrium state in which the system's dynamics bifurcate, the direction taken depending on the specific fluctuations that occur. Note that therefore the "randomness" of the fluctuations (which we can only describe in terms of probabilities) influences the behavior of the system of macroscopic equations most critically at specific bifurcation points, while other areas of the state space may be perfectly well approximated by macroscopic equations. Intrinsic noise from thermal fluctuations or transcriptional control could determine how the system at the macroscopic level goes through a bifurcation. Looking at a population of genetically identical cells in a homogenous environment, this leads to variability of cell states that may well be exploited by the biological system [Rao, *et al.*, 2002; Kærn, *et al.*, 2005; Shahrezaei, *et al.*, 2008]. The obvious context in which randomness has a function is generating diversity in evolution.

Looking at a single gene in a single cell, the initiation of transcription at its promoter site is driven by the association and dissociation of a very small number of molecules. This very low copy number of molecules has two consequences: the time of reaction events can only be described in terms of probabilities and changes in the number of molecules are discrete, with no obvious trend that could be approximated with a differential equation (see [Paulsson, 2005] for a review). The expression of a gene does however serve a function; say during the cell cycle, growth, differentiation or apoptosis of the cell. For example, in response to external stimuli, the cell may produce large quantities of a protein. This response, measured as an apparently smooth/monotonic change in concentration, appropriately described

by differential equations. Small fluctuations around an obvious trend/mean are thus ignored. At this level we are aiming at a description of a pathway acting as a switch, filter, oscillator, amplifier, studying the network's behavior in terms of its robustness, responsiveness, sensitivity of the model to changes in parameters, transitions between steady states and bifurcations. A usual assumption in such rate equation models is that parameters (rate coefficients) are constants. Since these parameters are implicitly linked to environmental variables, such as temperature, pH level or water balance, fluctuations in these are considered negligible. The art of modeling is then to decide in the given context which modeling approach or combination thereof is most appropriate. Even if ordinary differential equations are chosen, noise can influence the onset of oscillations. An example, serving as a toy model for this, is the "Brusselator" [Blomberg, 2006]. Here one observes damped oscillations around the stationary point before the oscillation bifurcation occurs. Noise afflicts these damped oscillations, and this gives information about the bifurcation before it appears, in the region of a stable stationary point. Thus, noise provides information and details about the type of bifurcation that are not as clear in the basic differential equations.

As pointed out in previous sections, in experiments one can only study a limited number of components and generate data for them. The unavoidable conceptual closure in modeling and the neglect of system variables, will inevitably lead to uncertainty in the analysis of a complex system, providing an epistemological motivation for stochastic models.

David Bohm [1957] argued for the possibility that there might be an ever-recurring dialectic between causality and chance - or stochasticity and determinism in nature. If there could be an infinite number of levels of existence of matter then for each level which manifested itself stochastically there could be a level below to which that could be deterministically reduced: but, conversely, each deterministic level could reflect some average behaviour of a complex stochastic level below.

9 WHAT CAN WE KNOW ABOUT LIVING SYSTEMS?

As the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost exclusive characteristics.[24] Our understanding of complex systems arises from reducing one type of reality into another.[25] A complex system is by definition too complicated to be comprehended by just using everyday common sense. Studying complex systems through mathematical modelling is therefore to seek an understanding through abstraction. In other words, studying complex systems we put a high value on simplicity. The problem is that abstraction itself can be complicated to start with and thus abstraction is often not perceived as what it really is: simplification.

[24]The statement has been attributed to Lotfi Zadeh.
[25]The statement is attributed to Claude Levi-Strauss.

Dealing with complexity by reducing it in modelling suggests a loss of predictability. A model reduces a complex biological process to an essential aspect of its behavior, removing a general principle by which a cell functions from its experimental context of a particular culture, cell line or organism. All models are wrong, some are useful.[26] Abstraction lives from the fact that not everything that is there in a natural system needs to be modelled. To reduce or simplify a complex system into a tractable form requires however an understanding of the natural system in question. Simplicity appears thus to follow understanding; to understand a process we need to know it well. On the other hand, an understanding of a complex system requires simplification; we are dealing with an iterative, exploratory and creative process. Systems biology is indeed the *art* of making appropriate assumptions.

Systems biology signals a shift of focus away from molecular characterization towards an understanding of functional activity; away from studying the function of genes and proteins towards an understanding of cell function, supporting inferences about phenomena at the physiological level of cell populations, tissue, whole organs and whole organisms. A skeptic might argue that this is about the same as trying to predict the world economy from observations I make at my local superstore. While this endeavor seems impossible due to the complexity of cells, encouragement comes from the likes of Max Weber: "All historical experience confirms that men might not achieve the possible if they had not, time and time again, reached out for the impossible."; Mike Mesarovic: "It is less frustrating not to catch a big fish than it is not to catch a small fish - we might as well ask the big questions." and Richard Feynman: "We do not know where we are 'stupid' until we 'stick our neck out,' and so the whole idea is to put our neck out.".

If something seems impossible we improve our chances of success by trying it. In the meantime the interdisciplinary endeavor systems biology would benefit from the involvement of philosophers of science, discussing the process by which we model complex systems. First steps in this direction have been made [Fox-Keller, 2002], Boogerd et al.. 2007]. In an essay for the journal *Nature* Fox-Keller [2007] discusses the differences between physics and biology and asks whether biology does have physics-style laws that are universally applicable? When limits to the generality of findings are found in biology, this is usually not considered a problem and simply sets the context for the findings. Evelyn Fox-Keller asks whether exceptions to presumed laws are just a reminder of the complexity of biological systems or whether biologists should adopt a different attitude and systematically search for all-encompassing laws. She concludes: "Even though we cannot expect to find any laws governing the search for generalities in biology, some rough, pragmatic guidelines could be very useful indeed." System biologist may already benefit from reading the first pages of Popper [1959], where he quotes Novalis: "Hypotheses are nets: only he who casts will catch."

[26] The statement is attributed to George E.P. Box.

ACKNOWLEDGEMENTS

The authors would like to thank Cliff Hooker and Ludger Jansen for corrections and suggestions in the writing of this text. Discussions with Mike Mesarovic and Jannie Hofmeyr informed various parts of the essay. The result of our efforts is unfortunately little more than a documentation of our state of confusion about the consequences of cell biological complexity. We therefore take full responsibility for the remaining misinterpretations and misguided conclusions. O.Wolkenhauer acknowledges support from the ExCell project at the Center for Logic, Philosophy and History of Science, University of Rostock, funded by the regional government of Mecklenburg-Vorpommern, Ministry for Education, Science and Culture (Project Nr. 62327030). and support from the Stellenbosch Institute for Advanced Study (STIAS) in South Africa, providing funding and an exceptional environment to study the complexity of biological systems.

BIBLIOGRAPHY

[Alberts *et al.*, 2008] B. Alberts, A. Johnson, P. Walter, J. Lewis, M. Raff, and K. Roberts. *Molecular Biology of the Cell*, 5^{th} revised edn,. Taylor & Francis, London, 2008.

[Allen, 2001a] J. F. Allen. Bioinformatics and discovery: Induction beckons again. *BioEssays*, 23(1), 104–107, 2001.

[Allen, 2001b] J. F. Allen. In silico veritas. Data-mining and automated discovery: The truth is in there. *EMBO reports* 2(7), 542–544, 2001.

[Allen, 2001c] J. F. Allen. Hypothesis, induction and background knowledge. Data do not speak for themselves. Replies to Donald A. Gillies, Lawrence A. Kelley and Michael Scott. *BioEssays* 23(9), 862–862, 2001.

[Bertalanffy, 1969] L. von Bertalanffy. *General Systems Theory.* George Braziller, New York, 1969.

[Blomberg, 2006] C. Blomberg. Fluctuations for good and bad: The role of noise in living systems. *Physics of Life Reviews* 3, 133-161, 2006.

[Bohm, 1957] D. Bohm. *Causality and Chance in Modern Physics*, Routledge and Kegan Paul Ltd., London, 1957.

[Boogerd *et al.*, 2007] F. C. Boogerd, F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff, eds. *Systems Biology: Philosophical Foundations.* Elsevier, Amsterdam, 2007.

[Cassirer, 1981] E. Cassirer. *Kant's Life and Thought.* Yale University Press, New Haven, 1981.

[Cornish-Bowden, 2004] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*, 3^{rd} revised edn. Portland Press, London, 2004.

[Ehresmann *et al.*, 2007] A. C. Ehresmann , and J.-P. Vanbremeersch. Memory Evolutive Systems: Hierarchy, Emergence, Cognition. *Studies in Multidisciplinarity* 4, Elsevier, Amsterdam, 2007.

[Fell, 1997] D. Fell. *Understanding the Control of Metabolism.* Portland Press, London, 1997.

[Fox Keller, 2002] E. Fox Keller. *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines.* Harvard University Press, Harvard, 2002.

[Fox Keller, 2007] E. Fox Keller. A clash of two cultures. *Nature* 445(8), 603, 2007.

[Gillies, 2001] D. A. Gillies. Popper and computer induction. *BioEssays*,23(9), 859–860, 2001.

[Kærn, *et al.*, 2005] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in Gene Expression: From Theories to Phenotypes. *Nature Reviews Genetics* 6, 451-464, 2005.

[Kalman *et al.*, 1969] R. E. Kalman, P. L. Falb, and M. A. Arbib. Topics in Mathematical Systems Theory. McGraw-Hill Book Company, New York, 1969.

[van Kampen, 1992,] N. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland/Elsevier, Amsterdam, 1992.

[Kampis, 2003] G. Kampis. *Self-Modifying Systems in Biology and Cognitive Science.* Pergamon Press, New York, 2003.

[Kant, 1892] I. Kant. *Critique of Judgement*. Translated with Introduction and Notes by J.H. Bernard, 2^{nd} revised edition, Macmillan, London, 1914.
[Katok and Hassselblatt, 1995] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1995.
[Kell, 2003] D. Kell. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* 26(1), 990–105, 2003.
[Kelly and Scott, 2001] L. A. Kelley and M. Scott. On John Allen's critique of induction. *BioEssays* 23(9), 860–861, 2001.
[Kholodenko, 2006] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology* 7, 165-176, 2006.
[Klir, 1991] G. J. Klir. *Facets of Systems Theory*. Plenum Press, New York, 1991.
[Luisi, 2006] P. L. Luisi. *The emergence of Life: From chemical origins to synthetic biology*. Cambridge University Press, Cambridge, 2006.
[Maturana and Varela, 1987] H. R. Maturana and F. J. Varela. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala Publications, New York, 1987.
[Mayr, 2004] E. Mayr. *What Makes Biology Unique?* Cambridge University Press, Cambridge, 2004.
[Mesarovic, 1968,] M. D. Mesarovic. Systems Theory and Biology—View of a theoretician. *System Theory and Biology*, 59–87, 1968.
[Mesarovic *et al.*, 1970] M. D. Mesarovic, M. Macko, and Y. Takahara. *Theory of Hierarchical, Multilevel, Systems*. Academic Press, New York, 1970.
[Mesarovic and Tkahara, 1975] M. D. Mesarovic and Y. Takahara. *General Systems Theory: Mathematical Foundations*. Academic Press, New York, 1975.
[Mesarovic *et al.*, 2004] M. D. Mesarovic, S. N. Sreenath, and J. D. Keene. Search for organising principles: understanding in systems biology. *IEE Systems Biology* 1(1), 19-27, 2004.
[Morgan, 2006] D. Morgan. *The Cell Cycle: Principles of Control*. Oxford University Press, Oxford, 2006.
[Muir, 1982] A. Muir. Holism and reductionism are compatible. In S. Rose (ed.), A*gainst Biological Determinism*. Alison and Busby, London, 122–135, 1982.
[Nicolis and Prigogine, 1989] G. Nicolis and I. Prigogine. *Exploring Complexity*. W.H. Freeman and Company, New York, 1989.
[Novák and Tyson, 2008] B. Novák and J. J. Tyson. Design principles of biochemical oscillators. *Nature Reviews Molecular Biology* 9, 981-991, 2008.
[Padulo and Arbib, 1974] L. Padulo and M. A. Arbib. *Systems Theory*. W.B. Saunders, Philadelphia, 1974.
[Paulsson, 2005] J. Paulsson. Models of stochastic gene expression. *Physics of Life Reviews* 2, 157-175, 2005.
[Popper, 1959] K. Popper. *The Logic of Scientific Discovery*. Routledge, London (first published in 1959), 1992.
[Raj and van Oudenaarden, 2008] A. Raj and A. van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135(17), 216-226, 2008.
[Rao *et al.*, 2002] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231-237, 2002.
[Rosen, 1985] R. Rosen. *Anticipatory Systems*. Pergamon Press, New York, 1985.
[Rosen, 1991] R. Rosen. *Life Itself*. Columbia University Press, New York, 1991.
[Rosen, 2000] R. Rosen. *Essays on Life Itself*. Columbia University Press, New York, 2000.
[Shahrezaei *et al.*, 2008] V. Shahrezaei, J. F. Ollivier, and P. S. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology* 4 (196), doi:10.1038/msb.2008.312008
[Smalheiser, 2002] N. R. Smalheiser. Informatics and hypothesis-driven research. *EMBO reports* 3(8), 702, 2002.
[Solé and Goodwin, 2000] R. Solé and B. Goodwin. *Signs of Life — How Complexity pervades Biology*. Basic Books, New York, 2000.
[Tyson *et al.*, 2003] J. J. Tyson, K C. Chen, and B. Novák. Sniffers, buzzers, toggles, and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology* 15, 221-231, 2003.
[Ullah and Wolkenhauer, 2007] M. Ullah and O. Wolkenhauer. Family tree of Markov models in systems biology. *IET Systems Biology* 1(4), 247-254, 2007.

[Ullah and Wolkenhauer, 2008] M. Ullah, O. Wolkenhauer. Investigating the two-moment characterization of subcellular biochemical networks. *Journal of Theoretical Biology* doi:10.1016/j.jtbi.2009.05.022 (in press).

[Weaver, 1948] W. Weaver. Science and Complexity. *American Scientist* 36 (4) 536-44, 1948.

[Wiggins, 2003] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, 2^{nd} edn. Springer, New York, 2003.

[Wolkenhauer and Mesarovic, 2005] O. Wolkenhauer and M. Mesarovic. Feedback Dynamics and Cell Function: Why Systems Biology is called Systems Biology. *Molecular BioSystems* 1(1), 14-16, 2005.

[Wolkenhauer, 2007] O. Wolkenhauer. Defining Systems Biology: An Engineering Perspective. *IET Systems Biology* 1(4), 204-206, 2007.

[Wolkenhauer and Hofmeyr, 2007] O. Wolkenhauer and J.-H. S. Hofmeyr. An Abstract Cell Model that describes the Self-Organization of Cell Function in Living Systems. *Journal of Theoretical Biology* 246, 461-476, 2007.