20

OUTSTANDING ISSUES IN SYSTEMS AND SYNTHETIC BIOLOGY

Pengcheng Fu[1] and Cliff Hooker[2]

[1]*Faculty of Chemical Science and Engineering, China University of Petroleum, 18 Fuxue Road, Changping District, Beijing 102249, People's Republic of China*
[2]*Faculty of Education and Arts, School of Humanities and Social Science, The University of Newcastle, NSW, Australia*

In this book, we have discussed a number of applications of systems biology and synthetic biology. In fact, the scope and potential applications of systems biology and synthetic biology are not yet fully defined. As we ponder the future directions in biology research, there remain many open issues, including those that are discussed.

20.1 OUTSTANDING SPECIFIC ISSUES

20.1.1 Systems Biology and Synthetic Biology for the Investigation of Nonprotein-Coding RNAs

The epoch of systems biology and synthetic biology began when whole-genome sequences for various organisms started to accumulate. The amount and precision of this information made it possible to map the coding and noncoding regions and the hierarchy of regulatory mechanisms, relationships among structural and functional assemblies, subcellular organelles and compartments, and interaction with external signals. One of the most important discoveries of the last few years has been the

Systems Biology and Synthetic Biology Edited by Pengcheng Fu and Sven Panke

identification of small, nonprotein-coding RNAs (ncRNAs) that act as integral regulatory components of cellular networks [1]. ncRNAs serve an astonishing variety of functions and thus play important roles in many intracellular processes, from transcriptional regulation, gene silencing, chromosomal replication, through RNA processing and modification, mRNA stability and translation, to protein degradation and translocation, and so on [2]. The size of ncRNAs range from about 20 nt for the large family of microRNAs (miRNAs) that modulate development in *Caenorhabditis elegans*, Drosophila, and mammals [3–8] to 100–200 nt for small RNAs (sRNAs) commonly found as translational regulators in bacterial cells [9,10] and up to 10,000 nt for RNAs involved in gene silencing in higher eukaryotes [11–13]. There are two approaches to searching for ncRNAs: computation methods that focus on intergenic regions and expression-based methods that examine expression levels of the transcripts [2]. Systematic identification and characterization of ncRNAs in genomes has become one of the most exciting challenges in cellular and development biology.

Among the noncoding RNA genes that produce functional molecules instead of encoding proteins, a large number of newly identified RNAs have been found to function as regulators [1]. These regulatory RNAs (reRNAs) impact all the steps in the genetic information pathways, and may serve as transcriptional regulators, translational regulators, modulators of protein function, or regulators of RNA and protein distribution. Study of many of these RNAs in bacteria and eukaryotes has shown a surprisingly high degree of similarity between regulatory RNAs in all types of organisms [1]. Therefore, insights gained by investigation of the regulatory role of reRNAs using one system are applicable to other systems as well. reRNAs are now recognized to play important roles as regulatory elements, yet little work has been done on a global scale to identify these intracellular regulators. Elucidation of correlations between expression levels of regulatory RNAs and cell metabolism such as photosynthesis and respiration may reveal the occurrence of hitherto unknown regulatory mechanisms. This information may clarify the mechanisms of gene expression and gene regulation. It may also facilitate rational engineering of the signaling, regulatory, and metabolic networks for desirable cellular functions.

Compared to protein-coding RNAs, ncRNAs are relatively small. ncRNAs are hard to find by classical mutational screens because they are inherently immune to frameshift or nonsense mutations [14]. Therefore, limitations exist for both computation-based and expression-based ncRNA detection methods. Recently, research efforts have been made to carry out systematic ncRNA gene-identification screens along three main lines: cDNA cloning and sequencing tailored to find new small non-mRNAs [15]; specially designed cDNA cloning screens for a new regulatory RNA gene family of miRNAs [3–5]; and comparative genome analysis for general ncRNA gene finding [16–18].

Systematic identification and characterization of ncRNAs in bacterial and eukaryotic genomes has become one of the most exciting challenges in cellular and development biology. There exists a controversial ‘‘introns-first’’ theory [19] that states that from the evolutionary point of view, the ncRNA molecules predate the origin of protein translation and therefore predate the exons surrounding them. It is believed that the contemporary introns housing functional RNAs are ancient relics of

the RNA world genome organization, and the newer protein regions surrounding them represent sequences that were originally noncoding and from which protein genes were eventually spawned [20]. On the contrary, search for new ncRNAs has resulted in finding many such ncRNAs with apparently well-adapted and specialized biological roles in the cellular transcription machinery [14].

Now can we use systems biology approaches to strive to understand how information flow in cells is adjusted by particular reRNAs, and how expression, function, and turnover of these reRNAs themselves are controlled. Elucidation of correlations between expression levels of reRNAs and metabolic flux distributions under different environmental perturbations may reveal the occurrence of hitherto unknown regulatory mechanisms. This information may clarify the mechanisms of gene expression and gene regulation. The next issue is whether we can facilitate rational engineering of the signaling, regulatory, and metabolic networks containing ncRNAs for desirable cellular functions.

20.1.2 Dimension Reduction in Systems Biology and Synthetic Biology Applications

Systems biology is inherently a universe in which every ''ome''—genome, transcriptome, proteome, metabolome, interactome, phenome, and so on, is another dimension. We have to reduce this dimensionality through integration in order to comprehend, evaluate, and make use of the information. Integrating and evaluating the knowledge bases with their highly disparate nomenclature and frames of reference is arguably the greatest methodological challenge in this new discipline. One example is the concisely described work of Toyoda and Wada, who have developed means of defining the dimensions of several data sets in common terms and projecting the intersections of these sets in two dimensions [21]. Their premise is that the intersections have four defining properties: data set, position, dynamics, and probability that the putative relationship actually exists. The implementation of their ''genome-phenome superhighway'' (GPS) for human, mouse, the worm *Caenorhabditis elegans*, and the mustard plant *Arabidopsis thaliana* may be found at http://omicspace.riken.jp/gps.

Each ''omic'' domain has its own unique annotation terminology and attributes, which has led to the development of unique ''markup languages'' compatible with the Internet HTML, Perl, and other computational data-handling conventions. These include ''G-language'' for the genomics environment [22], CellML, MathML, and SBML (systems biology markup language) [23], to name a few.

Given the huge amount of data produced in array-based studies, how does one (a) assess its reliability, (b) interpret it in a systematic, unbiased way, and (c) determine the completeness of the data set? A substantial literature has been developed just to address each of these questions. Chen et al. [24] provide a brief introduction and guide to the reliability and analysis aspects. Reliability is affected by the physical quality and composition of the array, stringency of the experimental conditions, background gene expression, and similarities among the probes. Nonlinear responses are inherent in biological systems, so appropriate nonlinear multivariate analysis is essential. Further research is needed to enable efficient database search, development

of programming language and data fusion for systems level understanding in biology and the integration of well-characterized biological parts into genetic circuits and metabolic networks for desired end products.

20.1.3 The Quest for the ''Minimal Organism'' and the Creation of Artificial Life Forms

Systems biologists and synthetic biologists are interested in determining the smallest set of genes, molecules, and structures for replication, growth, metabolism, and regulation that comprises life. Study of such a minimal gene set and its features may shed light on the basics of cellular function, help to determine the subset of essential genes in most species, and improve functionality. Theoretical and experimental efforts have been made using comparative genomics and systems analysis to determine the list of essential genes for a suite of minimal functions that many organisms have in common [25]. The smallest possible group of genes from small genomes is presumed necessary and sufficient for sustaining the functional growth of cells in the presence of a full complement of essential nutrients and in the absence of environmental stress [26].

Methods for making estimates of the ''minimal gene set'' by experimental biology include saturating transposon mutagenesis (gene knockout) [27] and gene silencing with antisense RNA [28], and so on. These genes can also be computationally identified from the well-studied organisms with small genomes by comparison of essential and nonessential proteins across related genera [29], and using a database of essential genes [30]. For example, *Mycoplasma genitalium* contains the smallest genome of any organism, and has a minimal metabolism. Glass et al. [31] have used global transposon mutagenesis to isolate and characterize the gene disruption mutants for 100 different nonessential protein-coding genes. They have identified 382 essential genes from the 482 *M. genitalium* protein-coding genes. Disruption of some genes accelerated the *M. genitalium* growth. The resulting *M. genitalium* mutants represent a close approximation to the minimal set of genes needed to sustain bacterial life, with little genomic redundancy [32]. Another study, analyzing viable gene knockouts in *Bacillus subtilis*, *M. genitalium*, and *Mycoplasma pneumoniae*, has resulted in a similar estimate [33]. It was found that approximately 80 genes out of the 250 in the original minimal gene set are represented by orthologs in all life forms. For $\sim$15 percent of the genes from the minimal number of genes, viable knockouts were obtained in *M. genitalium* [25]. *Escherichia coli* is also used as a model system for gene knockout to create a reduced ''clean genome.'' Fred Blattner's team [34] has removed about 750 ''redundant genes'' and planned to delete 500–600 more genes to approach the ''core genome'' that may be common to all organisms. It was claimed that after the gene removal, the constructs were observed to be more genetically stable and to exhibit increased protein synthesis and electroporation efficiency [34].

The quest for the minimal genome will improve our understanding of the workings of bacterial lives at systems level. On the other hand, the ultimate dream of the synthetic biologists is to create novel life forms that do not exist in nature. For this purpose, minimal organisms may be built up by designing a modular system from ''ground zero'' that can be given functions. It may involve the design and assembly of

genetic circuits and metabolic pathways and even whole chromosomes from chemical components of DNA. Researchers in Synthetic Genomics Inc. (http://www.syntheticgenomics.com/index.htm) have achieved this technical feat by chemically making DNA fragments in their laboratory and developing new methods for the assembly and reproduction of the DNA segments.

The goal is to obtain a synthetic chromosome, and eventually a synthetic cell for the construction of ''biofactories'' for the energy, chemical, and pharmaceutical industries. The synthetic chromosome created in Synthetic Genomics, Inc. was named *Mycoplasma laboratorium*. It can be transplanted into a living cell where it should ''take control'' of the cellular metabolism. Although *M. laboratorium* was claimed as a man-made bacterium, there exist some questions about it because the partially synthetic life form was composed of building blocks from already existing organisms. Even when the whole chromosome can be synthesized from chemical components, will we consider the engineered cells to be new life forms, or should we also require the synthesis of ribosomes and other components necessary for the expression of genetic information contained in the genome before accepting the result as an ''authentic'' new cell?

20.1.4 Systems Biology and the Evolution of Organelles

The properties, genomes, and functions of plastids and mitochondria are an obligatory part of systems biology studies of eukaryotes. Genomic and biochemical studies have established that mitochondria most likely evolved from the rickettsial group of α-proteobacteria [35]. The reRNA sequences in the genomes of aerobic mitochondria are most homologous to those of α-proteobacteria, specifically those of *Rhodospirillum*, *Bradyrhizobium*, and *Rickettsia* [36]. Homologues of 18 different rickettsial proteins are encoded in mitochondrial DNA, and in yeast, the nuclear genome encodes more that 150 mitochondrial proteins with homologues in *Rickettsiales* [37]. The rickettsial pathway for ATP production and that of aerobic mitochondria are virtually identical, and the individual enzymes are orthologs. The properties of rickettsia as an obligate intracellular pathogen, its ability to transport molecules in either direction across its cell walls, and other key factors firmly support the concept that aerobic mitochondria evolved from α-proteobacteria.

Anaerobic environments ranging from sea floor sediments to the gastrointestinal tracts of vertebrates and invertebrates are populated by extremely diverse communities of lower single-cell and multicellular eukaryotic life forms. Some eukaryotes have adapted to anaerobic life by using alternate mitochondrial respiratory pathways, such as reduction of fumarate to succinate, using rhodoquinone instead of ubiquinone as electron carrier [38,39]. These organisms retain their mitochondria and have been called ''Type I anaerobic eukaryotes.'' Additional organisms with ''anaerobic mitochondria'' include the fungus *Fusarium oxysporum* that uses a nitrate respiration pathway, platyhelminthes that utilize fumarate respiration, and trypanosomes that produce succinic acid while making ATP (summarized by Rotte et al. [40]). Perhaps most unusual is the mitochondrion of the anaerobic ciliate protist *Nictotherus ovalis*, which generates ATP with protons as the terminal electron acceptor, thus producing molecular hydrogen [41,42].

A second group of primitive anaerobic eukaryotes, most notably the parasitic trichomonads such as *Tritrichomonas fetus* in cattle, *Trichomonas vaginalis* in humans, some ciliated protozoa, and the cattle rumen chytrid fungi *Neocallimastix* and *Piromyces* have developed hydrogenosomes—an organelle with intriguing similarities and differences compared to anaerobic, as well as aerobic mitochondria [43]. Hydrogenosomes produce ATP as well as hydrogen. Although mitochondria use pyruvate dehydrogenase, the TCA cycle to regenerate CoA∼SH, and molecular oxygen as the terminal electron acceptor, hydrogenosomes have pyruvate-ferredoxin oxidoreductase, no TCA cycle, succinate-acetate CoA transferase and succinyl-CoA synthase to regenerate CoA∼SH, and protons as the terminal electron acceptor [40]. Hydrogenosomes and mitochondria use the same "transit peptides" for protein importation. Other proteins common to both organelle types include Hsp 10, Hsp 60, and Hsp 70, the succinyl-CoA synthase subunits α and β, and similar variants of ATP-ADP translocase. Early studies found no DNA in hydrogenosomes. In 1998, Akhmanova et al. [42] described genomic DNA in putative "hydrogenosomes" in the anaerobic ciliate *Nictotherus ovalis*. In retrospect, it appears more correct to describe the organelle in this species as an anaerobic mitochondrion.

The discovery of single-celled eukaryotes that had no mitochondria or hydrogenosomes originally suggested that these organisms were ancestors of eukaryotes that had the organelles (presumably endosymbionts of bacterial or archaean origin). These simple eukaryotes of four types—Metamonads, Microsporidia, Parabasalia, and Archamoebae—were grouped as a subkingdom called Archezoa [44] to distinguish them from Mitozoa, the subkingdom of all eukaryotes that contain mitochondria [45]. Ribosomal RNA sequencing indicated that the Archezoa predated the other known eukaryotes, and Archaezoan ribosomes were 70S, corresponding to those of prokaryotes. However, subsequently it was shown that several Archaezoa contained enzyme-coding mitochondrial DNA sequences [46]. Trichomonas, a Parabasalian, were found to contain hydrogenosomes, Microsporidia undergo meiosis and have tubulin genes that relate these taxa to fungi, and at least one type of Metamonad expresses a chaperonin immunochemically homologous to mitochondrial cpn60. In summary, the more recent and definitive research indicates that Archezoa are among the earliest eukaryotes, but they do not predate the endosymbiosis of mitochondria and hydrogenosomes [44]. That various Archezoa lack mitochondria, hydrogenosomes, peroxisomes, or other organelles as presumptive endosymbionts has been called, "...a secondary reduction caused by their parasitic lifestyle." [45]

In summary, a critical mass of genomic, proteomic, and phylogenetic data has finally accrued to support a comprehensive hypothesis for the origin of eukaryotes, consistent with the known properties of anaerobic and aerobic mitochondria, as well as hydrogenosomes. This hypothesis takes into account the metabolic pathways, electron transport chains, protein importation signals [47], gene loss and transfer to the host cell nucleus [48,49], assembly of Fe–S centers and their incorporation into apoproteins [50], chaperonins, enzymes, and lipids of the endoplasmic reticulum and nuclear envelope membranes.

How, and from what progenitors, did the first single-cell and multicelled eukaryotes develop? How did the nucleus, mitochondria, chloroplasts, and other organelles originate? What primal events led to the formation of chromosomes? Or the mechanisms of cell division? Or meiosis and sexual recombination? At the molecular level, the following questions include: How and when did RNA and DNA first develop? What transformed a world based on RNA as a carrier of genetic information and enzymatic activity into one in which the genetic information resided in DNA, and enzymatic catalysis was endowed in proteins?

A great deal of evidence acquired over the past 30 years supports the theory that mitochondria and hydrogenosomes originated as bacteria that developed an endosymbiotic relationship with eukaryotes. Various hypotheses have been forwarded to reconcile experimental data with how, and at what stage of evolution, the symbiosis occurred.

20.2 OUTSTANDING GENERAL ISSUES

An appropriate framework for systems and synthetic biology requires the construction of a naturalistic paradigm and philosophy of science for biological research. That project remains incomplete. In what follows a number of component issues in that project are discussed. Although progress toward a naturalistic understanding has been made in each case, there are also future unresolved challenges to the task.

20.2.1 Mechanism and Reduction

While the new high-throughput experimental technologies can profile all the chemical components within a cell as whole, and this has an interest in itself, the ultimate goal is to understand cellular physiology, that is, to understand how these components deliver cellular functioning. To do this it is necessary to study the dynamical interrelations among the components. And because the components must clearly interrelate in multiple ways to deliver function, it will be the complex dynamical system they jointly comprise that must be uncovered. It is this object that is the common core of systems and synthetic biology, and it is its representation as a complex dynamical system that forms the basis of their distinctively new and powerful modeling tools—and raises the issues in Section 20.1 (see Section 20.2.6 for a further issue within this claim)

This raises a basic ontological issue (i.e., one concerning what exists): what is the relationship between physiologically described function and biochemically described dynamical states and processes? The obvious response to make is that the two are one and the same; that, for example, aerobic cellular respiration is nothing but ATP synthesis through glycolysis, Krebs cycling, and electron transport. This is reduction by identification. The physiological function of respiration is identically reduced to, is identical to, and so nothing other than the dynamical system process. (All this assuming that the biochemical systems models involved are empirically supported

and predictively and explanatorily adequate; an assumption made throughout this discussion.[1])

There is a large philosophical literature on reduction, some of it proclaiming reduction and much arguing against reduction, especially in biology. Yet, from a scientific point of view it would be anomalous to claim anything less than a reduction, for example, to claim instead just a correlation between the occurrence of functional and biochemical systems properties, because this would leave unexplained duplicate realities, one functional and the other dynamical. Against the advice of Occam's razor, it would leave two realms mirroring each other but running in parallel, for no substantive reason. In what follows the state of philosophical debate is briefly summarized, from a commonsense scientist-friendly point of view, in order to focus on the specific issues at stake for systems and synthetic biology.

20.2.1.1 General Objections Perhaps surprisingly, one group of philosophical objections to reduction in general argues that correlation must be accepted because identification is impossible. These arguments largely turn on semantic (meaning) considerations: talk of functioning, for example, of respiring, the argument goes, has a very different meaning from talk of biochemical states and state transitions, so the two can never be identified, even if they are 1:1 correlated. The proper response to this kind of objection is to point out that it relies on *a priori* claims about semantics that are very unlikely in the face of what we know scientifically about language: roughly, that its recent evolutionary emergence, rapid dynamical shifts in vocabulary, syntax, and semantics as historical conditions change, and action-centered intentional basis, all suggest that current semantics are better treated as themselves shifting dynamical emergents, not *a priori* constraints.[2] There would need to be better reasons than these to defeat a general identification of the two subject matters described.

Another group of arguments turns on the fact that the mirroring is often not precise, that often there will be particular phenomenological conditions (e.g., ''respiration'') that do not nicely reduce to exactly corresponding underlying conditions (e.g., ''ATP synthesis'') of exactly the same scope. This is true, and not only because of the anaerobic organisms and other energy storage molecules, but also because of the complex dynamics. For instance, even Kepler's laws of planetary motion do not reduce exactly to a theorem of Newtonian mechanics because planet–planet interactions produce small deviations from Kepler's generalizations. This will be a common

[1] The issue of when and why that assumption is reasonable is just the general issue of the nature of scientific method at large. It turns out that scientific method is much more complex (and interesting!) than the neat logical models to which the philosophers had hoped to reduce it, and must itself be understood in dynamical systems terms, but this is another story—see Ref. [51] and, for example, Ref. [52].

[2] For those interested in the technicalities, a same-dynamical-role criterion of property identity is a useful small first step toward a more plausible alternative semantics and this already suffices to license identification of functions with dynamical processes, should other substantive requirements be met. This is argued in Part II of Ref. [53]. For a bioorganizational approach to the underlying intentionality, see Ref. [54].

situation wherever a more complex dynamics underlies more phenomenological observations. In such cases, surely, so long as the departures from strict correspondence can also be explained by the underlying (reducing) dynamics, the reduction can be considered successful. Call the last the explanatory principle.

This works well for cases where the departures are small. However, there are also large departures, such as in the relationship of phlogiston chemistry to oxygen chemistry, where we deny that phlogiston exists even if its postulation served to codify a number of chemical relationships that survive the replacement. And there are intermediate cases, for example, the imperfections of the thermodynamics–statistical mechanics relation. How are these to be treated? To decide we need to remind ourselves that for science reduction is not only about satisfying metaphysical curiosity, from a methodological point of view, but it is also primarily about extending explanation and evaluating the potential errors involved in using the phenomenological model to explain, in place of the underlying one. (Hence the explanatory principle above.) From this perspective, reduction is ultimately about the capacity to systematically replace one kind of description (the more phenomenological one) with another kind (the more basic, theoretical one) that is equally or more precise and equally or more predictively and explanatorily powerful. This satisfies the key cognitive aims of science. Reduction by identification then forms one extreme of a spectrum, where component ontology as well as relational structure is conserved under the replacement. The other extreme is occupied by cases like phlogiston where significant relational structure, but not ontology, is conserved under replacement.[3] Mismatch along the spectrum means that some nonconservation is melded with identificatory reduction. However, the key point remains that when the explanatory requirement holds, overall reduction is obtained.

20.2.1.2 Geneticism These general issues aside, an important part of the philosophical objection to specifically biological reduction has really been to geneticism, to the idea that organisms could be reduced to just a collection of genes and gene-determined traits. Modern biology agrees with this objection, DNA is one biochemical component among many—if with a distinguishable role—and it is the dynamical system of all of them that sustains function. But, conversely, the whole biochemical system now becomes the reduction candidate for physiology, so the objection to geneticism does not defeat reduction, but just shifts its focus. Setting aside that literature as well, there remains only those objections that are specific to reduction of functions to systems dynamics.

20.2.1.3 Reduction of Function to Dynamics Some objections to this have to do with the fact that our commonsense day-to-day function talk is rather

[3] Beyond that, sheer discontinuous replacement would occur, but it is hard to think of a substantial case in science. For the replacement view, see Part I of Ref. [53] and, more informally and accessibly, Ref. [55]. P. M. Churchland's elegant overall strategy, more subtle but powerful than it may appear, is itself explained in Ref. [56].

imprecise for marrying up to dynamical systems specifications, while others stem from the related problem that vague function descriptions can seem to cut across what turn out to be the dynamical process distinctions. These can all be resolved through a little careful analysis of language.[4] This is useful to know in a field like biology where talk of functions is ubiquitous, but often more pragmatic than precise, especially considering that only features that have functional consequences are likely to be modeled.

Setting general objections to reduction from function talk aside as well, brings us at last to the substantive conditions for function to system process reduction. First, we specify a function as a map from inputs to outputs. For example, cellular respiration, crudely globally specified, is the function that takes food and water molecules as inputs and outputs carbon dioxide. Note that more specific functional maps capturing the process detail can clearly be constructed as required. Corresponding to this in the molecular description is a dynamical process—that is, a metabolic map carried by (biochemical) dynamical laws—that takes oxygen and glucose as inputs and yields ATP (and perhaps other energy storage) and carbon dioxide as outputs. Then the obvious requirement for identificational reduction is that the respiration functional map be embeddable into the corresponding biochemical process map without distortion (homomorphically embeddable). A further coherence condition is equally obvious: the collection of all such embedded dynamical maps, together with any nonfunctional data concerning the system, should provide a single coherently unified biochemical model of the cell genome that preserves or increases predictive and explanatory power.[5] The embedding criterion essentially captures recent conceptions of a function to mechanism reduction, reducing both the cell and multicellular organisms to complexes of mechanisms.[6]

There is an inherent underdetermination by any function, taken in isolation, of its correct embedding. Although this has sometimes been taken as a fundamental objection to reduction, it ultimately reduces to a pragmatic issue of sufficient data. The problem is nicely illustrated in the case of the output of a network of electrical generators having a frequency variation less than that of any one generator; some kind

[4] See Part III of Ref. [53] and, briefly, Ref. [57], Part V, case I and case II end.

[5] See Part III of Ref. [53] and, briefly, Refs [56] and [57]. The basic reduction requirement, that functional maps are mirrored by dynamical maps, is in fact just the application of Nagel's [58] deductive reduction conception, rightly understood. Nagel shows how scientists arrive at reduction of a law L_2 or property P_2 of theory T_2 respectively to a law L_1 or property P_1 of theory T_1 by first showing how to choose conditions (real or idealized) under which it is possible to construct in T_1 a law L_1 or property P_1 that will mirror (be a relevantly isomorphic *dynamical* image of) the dynamical behavior of L_2 or P_2. From that the reduction is shown to be possible through the identification of L_2 or P_2 with the mirroring L_1 or P_1. Indeed, the requisite "bridging" conditions can be deduced from the mirroring condition, and then asserted as identities on the basis that doing so will achieve a reduction, supported in that light by claims of spatiotemporal coincidence or appeal to Occam's razor.

[6] See especially Refs [59, 60]. However, the conception of mechanism here does not yet adequately reflect the importance of process *organization* to cellular function [61], an outstanding issue for future development. Further on organization, see Sections 20.2.3 and 20.2.4.

of feedback governing process is at work, but is it a real governor or simply the functional appearance of one at the network level? The latter is possible because connecting the electrical generators in parallel automatically creates a phase-stabilizing mutual interaction among them without the need for a real governor.[7] This question is resolved by gathering other data about the network–this is the point of the unification criterion above.

Nonetheless serious issues remain with the overall position. Rosen, for example, argued that organisms could not be complexes of mechanisms in any compositional sense and that they were indeed not mechanisms.[8] Disentangling the aspects involved, there remain these systems issues that must be resolved: (1) self-organization and emergence, (2) the nature of the complexity in ''complex of mechanisms,'' and (3) the specific implications of self-regeneration for (1) and (2). Of these (1) and (3) will pose specific challenges for reduction. Conversely, however, a thoroughly dynamical systems approach will allow us to understand the subtle intertwining of reduction and its failure in emergence within a unifying framework, providing a full, naturalist account of reduction and emergence in systems and synthetic biology. These three issues are now discussed separately and in order.

20.2.2 Self-Organization and Emergence

In all systems it is true that the interacting components together create a dynamics that would not otherwise be present. When the outcome is surprising or unexpected or too complex to be readily understood, scientists are apt to talk about self-organized emergent patterns. There are many reasons why leaving things like that is unsatisfactory, among them that (i) no significant feature is addressed, our subjective surprise, and so on, keeps shifting and has no substantive association with reality, and (ii) this criterion is dynamically so weak as to trivialize these ideas. But when it comes to strengthening the requirement, there is currently huge diversity of opinion about both the concepts, self-organization and emergence. Two broad approaches to identifying something more penetrating can be distinguished, one epistemic and the other causal or dynamical.

The epistemic approach tightens up the subjectivity by adding a clause along the lines that self-organization occurs when the resulting system dynamics could not have been predicted from the known interaction rules of the components. Since the dynamics is entirely internal to the system, it is properly referred to a self-organized.

This approach is attractive because there are many complex behavioral patterns that arise from the simplest interaction rules, for example, with social insects (hives of bees and termite mounds), city traffic, and even simple population dynamics as reflected in the logistic equation. However, it still ties the definition of evidently physical

[7] For this example see Ref. [62] and further Part III of Ref. [53].

[8] See Refs [63, 64]. Rosen's objections have to do with the role of global organizational constraints on organisms and are discussed under Section 20.2.4 below.

properties to a cognitive test, and anyway proves difficult to formulate satisfactorily.[9] So we pass to the option of a causal/dynamical criterion.

One causal/dynamical distinction stands out, and fixing on this avoids a long detour through a tortuous literature. The distinguished difference is between patterns that dynamically constrain their components—that show "top-down" dynamical constraints—and those that do not. Consider the formation of an iron bar from cooling molten iron. In this phase transition a macroscopic pattern of intermolecular relations is formed, the iron crystal, which does thereafter have the power to constrain the movements of its molecular components through the formation of a new macroscale force constituted in the ionic lattice bonds formed. Its formation alters not only individual component behavior but also the specific dynamics under which they are now able to move: there are lattice vibrations and a Fermi conduction band in place of liquid molecular dynamics, that is, the phase change alters the force form of the dynamical equations that govern component behavior. The new macroscale force is able to retain the constraint relationship invariant under component fluctuations and exogenous perturbations, through lattice dissipation of these perturbing energies as sound and/or heat.[10]

By contrast, from intersecting shallow waves on a gently undulating beach there emerges the most beautiful and intricate patterns, but there is no comparable constraint formed by their interaction; shift the underlying sand structure and the dynamics can shift to entirely other patterns. Similarly, there is no dynamical constraint internal to social insect societies comparable to the ferric crystal force and compelling their insect members to satisfy hive and mound laws, or compelling city drivers to create traffic jams, and so on. All of these patterns are produced by dynamical interactions of components and thus reflect their "bottom-up" dynamical constraints, but only some also express top-down dynamical constraint.

It is natural to choose the formation of a new top-down constraint as a criterion of emergence for just this characterizes the coming into being of a new dynamical existence. The iron top-down constraint formation constitutes the coming into being of

[9] As it stands, the text formulation is intolerably vague: Predicted by whom? Knowing what? Using what tools? And it makes an apparently ontological distinction (the existence of emergent behavior) depend on a cognitive condition (human predictive capacity). If, in response, the criterion is instead formulated along the lines of "cannot be derived from the set of interaction rules," then these problems are lessened, but only to be replaced by the problem of what counts as an acceptable derivation. If derivation includes computational modeling of collective dynamics then almost all dynamics counts as derivable and nothing self-organizes. (Perhaps noncomputable dynamics might be considered an exception, but since this occurs in quantum theory and other "wave" dynamics, it seems a peculiar boundary.) If instead derivation is restricted to logical deduction then almost everything self-organizes since the demand for analytic closed-form solutions fails for almost all sets of differential equations. No satisfactory criterion of in-between scope is readily formulable.

[10] The iron bar is a new macroscale level with respect to its molecular constituents because it has its own characteristic dynamical interaction form. All other talk of levels either concerns measurement (liquid level), gravitation (level surface), or is metaphorical (semantic, social, abstraction, theory . . . levels) and can thus be paraphrased away—or is confused. Note that the presence of a top-down constraint does not fully determine the specific dynamical form of the system; both the virtual and real electrical governor arrangements (see footnote 7 and text) exhibit the same phase-stabilizing top-down constraint. Distinguishing between them is the electrical engineering "system identification" problem.

a new, individuated capacity to do work, expressed both endogenously in dissipation of perturbations and exogenously in rigid body action. It is the arrival of a new dynamical individual characterized by a new dynamical form.[11] The character of the new individual is constituted by its capacity to do new work. To broaden the criterion further would be to conflate genuine interactive emergence with the mere emerging in time of a pattern (as "from concealment").

Real emergent dynamical filtering insures that macroscopic properties have the stability we find them to have, making the macroscopic world as viably simple to survive in as it is for macroscopic creatures like us. But it also applies to smaller-scale structures; cellular metabolic regenerative organization and the cellular structures it sustains, for example, are emergent top-down constraints and cellular function would not be stable without them.[12] But by providing higher level structure for lower level processes, all these constraints actually underpin the reduction of the functions served to dynamical processes (and of course the constraints themselves and attendant structures to dynamical compounds of the components whose interactions constitute them).[13] Emergence heralds the presence of an irreducibly new dynamical existent; reduction to the components alone fails. Yet, contrary to the standard view of reduction and emergence as opposed, this discussion shows that emergence and reduction are intricately interwoven and mutually supportive.

There is no physical mystery about this when a dynamical model of emergence is to hand since it is precisely what the filtering consequent upon formation of a new dynamical constraint provides. In this way, we naturalize emergence for science. And it is precisely on that general basis, and only on that basis, that we can track causal paths "up" and "down" through the component/supracomponent levels and thus, come to

[11] In a more traditional philosophical language, the iron bar is supervenient on its molecules; nothing about the bar can change without the change being dynamically grounded in appropriate molecular changes. But dynamical analysis provides a much richer language in which to discuss the possibilities. First, it specifies top-down behavioral constraint formation in terms of change in dynamical form, the change in form describing the causal power this novel constraint possesses. (This also distinguishes such effects as nonepiphenomenal.) Second, the dynamics itself shows how the constraint, a (relatively) macrolevel state/property, is determined by the states/properties of its microconstituents and so is supervenient on them, yet can nonetheless also constitute a constraint on them. Here dynamics gives the constraint a subtle status that eludes conventional formal analysis, combining what common philosophical assumption opposes. (See Refs [57] and [65].) Thus, dynamical determination, there being only one dynamical possibility for the collective dynamical state/property, cannot be equated with logical determination—the collective dynamical state/property is logically derivable from but can be expressed as a logical sum of its constituent states/properties. The former is specified as the constituents fixing all space–time trajectories so as to allow only one macropossibility, but these trajectories may be computationally strongly inaccessible, for example, through all critical point phase transitions. The neuroscientist Roger Sperry was among the early adopters of a top-down constraint model of mind emergence, see in later summary [66].

[12] For a systems biology illustration and discussion see Refs [67] and [68].

[13] Metaphysical aside. If there are unique, unchanging, spatiotemporally local, fundamental dynamical entities (e.g., chemical ions as biochemical atoms) then there is no fundamental emergence, only existential emergents having these entities as ultimate components in various dynamical compounds. But top-down constraint formation of itself does not require this. Fundamental nonlinear fields would yield the same emergent result and there are no such local components, while mutant spatiotemporally local fundamental components would issue in fundamental kind emergence.

understand cellular and multicellular organization. However, there remains a challenge for science, though not specifically for biological science, to find a full analytic mathematical treatment of the top-down formation process that permits a more rigorous and general discussion of when, where, and how it occurs, in biological systems in particular. The ''how'' is the difficult part.

While this seems the proper way to deal with emergence, it might be allowed that self-organization should be more broadly defined to capture simply the central idea that the resulting pattern is brought about through the interactions of the system components. The colloquial term ''organize,'' as in ''get organized,'' encourages this wide connotation. This position is permissible; all that then matters is that the definition of the term is clear, as Alice's Humpty Dumpty allowed. Under the wider usage self-organization is coextensive with organization (widely interpreted) but neither coincides with emergence, while under the narrower constraint-formation usage, self-organization coincides with emergence, but neither coincides with organization. As noted at the outset, there is no worthwhile definition to be had that sits between these two options. In my view, it leads to clearer, stronger, more scientifically useful conceptions of organization, self-organization, and emergence to adopt the latter usage. For instance, the formation of a crystal is a clear case of emergence, but not of any significant organization (see subsequently), yet it is a paradigm self-organizing process in the sense of top-down constraint formation.[14]

An immediate consequence worth noting is that self-*organization* need have little to do with organization proper. This is as it should be. Organization is a relational condition of systems where components play distinct roles but the roles are so interrelated as to produce a coherent global outcome. A simple illustration is found in the way the parts of a car engine are interrelated so as to deliver torque from fuel ignition; a profound example lies in intracellular organization. Self-organization is simply constraint formation and, as the case of crystallization shows, need not involve the emergence of any organization. Crystal formation is, rather, an instance of von Feurster's correctly named principle of order-(not organization)-from-noise (i.e., from random reassortment). von Feurster's own example of shaking coins down through successively smaller size filters orders them by size but does not organize them in any interesting sense. The unfortunately wide colloquial connotation of ''organize'' conflates order and organization, which are important to distinguish in understanding what is distinctive of living systems (see subsequently).

20.2.3 Organization and Complexity

Complex systems are complex, not only because they have many components, but fundamentally also because they are organized.[15] This raises two complementary

[14] For this position see Ref. [69].

[15] All of the further properties they may show—see 19.4.1—are forms of organization.

issues, the nature of organization as a form of complexity and the nature of the constraints that ensures overall functionality. Here we focus on the former issue and address the latter issue in the next section.

We begin with order, the basic relational notion of which organization is a special form. The root notion of complex order is that derived from algorithmic complexity theory: the complexity of the order in a pattern is measured by the length of its shortest, most compressed, complete description. A crystal lattice is simply ordered: it has a short compressed description given by fixing the locations of all ions as multiples of crystal plane distances away from any one reference ion. A gas, by contrast, has a very long minimal description and hence is maximally complexly ordered because its component molecules are all moving at random, so the position of each has to be separately specified. The crystal is highly ordered and the gas highly disordered. By contrast, an organized system is one where a number of distinct kinds of components playing unique roles nonetheless interrelate so that together they support one or more overall, global functions; a car engine and an organism are paradigm cases. The extremes of order are equally inhospitable to organization; a highly ordered system is too uniform, and a highly disordered system is too random, to support the variety of specific interrelationships required for organization. The relation of the piston rod movement in a car engine to that of the behavior of the fuel injector is very different from its relation to the exhaust muffler temperature, yet all combine to produce harmonious functioning. The variety in the relationships explains why too simple or high orderedness restricts organization, while the occurrence of the systematic interrelationships among the components explains why too complex or low orderedness equally restricts organization. Organization occurs in an intermediate "window" of ordered complexity between extremes. Thus, complex organization, as in living cells, is not straightforwardly complex in the sense of algorithmic complexity, but in some other sense.

Subtle, multiple different coordinations—that is, correlations—are required for complex organization if its very different component roles are to jointly serve a function. That is, it involves nested, higher order correlations of correlations. Very complexly organized systems, like cells, multicellular organisms, and cities are characterized by many layers or orders of correlations of correlations. Let us mean by a system's organizational depth roughly the number of nestings of subordering relations within it (cf. cells within organs within bodies within communities). Then the complexity of an organization is better measured by its organizational depth than it is by algorithmic complexity.

But it is still not a very satisfactory measure, primarily because it does not take into account the appearance of top-down constraints within nested systems, that is, it misses regulatory hierarchy and modularity. The class of all merely nested systems includes, but is much wider than, that of the organized systems, since organized systems must also sustain a global function. To achieve a global function an *organized* system exhibits a highest order global correlation expressing a global constraint (to performing its functions), with nested sets of lower order correlations within that, some of them modularized by lower order top-down constraints. We have as yet no way

to properly take all these features fully into account, and hence no satisfactory definition of organizational complexity.[16]

More importantly for science, we have as yet neither a real capacity to represent organization mathematically, nor a real capacity to investigate it experimentally. The mathematical framework for dynamical modeling in most of science, including systems and synthetic biology, is that of differential equations (d.e.s) as vector fields on differential manifolds, for example, on system phase space. But these modeling resources, powerful though they are for modeling the energetics of processes, do not explicitly describe the physical organization of the system—a metabolic cycle and a pendulum, for instance, may be modeled as equivalent dynamical oscillators. In a phase space only the global dynamical states and their time evolution are specified, not the organized processes that produce the dynamics; hence, it cannot capture organization. There is at present no obvious resolution to the general theoretical problem of how to incorporate organizational principles into dynamical models in a principled way. Correspondingly, the parameters we can measure are either component features—biochemical concentrations and the like—or higher order regulation parameters, such as respiration rate. There are no experimental techniques for detecting organization directly. Rather, it is reconstructed in retrospect after system relationships, in the genome for instance, have been reconstructed from what we can measure. Thus, a future challenge to systems and synthetic biology is to become more understanding of dynamical organization, both theoretically and experimentally.

20.2.4 Autonomy and Living Organization

The most basic global biological function is the regeneration of the body through metabolism, utilizing intakes of air, water, and food; for without this nothing else is possible. It is clearly a global function because it concerns the regeneration of the whole body. Autonomy, a form of recursive self-maintenance, is the name given to the global organizational constraint that must be met in order to support metabolic function. It is worth explicitly identifying autonomy because of its useful roles. For instance, it uniquely picks out the living systems from within the wider domain of complex, organized, nonlinear, dissipative (entropy increasing) and irreversible, chemical and biological systems, providing an unbiased, operational criterion of life hitherto missing and especially needed in exobiology. It also suffices to provide a naturalistic grounding for agency (see subsequently) and fruitfully frames the evolution of intelligence (see subsequently), thus also providing a framework for (organically) intelligent robotics. Let us explore the idea.

Finite systems sustaining dynamical equilibria far-from-(static)-equilibrium must do so by irreversibly taking in ordered or low entropy energy and material components from their environment and exporting it to material components carrying dissipated,

[16] Gell-Mann [70] discusses effective complexity and logical depth (see Ref. [71]) as other possibilities, but neither is satisfactory for various reasons he notices, but fundamentally because these too are general dynamical conceptions and do not directly include top-down constraints (for some further discussion see Ref. [69]).

less ordered, or higher entropy energy. These open systems must be organized: by the Morowitz theorem they must have at least one, and typically have many, closed-loop processes running within them.

For instance, a candle flame creates a thermodynamic asymmetry between itself and its environment, including an organizational asymmetry as it both preheats its own fuel supply (oil or wax) and creates a convection air current that delivers fresh oxygen to the flame. By supporting these two cyclical processes, the candle flame process contributes to the maintenance of the process temperature; in those partial respects, it is self-maintained (including of its self-maintenance capacity). But it has no self-regulatory capacity: should the flame die down, it does not cause more oxygen and wax vapor to flow in to revive it or cause a search to bring about delivery of other means to revive it, in contrast to hungry animals actively searching for food to revive themselves. The locus of regulation of these latter processes, if any, lies outside the flame process.

Living beings from single cells ''up'' are also among these open, irreversible, partially self-maintenance systems that maintain a state asymmetry with their environment. But unlike the candle they display a self-regulatory capacity that is extensive and active. Internally, as self-regenerating systems their cyclic processes must contribute to re-creating each other, that is, each process must partially regenerate the material constraints for themselves and/or others to work, requiring a highly organized web of cyclic process-constraint interdependencies.[17] Hence there must be strong mutual internal regulation of activity if internal coherence is to be maintained. Externally, organisms actively search for, and intake, requisite ordered energy and materials and excrete wastes, all the while avoiding or ameliorating damage. This requires active regulation of behavior. Even single cells regenerate themselves metabolically and partially regulate their environmental experience. Multicellular animals perform the same overall tasks, only with an expanded range of self-regulatory capacities, for both internal interaction (e.g., the cardiovascular resource delivery and waste removal system) and external interaction (e.g., neurally regulated sensory and neuromuscular motor systems, and so on) to match their expanded regenerative requirements.[18]

There are two broad cyclic processes involved in this activity, internal metabolic interaction and external environmental interaction, and these need to be coordinated: the environmental interaction cycle needs to deliver energy and material components to the organism in a usable form and at times and locations the metabolism requires to complete its regeneration cycles. The presence of these two thus synchronized cyclic processes resulting in system regeneration is the broadest functional sense of what is meant by a system's being autonomous. Though the detail, especially the dynamical boundaries and self-regulatory capacity, vary, this autonomy requirement picks out all and only living individuals—from cells, to multicellular organisms to various multi-organism communities, including many business firms, cities, and nations. In all

[17] These are what Kaufman [72] calls work-constraint cycles.

[18] They are models of self-regulation, including active self-maintenance of their self-maintenance capacities. Hence they are recursively self-maintained—see Ref. [73].

autonomous systems, the locus of living process regulation lies more wholly within them than in the environment—hence, the root sense of autonomy in the traditional sense.[19] Birds organize twigs to make nests, but twigs themselves have no tendency to organize nests or birds.

Autonomy is a subtle global constraint on the organization of interaction for whole organisms in their environmental context. In contrast to gases and crystals, dividing a cell in two typically does not produce two new cells because the fundamental global process organization that produces cell-type cohesion has been disrupted. Clearly, autonomy is an emergent property of the cell as a whole. In fact, emergence is a ubiquitous feature of the far-from-equilibrium systems. Comparing living systems to inanimate systems highlights the distinctive character of living interactive organization:

Comparative System Order			
Property	System Kind		
	Gas	Crystal	Cell
Internal bonds	None	Rigid, passive	Adaptive, active
Directive ordering[a]	Very weak, simple	Very strong, simple	Moderate, very complex
Constraints	None	Local	Global
Organization	None	None	Very high

[a]Directive ordering is spatiotemporally selective energy flow.

Entities are properly treated as genuine agents when they have a distinctive wholeness, individuality, and perspective on the world and their activities are self-regulated, normatively self-evaluated, willful, anticipative, and adaptive. Autonomous systems are inherently all of those things:

- *Self-Regulation.* We have already seen that autonomous systems are strongly self-regulated in both their internal and external interaction, making them the distinctive primary locus of their regulation. And because the self-regulation is in service of maintaining an internally coherent whole, they have a distinct, individual reference point for their activity that provides them a distinctive perspective on the world.
- *Normative Self-Evaluation.* Autonomous self-regeneration constitutes the fundamental basis for normative evaluation because it is the sine qua non and reference point for all else. Autonomy is the condition against which the outcomes of system processes are measured for success or failure. In single cells the measurement is simply continued existence or not. Multicellular systems have developed many internal, partial, and indirect surrogate indicators

[19] On autonomy see further Refs [51,54,64,75] and references therein. Self-governance lies at the core of our commonsense conception of autonomy. However, we are most familiar with the idea of autonomy as applied to persons and political governance, but these are sophisticated notions applied to sophisticated systems whose trappings may distract from fundamentals. We need to return to basic principles operating in all living systems to construct a naturalist notion that will ''grade up'' across the evolutionary sequence to our sophisticated concept.

for autonomy satisfaction and its impending violation, often based around closure conditions for their important subprocesses, for example, hunger (impending violation) and food satiation (satisfaction). It is these specific surrogate signals (cf. also thirst/fluid satiation, pain/pain-freeness) we think of as the basic, primitive norms guiding behavior, but they are literally grounded in turn in the obtaining of autonomy, from which they derive their normative character.

- *Willfulness.* A will is the capacity to do work (i.e., transform energy) in relation to the self whose will it is. The constitution of the autonomy constraint, which focuses directive organization on the generation of behavior to achieve self-regeneration, constitutes just such a distinctive capacity.
- *Anticipation.* To anticipate is to act now in relation to some future state, event, or process. Anticipation is thus an integral feature of autonomous systems because of their need to interact with their environment in ways achieving future closure outcomes that contribute to maintaining autonomy. The interactive relationship between the present action performed and the future, autonomy-evaluated outcome required is the most basic form of anticipation.[20] The willful performance of anticipative interactive activity against a normative evaluation criterion provides a root sense of action.
- *Adaptedness, Adaptiveness.* An organism is adapted when it possesses an autonomy-satisfying set of traits in its life environment. Conversely, an organism's ecological niche comprises the range of life environments for which its traits provide satisfaction of autonomy. An organism's adaptiveness is its capacity to alter its specific traits in mutually coordinated ways so as to adapt to, that is, satisfy autonomy in, a wider range of life environments than its current one.

20.2.4.1 Intelligence and Intentionality Agency of this kind provides an organizational platform for characterizing, and understanding the evolution of, intelligence and intentionality. There are three major aspects determining a system's anticipative capacities: the width of its interactive time window, the degree of articulation of the autonomy-related norms that it can use, and the high-order interactive relationships that it can effectively regulate. Between them, these features characterize the dimensions of intelligent/intentional capacity, and their roughly joint

[20] The root notion of anticipative action for Rosen [63] is that of a sequence of subactions that together achieve a closure condition and for which each subaction exists only because it is a member of that closure-achieving sequence. Each element then anticipates the next and the sequence anticipates the closure outcome. While this is too broad to provide any distinctively agency sense of anticipativeness, since any cyclically regenerating system (e.g., an autocatalytic polymer) counts as acting anticipatively, it does capture the central functional character of anticipation. Elementary systems like single cells will only exhibit action sequences where what anchors the repeated activation of the elements is just their belonging to a closure-achieving sequence. A distinctive agency sense of anticipativeness emerges when Rosen's root condition is applied to autonomous systems, since only these define a principled sense of it being the system itself that is anticipatory.

evolution traces the emergence of mind. And because of their preceding properties, autonomous systems can also be provided with action-centered informational and semantic characterizations, to complete the sense of agency. Organism information is modeled as reduction in downstream process regulation uncertainty. (''Shall I do A or B? Given the result of my last interaction, B is the thing to do.'') Organism semantics is that of the anticipated norm-referenced, autonomy-satisfaction provided by an action. These conceptions of information and semantics grade back to the actions of single cells, though the stronger the self-directed anticipative organization involved, the richer the semantic and informational structures sustained. In this context intentionality is conceived as a high-order regulatory capacity for fluid, meaningful goal-directed management of interaction. Intelligence and intentionality coevolve making use of a common self-regulatory apparatus. This avoids the common but implausible split between the two, respectively into problem solving and referential capacities.[21]

In sum, autonomy promises to provide the broad organizational framework from within which a fully naturalized conception of organisms can be developed in terms of the naturalistic intertwined emergences and mechanistic reductions that reveal their biochemical organizational depth. Of course, from a scientific point of view, the devil lies in providing the details. And the challenges in doing so are not only to do with coping with complications, but they also run deeper.

20.2.4.2 Challenges Posed by Autonomy Science, as discussed in Section 20.2.3, has only weak tools for studying organization. It has equally weak tools for studying global constraints, especially spatiotemporally extended global constraints like autonomy. These are at present not representable in the differential equation/phase space formalism. Although autonomy, like any dynamical constraint, must in principle be representable as a limitation on system accessibility to dynamical states (viz., constraint to those satisfying autonomy), there is at present no modeling methodology for constructing its constraint representation. So, while it is always possible to capture the dynamical consequences of internal organization by modeling system plus environment as a system of coupled component subsystems, there is no principled, internally motivated basis for reversing the process to extract organization from the dynamics, that is, for individuating the system in a principled way.[22] This is as much a challenge for theoretical robotics as for theoretical biology.

Dually, the challenge posed to practical construction and regulation/control in biology and robotics is equally deep because, if the account of autonomy (and of autonomy-based cognition) is even roughly correct, it provides a set of organizational requirements for this task that will prove far from simple to meet. For instance, despite using the label ''autonomous agent,'' there are at present no truly autonomous robots in

[21] This interaction-centered semantics is very different from, and more powerful than, standard direct referential semantics, for it captures directly the unlimited implicit possibility content in our action-differentiated grasp on reality. Bickhard argues that in this way it resolves the frame problem and is anyway ultimately the only coherent naturalist semantics, see for example Ref. [76]. Further see Refs [54] and [74] and Section 20.2.4.

[22] One thinks instinctively of the coupling of equations as the requisite tool, but so far as I am aware there is as yet no well-defined way to characterize either organization or globalness of constraints in these terms.

this organizational sense. Robotics uses a very limited formal notion of autonomy (something like invariant dynamical form) and limited performance criteria (typically confined to a single task) and an equally limited satisfaction method. This is as yet very far from even incorporating normative signals into the body coherence of robots, let alone the complexity required for self-regeneration and the capacity for fluid management of multidimensional environmental and internal interaction processes in relation to that (cf. Ref. [4], footnote 17). Similar constraints currently apply to our capacity to understand, much less synthesize, real biological systems. Despite calls for the simulation of biological autopoietic cells, we remain far from being capable of doing so.

Robert Rosen argued that living systems were not mechanical, that they could not be reduced to congeries of mechanisms (see Section 20.2.1), not simply because reduction in general failed, but for deeper structural reasons. Yet reduction to mechanisms is evidently what systems and synthetic biology aim to do. The gist of Rosen's objections (their final 1991 version [14] is couched in an arcane modeling language) is that holistic, organizational features like autonomy are central to being alive and these cannot be captured by analysis into mechanisms—indeed, our present general modeling tools must necessarily fail to adequately capture such features. He argued that these limitations, largely unrecognized and unexamined, represented a powerful limitation on the development of biological science. Cloning is hailed still, even while the profession knows that, though a technical feat, it is limited to intercellular nuclear transfer, and the entire cytoplasmic apparatus of the globally coherent regenerative cell is simply ignored.

There is some point to Rosen's line of objection. Metabolic regeneration is central, does exhibit autonomous organization, and currently cannot be adequately modeled dynamically. The emergence of high-order global functional coherence expressed in adaptive intelligence offers another version of this challenge. Rosen argues that this difficulty is made more pointed by the fact that often the components in metabolism are only thus because of the character of the whole (cf. Rosen on anticipation, footnote 20). This seems to make it impossible to understand such systems without postulating the global dynamical organization at the outset, stymieing attempts to synthesize the organization from its components. If, in addition, the components are formed during the self-organization of the whole process, then the argument is reinforced.

Especially these last cases are real challenges to substantive biological theory. However, the scientists involved might argue that new tools to understand them are being developed, albeit slowly, and this shows that they should be recognized as so many methodological challenges rather than overwhelming *a priori* demonstrations of the separation of biology from natural science. Hogeweg, for example, has pioneered the use of computational models to understand the ways in which spatial segregation processes can lead to the survival of entities, whether molecules, viruses, organisms, or even prebiotic entities, where an unsegregated model would predict their extinction, and also illuminate multilevel selection and evolution processes. Hogeweg employs cellular automata (CA) models to capture the spatial organization necessary to explain the outcome. This is possible because CAs are inherently and explicitly relationally

organized (though in a generalized, not narrowly spatial, way), even while able to incorporate some aspects of local dynamical interaction. Dynamic networks are similarly inherently and explicitly relationally organized, while currently they are largely used to express functional relationships, they too can be adapted to express spatial relationships. In ecology, for example, there is increasing attention to modeling spatial organization and these kinds of modeling tools can be used to model intracellular spatial relationships.[23] Even so, there is no inherent capacity in any of these tools to represent either organization per se or globalness of constraints. But it may be that in future, as need and capacity to model spatial organization grows, more and powerful such tools will alleviate these problems.

20.2.5 Condition-Dependent Laws and the Unity of Science

Scientists in systems and synthetic biology often regard their approach more as ''model building'' than as ''theory'' or ''law'' centered; this is understandable in a domain where nearly every variation results in differing functional capacities and behavioral patterns. Compared to the grand universal, invariant laws of physics, these local idiosyncratic behavioral patterns do not count as laws; so, especially when most biological systems are yet too complex to predict, it is more useful to simply model each system and try to understand it on that basis. But of course biologists do use laws in constructing their models, the laws of (bio-)chemistry; if these did not operate the same-everywhere biology would be much harder than it already is. Even so, the complication arises from the fact that the operational invariance largely occurs at the ion–ion interaction level. How n-body, k-component ion systems operate is often a strong and sensitive function of the initial and boundary conditions, especially organizational conditions, obtaining and that is why no simple set of laws can be deduced in advance. Indeed, self-organization precisely occurs because of the sensitivity of dynamical form to dynamical initial and boundary conditions (see Section 20.2.2).

But the last equally provides license to extend the notion of law to such cases. For since self-organization involves a new dynamical form, it is reasonable to say that it obeys new dynamical laws characteristic of that form. Moreover, the idea that true laws have to be specified independently of any initial and boundary conditions is a conceit of physics, and perhaps ultimately not true there either considering that even fundamental laws evidently changed form as the big bang cosmos cooled. But once that independence requirement is dropped we are free to see biology as replete with real behavioral laws, it is just that they will be condition-dependent, or ''special'' (as some philosophers say).[24] For instance, condition—a cooling mould of liquid iron in contact with a heat reservoir of lower temperature, emergent laws—rigid body (not fluid) dynamics, crystalline (not fluid) conduction of electricity, heat, and

[23] See, for example, Hogeweg [77–79]. On spatial modeling in ecology see, for example, Refs [80] and [81] and references therein.

[24] See Ref. [82] for an early insight of this kind. We now see that this condition is not unique to life, for instance, it characterizes at least all dynamics that shows self-organization.

sound. Put that way, condition-dependent laws are commonplace even in physics, and certainly throughout all the other sciences. It is just that condition-dependent laws are often on that account hard to predict or use for prediction, but that is a different, epistemic issue.

Why not push condition-dependence further to include every instance of change in initial and/or boundary conditions? For instance, the specific force of gravity changes between the Sun–Saturn and Sun–Earth subsystems because of the changing masses involved. Why not claim all these as equally condition-dependent laws? Well, because it is the same general law that is involved; the diverse cases are unified by a single lawful interaction form. This is not so for the iron bar and other cases involving self-organization. However, surely the self-organization cases are equally a consequence of the underlying universal dynamics, and simply produced under specific initial and/or boundary conditions; if it is just that at present we cannot analytically represent self-organization then that should not stop us from allying them to the previous simpler cases. This is so, but there are two important differences marking off the self-organization cases: (i) it heralds the presence of an irreducibly new dynamical existent, (ii) the dynamical form itself alters accordingly, so there is no common universal law form. Thus, they represent a genuinely interesting set of conditions.

However, there are also interesting mid-way cases. Self-organization through something as radical as phase change is not the only way to induce the formation of a new constraint condition; inducing a Hopf bifurcation of the dynamics (where a smooth parameter change alters the dynamic attractor landscape) is another, as is simply being moved from one local energy well to another in an unchanged dynamical landscape of a system. In each case changed initial and/or boundary conditions lead to changed dynamical laws. Although the well-shift cases may be set aside on the grounds that they too are unified by a common dynamics (represented by the landscape), the former Hopf-bifurcation cases also manifest a changed dynamical form and deserve to belong to the self-organization cases, as do other kinds of dynamical bifurcations. Polanyi once argued, in effect, that what was distinctive of living systems was that their governing laws were so strongly dependent on initial and/or boundary conditions [32]. Polanyi had in mind at least the way that information can alter the basis of behavior in living systems. If the impact of an information-conveying signal on an organism is dynamically equivalent to a Hopf or other bifurcation, then Polanyi's living systems can all be brought under the same dynamical paradigm.

The same considerations apply to dynamical models of the genome. There may be a wide variety of dynamically different forms that a genome can take up as various of its processes alter its own initial and/or boundary conditions so as to induce a dynamical bifurcation, for example, create and insert a new catalyst into the protein dynamics, thus forming special laws for that condition. These effects can propagate historically. The emergence of a new constraint with new dynamics may lead to the subsequent dynamical formation of still further top-down constraints, and so new entities, that would not have been dynamically possible without that preceding formation event. Indeed, something like that must be the overall dynamical form of development. Moreover, this cascade of dynamical consequences is marked by its initiating formation event and thus exhibits dynamical fixation of (these) historical constraints.

Such path-dependent dynamics occur throughout ecology (e.g., somatic and niche symbioses), economics (e.g., in off-highway development and technological learning), and the social sciences (fashion, and so on), but are not unknown in physics, for example, in hysteresis.

Needless to say, biology will not splinter into an unprincipled disunity under these complex dynamics.[25] Once again, these bifurcations will still be dynamically determined by, and identified in terms of, their dynamical constituents and governed by laws that themselves are thus grounded in the underlying universal dynamics. This is precisely what the biochemistry of the dynamical network models is meant to show. It will also encompass the many changes that consist of less profound dynamical transitions falling under the same dynamical form (even shifting between strange and other attractors). And the requirement to "match-up" the dynamics of different spatiotemporal scales and domains provides a further important unifying component. For example, unifying molecular chemistry and cellular biology requires interpreting cellular processes in biochemical terms that immediately generate many penetrating tests because of the requirement to match up the two descriptions—for instance all of the function to process reductions.

All this provides a shared dynamical framework interconnecting emergent variety in intimate ways that make it possible to successfully model complex genome dynamics and even development, navigating through the complex world of emergent but interconnected cellular and intercellular levels and laws. This gives a strong sense in which biological science remains unified even while acknowledging more strongly initial and/or boundary condition-dependent laws than simple physics and chemistry was wont to consider. The challenge to biological science is to recognize explicitly and better understand this plethora of law types and shifts, so as to make explicit their basis and their theoretical and methodological implications.

20.2.6 Limits of Knowability

The advent of complex systems models introduces new considerations concerning the manner and limits of scientific knowability. By this it is not meant the pragmatic fact of vastly more complex systems generating vastly more extensive sets of data than can practically be managed (cf. Section 20.1.2). Rather, the interest here is in principled limits on knowledge. Discussion is limited to knowledge of complex systems and is even so preliminary.[26]

An immediate consideration is the limit on analytic solvability to achieve "closed form" symbolic representation of dynamics, that is, a single formula giving the

[25] See, for example, Ref. [83], and for the complex dynamical unity response presented here see Ref. [57], footnote 4, and Section 20.2.5. It should be added that the conception of laws as simple universal generalizations, common among philosophers and scientists alike, is simplistic, science shows a far more complex and rich spectrum of laws—see Ref. [84].

[26] For more fundamental limits on knowability deriving from quantum theory, see for example Ref. [70], and for something of the variety of forms knowledge limits can take, see Ref. [85].

universal solution to a set of dynamical equations.[27] But as we move beyond simple sets of independent linear differential equations toward nonlinear, partial differential equations in interdependent coupled equation sets, we find that the dynamics they represent rapidly becomes very complex and the equation sets lack analytic solutions. Beyond this again lie bifurcations; these have no analytic representation within which their dynamics is exhibited, as the standard dynamical systems do, and cannot have one in standard dynamical terms precisely because they change their dynamical form, that is, change their dynamical representation, and as a function of their own initial/boundary conditions. In all these cases, as noted in Chapter 19 (see footnote 36), it is then necessary to explore their dynamics through numerical approximation and temporal iteration. Their dynamics is exhibited in extended form in space and time, rather than being condensed into a single abstract relation among symbols. This places computational modeling at the center of their scientific investigation in a strong manner and highlights the huge, and unique, contribution of computers to scientific knowledge.

However, it should not be forgotten that in most cases computational modeling provides only a numerical approximation, not exact values. Again, in most cases this is not a problem since the degree of approximation can be increased at will. But mathematical science contains many noncomputable functions,[28] that is, functions where information crucial to identifying it is lost at any level of finite approximation. Many superposition or ''wave'' phenomena (classical and quantum) are of this kind where wavelet information at indefinitely small scales is important to identifying the whole function. A comparable situation occurs when chaos (a ''strange'' attractor) is involved. Because nearby chaotic trajectories diverge exponentially from one another (at all points along their trajectories), any approximation will be invalidated by some trajectories within the approximation range—often quickly, one of the earliest discoveries of chaos (by Lorenz, using a coupled triad of partial differential equations) concerned just such divergence brought about from slightly different rounding errors. Thus, though computational numerical approximation represents a huge expansion of our capacity to know complex dynamics, it also represents a selective, but important, diminution in our knowledge capacity.

The exponential divergence of dynamical trajectories characteristic of chaotic attractors manifests sensitivity to initial conditions. Small differences in the conditions determining the initial dynamical state are eventually amplified into large divergences. This can happen with nonlinear dynamics generally, it does not require chaos; bifurcations are examples. In such circumstances, prediction is limited by the accuracy of knowledge of the initial system state, that is, of the initial conditions. This is so even

[27] Curiously, this is equivalent to science constructing a compressed symbolic description of reality, in the sense of algorithmic complexity theory. Could the latter's difficulties with defining organization be reflected in some characteristic of the former? And what has this to do with Rosen's [64] more abstract concerns with modeling?

[28] That is, mathematical functions, not biological functions. Mathematical functions are many: one maps from a domain to a range, hence unique on the range. One distinctive merit of the proposal to model biological functions as input/output maps is that this relates them directly to mathematical functions and hence, via modeling, to dynamical maps and so to biochemical processes.

though the system dynamics is deterministic, and so exactly one precise trajectory happens. Since all human knowledge (indeed all creaturely knowledge) has finite resolution, prediction is permanently parted from determinism. Self-organization ensures that this extends to prediction of condition-dependent laws—in particular to predicting self-organized intracellular dynamics.

Limits on predicting the behavior of intelligent agents provides a further class of special cases. Even simple sensory agents can on occasion amplify very small signals (perhaps a few light quanta) into large behavioral differences, so that even smaller uncertainties in those signals or in the internal state created will place limits on predicting behavior. Even where these details are knowable in principle another limit typically bites hard: such complex systems have considerable logical depth and for such systems the time required to make a prediction is in principle large,[29] added to which is the time required to obtain all the relevant state and process information to do so. Often, those times are much longer than the time horizon for relevant prediction and action, whence one's interaction with them is always on the basis of some uncertainty. Slow, long-running "agendas" in human personal development can produce surprising behaviors that defeat even decades of contrary data about a person. Another version of the same limit applies when an agent alters its own environment on too fast a timescale for it to know the consequences of its past actions before it acts again. Humans have always been in this predicament and continue notoriously to be so, as witness climate change, peak oil, nuclear proliferation, stability of financial markets, and so on.

Finally, an important methodological issue has recently opened up concerning the most effective statistical means of extracting knowledge of genome organization and dynamics from the large data sets generated by contemporary high-throughput experimental technologies, data often sparsely distributed in large-dimensional parameter spaces. A classic paper by Breiman [36] opposes two approaches to the data: model learning and machine learning. In model learning, a class of mathematical models specified by parameter values is chosen as a presumed model of the underlying reality from which the data is taken and its parameters are interpreted in terms of the entities and potential dynamical processes thought to constitute the underlying reality. The problem for statistical methodology is then to use the data in an unbiased way to estimate the parameter values and so fix the particular model involved. This model can subsequently be tested by its prediction of new data and the parameter values re-estimated as required.

In machine learning, by contrast, no model is specified, rather the data are used to "tune" a machine-learning process (some one of a large class of convergent mathematical adaptation or self-correction processes, for example, neural nets, run on computers). The tuned machine is then used to predict further data and is tested against it. The tuned machine state may have no obvious understanding in physical model terms; indeed its state dimensions, hence parameters, emerge from the tuning process and may be very large. Nonetheless, in a variety of situations it provides superior predictive performance and, with modeling goodness-of-fit tests often too weak to select among a variety of models, it emphasizes prediction as a self-sufficient goal of

[29] For discussion see Refs [69] and [70].

science. This threatens to pull apart prediction from ontological-dynamical understanding as epistemic goals of science and thus, represents a distinctive constraint on scientific knowledge. Where it provides intelligible insight into underlying processes, it is essentially a form of induction from data, just as the previous model learning approach can be considered a form of hypothetico-deductive falsification, linking these alternatives to a much more general and venerable debate about scientific methodology (empiricism versus Popperianism, and so on). There is a lively version of the debate in systems biology [37,38].

To the philosophically minded the machine-learning language may, however, suggest much more, specifically the prospect of either (1) a new phenomenological empiricism, where data are again uncritically glorified and last century's well-abandoned extreme claims of solely reconstructing theory from it (cf. behaviorism) reemerge, or (2) a kind of postmodern antirealism in which scientific investigations each use their own separate machine-learning states, each state employed while ever it is predictively adequate, with discussion of underlying reality considered a sign of nostalgia for grand schemes, implicit attempts at ideological hegemony or mental confusion. This is not the place to discuss either the foundations of statistical method or the fate of grand conceptions. Rather, abandoning the extremes represented by 1 and 2 above, I want to briefly suggest consideration of a middle-ground approach to method recognizing the utility of both induction and hypothetico-deduction in context.

It is surely a false dichotomy to oppose prediction to understanding, since each is necessary to the other: understanding without prediction is ignorant and uncritical, prediction without understanding is weak and fragmented. The former is obvious for any finite, comprehensively fallible species like us commencing research in ignorance. The outcomes of predictive tests underpin acceptance/rejection of any proposed models and hence of improved understanding. The latter rests on the way a confirmed dynamical model can direct research much more effectively than simply trying to collect more data per se. For instance, such models distinguish law-like relations (as energy transform processes) from mere correlations or noisier relations, and also identify the sources of noise and bias, including in the interaction between system and data-gathering instruments—all of which structure future-testing regimes and assessment regimes, including the filtering and correction of data itself. And as before, model matching across scales and domains widens and focuses this role.

In addition, the machine-learning approach still relies on the choice of data categories, experimental setup, and appropriate instruments and probes to generate its data. But all such choices make presumptions about the character and salience of features in the underlying reality. For instance, instruments have systematic errors and limitations and unless we have sound insight into how the instruments work we cannot know what their defects are and hence how to process data. (A striking demonstration of this comes from the quantum theory proof that even core classical measuring devices have inherent error rates, of which science had been entirely unsuspecting.) Instruments themselves are understood through empirically validated theoretical models.[30]

[30] Often enough using the very theories they are used to test. But this is largely OK, see Ref. [89]. On data choices, see, for example, Ref. [90].

By comparison, in all these cases machine learning can only combine the data pools without direction, there being no methods within data lists alone for simulating dynamical discrimination and unification, systematic data errors and data limitations. Even identifying random data errors may cause problems here, since these have somehow to be distinguished from inherent dynamical fluctuations in the system, the latter behaving as noise except near bifurcations where their form may be critical to understanding system dynamics. All this leads to insatiable demands for sufficient data, ultimately extending to encompass all science and the entire universe as a block—not a good theory of epistemic (learning) strategy for finite agents beginning in ignorance.

On the other hand, machine learning often finds patterns in high-dimensional data where our knowledge of models is initially poor and complex dynamical process lie behind the data. All of which suggests that a pragmatic mixed strategy is called for, reinforced by the many approaches in use that combined parametric and nonparametric modeling. If you know nothing about the domain but have enough data (data rich, hypothesis poor), then machine learning may be the best approach, while if you know a lot about the domain then, especially if only a small range of data is available (hypothesis rich, data poor), model learning is surely the best bet. And in between, knowledge-wise and data-wise, the features of the best-mixed model will no doubt vary complexly with context.

REFERENCES

1. Storz G, Altuvia S, Wassarmann KM. An abundance of RNA regulators. *Annu Rev Biochem* 2005;74:199–217.
2. Storz G. An expending universe of noncoding RNAs. *Science* 2002;296:1260–1263.
3. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001;294(5543):853–858.
4. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001;294(5543):858–862.
5. Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001;294:862–864.
6. Mourelatos ZZ, Dostie J, Paushkin S, Sharma M, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* 2002;16(6):720–728.
7. Ruvkun G. Molecular biology: glimpses of a tiny RNA world. *Science* 2001;294:797–799.
8. Grosshans H, Slack FJ. Micro-RNAs: small is plentiful. *J Cell Biol* 2002;156(1):17–22.
9. Wassarman KM, Zhang A, Storz G. Small RNAs in *Escherichia coli*. *Trends Microbiol* 1999;7(10):37–45.
10. Altuvia S, Wagner EGH, Switching on and off with RNA. *Proc Natl Acad Sci USA* 2000;97:9824–9826.
11. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002;415(6873):810–813.
12. Erdmann VA, Szymanski M, Hochberg A, de Groot N, Barciszewski J. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res* 2000;28(1):197–200.

13. Avner P, Heard E. X-chromosome inactivation: counting, choice and initiation. *Nature Rev Genet* 2001;2:59–67.
14. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Gen* 2001;2:919–929.
15. Hüttenhofer A, et al. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J* 2001;20:2943–2953.
16. Argaman L, et al. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 2001;11:941–950.
17. Rivas E, Klein RJ, Jones TA, Eddy SR. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 2001;11:1369–1373.
18. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 2001;15:1637–1651.
19. Jeffares DC, Poole AM, Penny D, Pre-rRNA processing and the path from the RNA world. *Trends Biochem Sci* 1995;20:298–299.
20. Poole AM, Jeffares DC, Penny D. The path from the RNA world. *J Mol Evol* 1998;46:1–17.
21. Toyoda T, Wada A. Omic space: coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* 2004;vol. 20:1759–1765.
22. Arakawa K, Mori K, Ikeda K, et al. G-language genome analysis environment: a workbench for nucleotide sequence data mining. *Bioinformatics* 2003;19:305–306.
23. Hucka M, Finney A, Sauro HM, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;19:524–531.
24. Chen Y, Bittner ML, Dougherty ER. Issues associated with microarray data analysis and integration (supplementary information to article by Bittner, M, Trent, J, and Meltzer, P.) *Nat Genet* 1999;22:213–215.
25. Luisi PL, Oberholzer T, Lazcano A. The notion of a DNA minimal cell: a general discourse and some guidelines for an experimental approach. *Helv Chim Acta* 2002;85:1759–1777.
26. Koonin EV. How many genes can make a cell: the minimal-gene set concept. *Annu Rev Genomics Hum Genet* 2000;1:99–116.
27. Hutchison CA III, Peterson SN, Gill SR, Cline RT. Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 1999;286:2165–2169.
28. Ji Y, Zhang B, von Horn SF, Warren P, et al. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001;293:2266–2269.
29. Chandonia J-M, Konerding DE, Allen DG. Computational structural genomics of a complete minimal organism. *Genome Inform* 2002;13:390–391.
30. Zhang R, Ou H-Y, Zhang CT. DEG: a database of essential genes. *Nucleic Acids Res* 2004;32:D271–D272.
31. Glass JI, et al. Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* 2006;103 (2):425–430.
32. Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 1996;93:10268–10273.
33. Arigoni F, Talabot F, Peitsch M, Edgerton MD, Meldrum E. A genome-based approach for the identification of essential bacterial genes. *Nat Biotechnol* 1998;16:851–856.

34. Salisbury MW. Get ready for synthetic biology. *Genome Technol* 2006;26–33.
35. Gray MW. Evolution of organellar genomes. *Curr Opin Genet Dev* 1999;9:678–687.
36. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson A-S, Winkler HH, Kurland CG. The genome sequences of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 1998;396:133–140.
37. Müller M, Martin W. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *Bioessays* 1999;21(5):377–381.
38. Tielens AGM, van Hellemond JJ. The electron transport chain in anaerobically functioning eukaryotes. *Biochim Biophys Acta* 1998;1365:71–78.
39. Tielens AGM. Energy generation in parasitic Helminths. *Parasitol Today* 1994;10: 346–352.
40. Rotte C, Henze K, Müller M, Martin W. Origins of hydrogenosomes and mitochondria. *Curr Opin Microbiol* 3(5):481–486.
41. Embley TM, Martin W. A hydrogen-producing mitochondrion. *Nature* 1998;398:517–518.
42. Akhmanova A, Voncken F, val Alen T, van Hoek A, Boxma B, Vogels G, Veenhuis M, Hackstein JHP. A hydrogenosome with a genome. *Nature* 1998;396:528–529.
43. Embley TM, van der Giezen M, Horner DS, Dyal PL, Bell S, Foster PG. Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* 2003;55(7):387–395.
44. Keeling PJ. A kingdom's progress: Archezoa and the origin of eukaryotes. *Bioessays* 1998;20:87–95.
45. Cavalier-Smith T. Eukaryotes with no mitochondria. *Nature* 1987;326:332–333.
46. Clark CG, Roger AJ. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc Natl Acad Sci USA* 1995;92:6518–6521.
47. Pfanner N, Geissler A. Versatility of the mitochondrial protein import machinery. *Nat Rev Mol Cell Biol* 2001;2:339–349.
48. Mourier T, Hansen AJ, Willerslev E, Arctander P. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol Biol Evol* 2001;18 (9):1833–1837.
49. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 2003;29:380–395.
50. Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancis A, Delgadillo-Correa M, Johnson PJ, Muller, Miklos T. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* 2004;101 (28):10368–10373.
51. Hooker C, *Reason, Regulation and Realism*. Albany, NY: SUNY Press, 1995.
52. Shi Y. *An Economic Theory of Knowledge*. London: Elgar, 2001.
53. Hooker C, Towards a general theory of reduction. Dialogue XX, Part I: Historical framework, pp. 38–59; Part II: Identity and reduction, pp. 201–236; Part III; Cross-categorical reduction, pp. 496–529; 1981.
54. Christensen W, Hooker C. Self-directed agents. In: MacIntosh J, editor. *Can J Philos* (Naturalism, Evolution and Intentionality, Special Supplementary, Ottawa, Canada) 2002;27: 19–52.
55. Churchland P. *Scientific Realism and the Plasticity of Mind*. London: Cambridge University Press, 1979.

56. Hooker C. Reduction as cognitive strategy. In: Keeley B, editor. *Paul Churchland*. New York: Cambridge University Press, 2005.
57. Hooker C. Asymptotics, reduction and emergence. *Br J Philos Sci* 2004;55:435–479.
58. Nagel E. *The Structure of Science*. London: Routledge & Kegan Paul, 1961.
59. Bechtel W. *Discovering Cell Mechanisms: The Creation of Modern Cell Biology.* Cambridge, UK: Cambridge University Press, 2006.
60. Bechtel W, Abrahamsen A. Explanation: a mechanistic alternative. *Stud Hist Philos Biol Biomed Sci* 2005;36:421–441.
61. Bechtel W. Biological mechanisms: organised to maintain autonomy. In: Boogerd F, Bruggeman F, Hofmeyr J-H, Westerhoff J, editors. *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier, 2007.
62. Dewan E. Consciousness as an emergent causal agent in the context of control system theory. In: Globus G, Maxwell G, Savodnik I, editors. *Consciousness and the Brain*. New York: Plenum, 1976.
63. Rosen R. *Anticipatory Systems*. New York: Pergamon, 1985.
64. Rosen R. *Life Itself*. New York: Columbia University Press, 1991.
65. Collier J. Supervenience and reduction in biological hierarchies. In: Matthen M, Linsky B, editors. *Can J Philos*, (Philosophy and Biology, Supplementary Volume, Ottawa) 1988;14: 209–234.
66. Sperry R, *Science and Moral Priority: Merging Mind, Brain and Human Values*. New York: Columbia, 1983.
67. Bruggeman FJ, Westerhoff HV, Boogerd FC. BioComplexity: a pluralist research strategy is necessary for a mechanistic explanation of the 'live' state. *Philos Psychol* 2002;15(4): 411–440.
68. Boogerd FC, Bruggeman FJ, Richardson RC, Stephan A, Westerhoff HV. Emergence and its place in nature: a case study of biochemical networks. *Synthese* 2005;145:131–164.
69. Collier J, Hooker C. Complexly organised dynamical systems. *Open Syst Inf Dyn* 1999;6:241–302.
70. Gell-Mann M. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. New York: Henry Holt, 1994.
71. Bennett C. Dissipation, information, computational complexity and the definition of organization. In: Pines D, editor. *Emerging Syntheses in Science, Proceedings of the founding workshops of the Santa Fe Institute. Redwood California. Addison-Wesley*, 1985.
72. Kaufman S. *Investigations*. New York: Oxford University Press, 2000.
73. Bickhard M. Representational content in humans and machines. *J Exp Theor Artif Int* 1993;5:285–333.
74. Christensen W, Hooker C. An interactivist-constructivist approach to intelligence: self-directed anticipative learning. *Philos Psychol* 2000;13:5–45.
75. Moreno A. A systemic approach to the origin of biological organization. In: Boogerd F, Bruggeman F, Hofmeyr J-H, Westerhoff J, editors. *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier, 2007.
76. Bickhard M, Terveen L. *Foundational Issues in Artificial Intelligence and Cognitive Science—Impasse and Solution*. Amsterdam: Elsevier Scientific, 1995.
77. Hogeweg P. Computing an organism: on the interface between informatic and dynamic processes. *Biosystems* 2002;64:97–109.

78. Hogeweg P. Multilevel processes in evolution and development: computational models and biological insights. In: Lässig M, Valleriani A, editors. *Biological Evolution and Statistical Physics*, Springer lecture notes in physics 585. Berlin: Springer Verlag, 2002;pp. 217–239.
79. Hogeweg P, Takeuchi N. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Orig Life Evol Biosph* 2003;33:375–403.
80. Dieckmann U, Law R, Metz J, editors. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge, UK: Cambridge University Press, 2000.
81. Cash D, editor. Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World. Available at http://www.ecologyandsociety.org/.
82. Polanyi M. Life's irreducible structure. *Science* 1968;160:1308–1312.
83. Dupré J. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press, 1993.
84. Hooker C, Laws, natural. In: Craig E, editor. *Routledge Encyclopedia of Philosophy*. London: Routledge, 1998.
85. McDaniel R, Driebe D, editors. *Uncertainty and Surprise in Complex Systems*. Berlin: Springer, 2005.
86. Breiman L. Statistical modelling: the two cultures. *Stat Sci* 2001;16:199–215.
87. Kell D, Knowles J. The role of modelling in systems biology. In: Szallasi Z, Sterling J, Periwal V, editors. *System Modelling in Cellular Biology*. Cambridge, MA: MIT Press, 2006.
88. Westerhoff H, Kell D. The methodologies of systems biology. In: Boogerd F, Bruggeman F, Hofmeyr J-H, Westerhoff J, editors. *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier, 2007.
89. Hooker C. Global theories. *Philos Sci* 1975;42:152–179. (Reprinted in Hooker, C., *A Realistic Theory of Science*, Albany, NY, SUNY Press, 1987).
90. Lewontin R, Levins R. Let the numbers speak. *Int J Health Services* 2000;30(4): 873–877.