Fridolin Gross

The Sum of the Parts: Heuristic Strategies in Systems Biology

Ph.D. Thesis in
Foundations of the Life Sciences and Their Ethical Consequences
(FOLSATEC)

Matricola n. R08430
Settore Disciplinare: FIL/02

Submitted on January 7, 2013

Supervisor:
Prof. Dr. Mark A. Bedau
Professor of Philosophy and Humanities, Reed College

External Co-Supervisor:
Prof. Dr. William C. Wimsatt
Professor Emeritus of Philosophy, University of Chicago

Lab Supervisor:
Dr. Andrea Ciliberto
Firc Institute for Molecular Oncology (IFOM)

European School for Molecular Medicine (SEMM) and the University of Milan, Italy

CONTENTS

LIST OF FIGURES

ABSTRACT

This thesis addresses philosophical issues regarding the young field of systems biology. Systems biologists commonly present their approach as a superior alternative to 'traditional' molecular biology that they describe as being overly 'reductionist.' However, the heterogeneity of systems approaches makes it difficult to understand what 'the' approach of systems biology exactly consists in.

Here I propose a framework for the systematic comparison of different scientific approaches in biology. I argue that the relevant issues arise at the level of strategies of mechanistic discovery. These strategies are best understood as 'heuristic,' that is, as tools to reduce the complexity of a given research task. While having the virtue of making the search for mechanisms more efficient, heuristic strategies rely on particular assumptions about the system under study. This can introduce bias and lead biologists to underestimate the actual complexity of the system. Framing the analysis in terms of heuristic strategies provides a precise way to distinguish between different approaches and to better understand the ongoing rhetoric battles.

I discuss a number of case studies, both from molecular biology and from systems biology. I argue that the traditional approach of molecular biology relies on a relatively well-defined set of heuristics that corresponds to a particular idea of the organization and complexity of living systems. Approaches in systems biology relax some of the underlying assumptions of the traditional approach, notably by applying tools of mathematical modeling, but they have to make use of alternative heuristics in order to be efficient. As a result, they rely on different assumptions about organization and complexity.

My detailed discussion of case studies reveals that there are a number of different systems approaches that can be distinguished by analyzing their heuristic character. The ambition of systems biologists to build formal models of biological mechanisms, however, has the virtue of making many of the underlying assumptions explicit which helps to recognize and reduce bias, and moreover facilitates the integration of different approaches.

Some of the issues touched upon also have relevance for more general questions in the philosophy of biology. Assumptions about biological organization and complexity can heavily influence what we think of as a good scientific explanation. Since systems biology puts into question some of these assumptions, we might be forced to revise our ideas about mechanistic explanation. I argue that notably the concept of biological robustness has to be taken into account by philosophers who are thinking about mechanisms in biology.

INTRODUCTION

The recent popularity of systems approaches in molecular biology is perhaps best understood as a reaction to technological developments beginning in the 1990s, notably the large sequencing projects such as the Human Genome Project. Technological advances are usually considered to be important drivers of scientific progress. The light microscope revolutionized the study of living structures in the seventeenth century; the steam engine had considerable influence on the development of thermodynamics in the nineteenth century; and more recently, the construction of fast computers has given a boost to almost every scientific discipline. Obviously, countless other examples could be named. However, the recent technological developments in molecular biology are often perceived almost as a threat, or at least as a big challenge for scientists. The tools of genomics in particular, the large sequencing projects and their successors, are described as overwhelming biologists by producing large and unmanageable amounts of data. Very often, the assessments of the current situation in molecular biology are rendered more dramatic with the help of aqueous metaphors (speaking, for instance, of a *flood*, *deluge*, *spate*, *shock wave*, or *tsunami* of data), and typically they conclude by expressing the need for radical change at the conceptual level. Systems biology is often presented as an alternative and superior way of doing biology:

> Perhaps the most important consequence of the Human Genome Project is that it is pushing scientists toward a new view of biology—what we call the systems approach. Systems biology does not investigate individual genes or proteins one at a time, as has been the highly successful mode of biology for

> the past 30 years. Rather, it investigates the behavior and relationships of all of the elements in a particular biological system while it is functioning. These data can then be integrated, graphically displayed, and ultimately modeled computationally. (Ideker et al. 2001, 343)

Many advocates of systems biology consider the traditional approach of molecular biology, which during the second half of the twentieth century proved successful at explaining some of the fundamental mechanisms of life, inadequate to respond to the challenges of the so-called 'postgenomic era' (Winnacker 1997).[1]

Most commonly, this inadequacy is traced back to the allegedly 'reductionist' spirit of molecular biology:

> Without question, the reductionist pursuit of molecular biology has been a tremendous success story. Systems biology today would not be possible without the tools and knowledge that the reductionistic approach to identifying system components has provided. But it is not always possible to understand the behavior of a complex system simply by scaling up the properties of its individual parts. (Levesque and Benfey 2004, R179)

> Biological systems are extremely complex and have emergent properties that cannot be explained, or even predicted, by studying their individual parts. The reductionist approach—although successful in the early days of molecular biology—underestimates this complexity and therefore has an increasingly detrimental influence on many areas of biomedical research, including drug discovery and vaccine development. (van Regenmortel 2004, 1016)

> Molecular biology requires a certain way of thinking. It is about the naming and behaviour of the parts. We reduce each whole to its component parts and define them exhaustively. Biologists are now perfectly used to that thinking and the interested lay public has caught up, too. So we are now ready to move on. Systems biology is where we are moving to. Only, it requires a different

[1] Throughout this thesis, the term "molecular biology" is understood in a rather broad sense, without clearly delineating it from disciplines like cell biology, immunology, etc. This is in line with the usage of many systems biologists, but not necessarily of biologists in general. I thank Francesca Ciccarelli for pointing this out to me. For more details, see my discussion in Chapter 2.

> mind-set. It is about putting together rather than taking apart, integration rather than reduction. (Noble 2006, x–xi)

> Reductionism, which has dominated biological research for over a century, has provided a wealth of knowledge about individual cellular components and their functions. Despite its enormous success, it is increasingly clear that a discrete biological function can only rarely be attributed to an individual molecule. (Barabási and Oltvai 2004, 101)

These quotes give the impression that molecular biology and systems biology are two distinct and well-defined ways of doing biology, and the history of biology at the turn of the millenium is perceived as unfolding with almost Hegelian necessity: the obsolete approach of molecular biology is saluted for preparing the stage and giving way to the new era of systems biology. But what exactly is systems biology and how does it differ from the 'traditional' approach of molecular biology? These are the main questions I want to address in this thesis.

The term 'systems biology' appeared towards the end of the 1990s,[2] and gained widespread use in the early 2000s. Very early on, systems biology showed the characteristic features of an institutionalized discipline: research institutes for systems biology were founded, starting with the Institute for Systems Biology (ISB) in Seattle and the Systems Biology Institute (SBI) in Tokyo in 2000; conferences about systems biology began to be held regularly around the same time; and several journals specifically dedicated to systems biology were created in the following years, such as *Systems Biology* (2004), *Molecular Systems Biology* (2005), and *BMC Systems Biology* (2007).[3] In spite of this concretization at the institutional level, no clear and unique characterization of systems biology has crystallized up to now. The field shows a considerable heterogeneity of approaches that have their historical roots in different traditions of theoretical biology and other theoretical fields studying complex systems. To be sure, there are a number of distinctive features that are commonly cited, such as the combination of mathematical methods with experimental approaches, the investigation of quantitative and dynamic properties of living systems, and a focus on interdisciplinarity and integration. However, aside from these very

[2] Some scholars prefer to speak of the 'new systems biology' in order to distinguish the term from earlier use in discussions about the application of a general systems theory to biology (e.g. von Bertalanffy 1950).

[3] For a more exhaustive historical overview, see Braillard (2008).

general attributes, different systems biologists often highlight different aspects as being central to the new field. Some see the main goal of systems biology in the integration of different levels of biological information (Ideker et al. 2001), while others emphasize the continuity with earlier systems theories (e.g. Wolkenhauer 2001, Westerhoff and Palsson 2004), still others stress the importance of engineering concepts, like robustness, modularity, or feedback (e.g. Kitano 2002b). Given this multitude of accounts and characterizations, it remains unclear what it is that different work labeled as systems biology has in common—whether it is simply the 'continuation of molecular biology by other means,' or a radically different epistemic approach to the study of living systems.

With respect to this philosophical issue, two very different kinds of views are commonly found among biologists themselves. According to skeptics, the new interest in systems approaches is simply a hype. 'Systems biology' for them is a fancy label that helps attracting research funds while doing largely the same thing as before, although perhaps on a larger scale and with more fashionable tools. Advocates of systems biology, by contrast, point out that there are substantial epistemic differences between molecular biology and systems biology. Usually, they invoke the opposition between 'reductionism' and 'holism' to argue for the superiority of their approach (Calvert and Fujimura 2011). It should be obvious, however, that scientists when commenting on these issues are rarely neutral observers. Especially those who identify themselves as systems biologists have a strong interest in justifying and promoting their own perspective. As a consequence, they are prone to equate 'systems biology' with the particular scientific approach they are pursuing, and, on the other hand, to give an oversimplified account of the approach of traditional molecular biology.[4] The consensus that has emerged from the rhetorics put forward by systems biologists, is that traditional molecular biology has confined itself to the study of the parts of living organisms, whereas systems biology aims at understanding how those parts interact to produce phenotypic properties and behavior. This schematic distinction enables them to equate the two labels of 'molecular biology' and 'systems biology' with competing philosophical perspectives: Molecular biologists dissect organ-

[4]An interesting non-standard view is expressed by the molecular biologist Sidney Brenner. Contrary to most 'traditionally minded' molecular biologists, he thinks that systems biology is something very different from molecular biology, and characterizes it as the attempt to solve 'inverse problems,' that is, problems of directly inferring the underlying causal structure of a system from given observational data. He argues that this goal cannot be achieved in biology (Brenner 2010). I doubt, however, that many systems biologists would agree with the way he describes their activities.

isms, list their parts, and try to explain biological function solely in terms of individual molecules or genes. They implicitly follow a reductionist perspective assuming that the whole is captured by the sum of its parts. Systems biologists, by contrast, realize that the interactions between the parts and the systemic context in which they are embedded have to be taken into account, and that biological systems show emergent behaviors in which 'the whole is more than the sum of its parts.' Even though the usage of philosophical terms, such as 'reduction' or 'emergence' is rarely clarified by biologists themselves, and the oversimplification of the dichotomy rather obvious, this consensus is echoed in works that purport to approach the issue from a purely philosophical perspective:

> The molecular biological revolution led to a characterization of the molecular constitution of organisms. Systems biology aims to decipher how the molecules jointly bring about cellular behaviour. The fact that the molecules are supposed to do this jointly suggests that studying them only individually without a focus on their interactions may not work. On the other hand, it is clear that a return to the holist physiology strategy will not work either. Perhaps some new strategy is needed, with unique philosophical foundations. (Boogerd et al. 2007, 8)

While molecular biology is very narrowly defined as the 'characterization of the parts,' the description of the alternative approach is rather vague and general. At times one gets the impression that systems biology is more of a vision of how biology could be done if it were freed of the insufficiencies of earlier approaches.

There are some philosophers of biology who have analyzed issues about systems biology in less vague and more neutral ways. Many of these analyses, however, focus on specific problems occurring *within* systems biology, thereby leaving unclear its relationship to the traditional approach of molecular biology. They deal, for instance, with the classification of different streams of systems biology (O'Malley and Dupré 2005), with the question of how different traditions of mathematical modeling and systems thinking are merged in new approaches (Krohs and Callebaut 2007), the relationship of systems biology to the framework of mechanistic explanation (Braillard 2010), or the role of integration in systems biology (O'Malley and Soyer 2012).

A number of authors have more specifically addressed the question of the relationship between molecular biology and systems biology. Pierre-Alain Braillard (2008), in his dissertation, has given a very detailed account of some of the philosophical problems arising in the context of systems biology. He argues that what distinguishes systems biology from molecular biology is not the study of emergent phenomena *per se*, but rather the *formal* study of such phenomena. Formal methods become increasingly relevant due to the high complexity of the processes uncovered by modern experimental techniques. However, it is not entirely clear whether Braillard wants to imply that the epistemic framework of systems biology is in continuity or in tension with molecular biology. Powell and Dupré (2009) argue that the classification of molecular biology as reductionist misses the philosophically interesting point. The more relevant issue, they argue, is that molecular biology's focus on simple molecular explanations risks to underestimate the real complexity of biological systems:

> [M]olecular biology showed that molecular details do count, and may be richly explanatory. This prosaic yet productive discovery becomes potentially distorting only when it is combined with a commitment towards the simple, since that commitment so easily slips into the simplistic. (Powell and Dupré 2009, 62)

The rise of systems biology, therefore, is the consequence of an "increasing recognition of complexity and context" (Powell and Dupré 2009, 62), and concepts like emergence, even though admittedly vague, might play a productive role as an "essential corrective to misleading philosophical assumptions grounded in traditions of reductionist thought" (Powell and Dupré 2009, 63). Even though it is more refined, this position essentially underwrites the consensus view according to which the traditional approach of molecular biology has to be replaced, or at least to be complemented, by a perspective that is more adequate to the actual complexity of living systems. Taking a somewhat different stance, De Backer et al. (2010) argue in a recent article that systems biology (SB) is in continuity with the traditional approach of molecular biology (MB):

> As such, SB can be considered as a next step in the development of MB, centred on the same biological question, and expanding its experimental toolbox

> with systemwide (omics) analyses and mathematical modelling. (De Backer et al. 2010, 40)

They admit that the problems studied by systems biology involve a higher complexity as far as the size of the system under study is concerned. Yet this does not imply a departure from molecular biology's 'reductionist' focus on molecular features:

> SB definitely realizes the shift from single-gene regulation to genomic regulation; from individual molecules to system-wide molecular interactions; from linear pathways to dynamic networks .... In this, SB takes the molecular view on biological organisms to its full potential. Hence, reductionism is methodologically maintained in SB. (De Backer et al. 2010, 40)

These philosophically more sophisticated analyses help at least to partially revise the simplistic picture of 'reductionist molecular biology' versus 'holistic systems biology.' However, the picture that arises is one in which the boundaries between systems biology and molecular biology are not very sharp: systems biology introduces more powerful tools to cope with complexity and overcomes some of the mental biases of molecular biologists. Is it possible to say anything more precise?

In this thesis I want to propose a framework of comparison that avoids both oversimplified dichotomies and the blurring of relevant differences. I want to argue that there are in fact relevant differences, and that these mainly arise at the level of strategies of discovery. What I mean by discovery in this context is the search for causal mechanisms in order to explain an object's properties and behaviors of interest. My starting point is that scientists dealing with complex systems in nature must generally assume that *what they study is not as complex as it could possibly be*. The reason is that initially they do not have sufficient information to get an idea of the actual complexity of the systems they are studying, and it would be highly impractical to work with the full set of possibilities of how the system *could* be organized. In order to make progress toward an adequate mechanistic explanation, they make use of specific strategies, that I call *heuristics*, whose role it is to reduce this set of possibilities. Heuristics introduce specific assumptions about the system that may or may not be justified. Thus what makes these strategies efficient at figuring out how a system works at the same time creates the risk of underestimating its complexity.

I will argue that the general approach of molecular biology is guided by a more or less well-defined set of heuristic strategies, among which figures prominently the assumption that living systems can be studied by dividing them into a set of relatively small and quasi-independent mechanisms. In addition, it makes use of more specific heuristics that license a focus on molecular properties and qualitative features of these mechanisms. If taken literally as features of reality and not just as tools of discovery, these assumptions can indeed lead to a simplistic perspective on life (which seems to have been Powell and Dupré's worry). The crucial and often neglected point, however, is that alternative approaches, such as the ones classified as 'systems biology,' must apply heuristic strategies as well in order to be efficient. The availability of genome-wide data and the additional power of mathematical methods do not enable scientists to pursue discovery in an unbiased way. For this reason, I propose that the relevant comparison should be in terms of alternative heuristic strategies. Investigating how particular assumptions are relaxed in systems biology while others are introduced, in other words, how specific heuristics are replaced by others, will allow me to identify with some precision both continuities with and deviations from the traditional approach.

My aim is thus not to establish the philosophical foundations of systems biology, as has been the ambition of other philosophical work (e.g. Boogerd et al. 2007), but rather to understand what systems biology is by investigating existing work that goes under the label. For this reason, my analysis makes heavy use of detailed case studies. In this way I avoid both giving an oversimplified account of molecular biology and talking about some idealized version of systems biology that might not be more than a largely unfulfilled promise. Moreover, in order to do justice to the heterogeneity of systems approaches, I have chosen to discuss several different examples from different areas of systems biology. For lack of space and time, I had to leave out large and important parts of the field. In particular, I have not discussed the various 'omics' approaches (genomics, proteomics, metabolomics, etc.) that are sometimes subsumed under the label of systems biology as well. But even if the results of my analysis might not be generalizable to all of systems biology, my general strategy can nevertheless serve as a template for further case studies and perhaps for establishing a more complete picture of the 'epistemic landscape' of current systems biology.

A clearer view of the differences between molecular biology and systems biology is, however, not the only goal of my analysis. Investigating the case of systems biology can also have consequences for some more general issues in the philosophy of biology. Our conceptions of what counts as a good biological explanation, for instance, has in the past shown to be heavily influenced by our assumptions about the complexity and organization of living systems. For this reason, I turn from discovery to explanation in the last chapter and discuss the impact of recent work in systems biology on philosophical models of explanation.

The thesis is structured as follows. In Chapter 1 I introduce a general pragmatic perpective on the philosophy of biology in which the particular explanatory aims of biologists are taken seriously. I argue that an important role for philosophers is to investigate and assess the strategies to reach these aims. Afterwards, I discuss the topics of mechanistic explanation and reductionism and conclude that the central issue regarding the relationship of systems biology to molecular biology is not about explanatory reductionism, but lies mostly at the level of strategies of discovery. I analyze the concept of complexity in some depth and introduce heuristics as tools to reduce the (epistemic) complexity of a given research task.

Chapter 2 discusses heuristic research strategies in traditional molecular biology. I start by discussing already existing work on the topic of discovery by Bechtel and Richardson (1993) and Darden (2006) who have proposed general research strategies applied in the life sciences. By analyzing two case studies, the discovery of the mechanism of protein synthesis and the more recent search for the spindle assembly checkpoint mechanism, I identify further and more specific heuristic strategies of molecular biology.

After having characterized the approach of molecular biology, I turn to systems biology in Chapter 3. Here I discuss several case studies in order to reveal specific differences from the traditional approach of molecular biology. The first example continues the discussion of the spindle checkpoint mechanism and is an instance of mathematical modeling of small mechanisms. These models retain some of the more fundamental strategies of molecular biology. Mathematical methods and quantitative data allow systems biologists to relax some of the more specific assumptions of molecular biology. It appears, however, that this increase in analytic power goes at the cost of introducing different as-

sumptions in the form of idealizations. Next, I discuss two approaches to understand the behavior of large networks. The approach of network motifs provides an alternative strategy to decompose large systems into functional units, but it introduces strong assumptions that require caution in the interpretation of results. The attractor perspective in stem cell biology, by contrast, envisions to forego functional decomposition completely based on the assumption that cellular behavior is simple and coherent at the level of the whole system. My analysis suggests that both of these network approaches should be integrated with investigations of smaller and more detailed models in order to be efficient and reduce potential bias. Finally, I analyze a very recent example of whole-cell modeling which proposes a new way to integrate different styles of mathematical modeling.

Chapter 4 takes up the issue of scientific explanation. Here I argue that recent work in systems biology can lead philosophers to reconsider their conceptions of mechanistic explanation in the life sciences. In particular I discuss the widespread idea that 'difference making' is central to scientific understanding and explanation. Dynamic modeling in systems biology draws attention to the explanatory role of 'non-difference making' relationships. By analyzing the concept of robustness as it is investigated in systems biology, I point to ways in which biological systems can be less complex than what is combinatorially possible, yet in a way that is unexpected from a traditional mechanistic perspective. The explanatory role of mathematical modeling in this context is not to explain complex behavior, but to explain simple behavior exhibited by potentially complex systems.

I have done my best to make this thesis readable to both scientists and philosophers. However, it is difficult at times to strike a balance in this regard, and I apologize in advance for passages that might be either too technical or too superficial for the taste of some readers.

1

PROBLEM SOLVING IN SCIENCE AND THE ROLE OF HEURISTICS

Summary

In this chapter I develop a general pragmatic perspective on the philosophy of biology that focuses on the strategies that biologists use to reach their particular epistemic aims. An important part of their activities consists in the discovery of mechanisms and the development and revision of proposed mechanistic explanations. I discuss the issue of reductionism in biology and argue that the central issue regarding the relationship of systems biology to molecular biology is not about explanatory reductionism, but lies mostly at the level of strategies of discovery. I analyze the concept of complexity in some depth and introduce heuristics as tools to reduce the (epistemic) complexity of a given research task.

1.1 Towards a Philosophy for a Pragmatic Science

Thomas Kuhn (1963) famously referred to most of the activity of scientists as puzzle-solving. And even if one may hold that the idea of 'normal science', that Kuhn essentially understood as the fabrication of expected results, does in general not fit the activities

within the life sciences very well, the comparison with puzzle-solving can nevertheless serve as a valuable analogy. The relevant point is that scientists belonging to the same discipline or field do not only work on related problems, but they usually also share an understanding of how to go about attacking these problems. In other words, they have common ends and make use of common means. They may not always be confident about whether they will find any solution to their problems at all, but they have relatively concrete ideas about the kind of solution they are looking for, and they know which methods and techniques will increase their chances of finding one. Describing scientific activity in terms of solving problems or puzzles, therefore, means conceiving of it as a rational activity, where 'rational' refers not only to the assessment of the eventual results of scientific research, but also encompasses the choice of effective means to achieve these results.

Before Kuhn the prevailing style of doing philosophy of science implied a very different conception of scientific rationality. The main concern of logical empiricism and its direct successors was the analysis of theories—the eventual *results* of scientific activity (e.g. Popper 1959, Nagel 1961, Hempel 1965). The main focus in this endeavor was on the inner coherence of theories and their relationships among one another as well as to the empirical facts provided by experimental observation.

Many philosophers have criticized the logical empiricist approach for promoting an ideal of scientific rationality that is not attainable for real cognitive agents. In the context of this discussion, William Wimsatt (2007b, Chapter 1) distinguishes between two types of rationality that we might refer to as *perfect rationality* and *pragmatic rationality*, respectively. Perfect rationality focuses on logical rigor and represents the ideal of the logical empiricists. Pragmatic rationality, by contrast, is concerned with optimal strategies to reach given aims. According to Wimsatt, "rigor is not a scientific-end-in-itself" (Wimsatt 2007b, 244), and he argues that perfect rationality is too narrow a concept to capture what is going on in most parts of contemporary science. Focusing on logical structure might be the right approach if the goal of science is seen in a fully explicated theory, but, as Wimsatt notices, "at least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them from or reducing them to other theories" (Wimsatt 2007b, 241)

Unless philosophers want to completely dismiss the scientific status of biology as ac-

tually practiced, they must acknowledge that its activity mainly consists in the discovery and description of mechanisms, and not of laws and theories. In comparison with the logical empiricists' ideal, this requires a much more local and pragmatic view on the goals of scientific activity in general. Logical rigor might still be desirable other things being equal, but when faced with the complex types of problems that are common in biology, attaining it will often be an unrealistic requirement. When rationally reconstructing and evaluating modes of scientific activity, philosophers should above all assess whether the strategies chosen by scientists are efficient ways of achieving their particular goals, irrespective of whether these strategies conform to the high standards of perfect rationality. Clearly, such an assessment cannot narrowly focus on theories as the results of scientific activity. After all, for many fields of research, like for biology, it is not obvious that the knowledge produced can even in principle be laid out in the form of one coherent theory—at least if the term is narrowly understood in the traditional sense of a formal axiomatic system. The more important point, however, is that, if philosophers give up the standard of perfect rationality, they have to accept that most of the results of science retain a somewhat tentative character and almost unavoidably carry traces of the process of their discovery. This 'path-dependence' of scientific results strongly suggests that science ought to be analyzed as an ongoing activity, in which *both* the results, in whatever form, and the strategies employed to attain them have to be taken into account. Only in this way can the rationality of scientific endeavors be judged properly.

When philosophers analyze issues like scientific discovery and explanation, they cannot ignore the current state of scientific knowledge. This also means that, if scientists themselves are not sure about certain empirical issues, philosophers should not pretend that they know better. Throughout this thesis, I want to argue that many of the current debates about systems biology can be framed in terms of diverging opinions about the complexity and organization of biological systems. These different conceptions entail different ideas about the way in which biological phenomena should be explained and translate into different research strategies to achieve such explanations.

I will start in the next section by discussing explanation in science in general and briefly present some of the basic accounts that have been proposed. From my assessment of these proposals, I conclude that in disciplines dealing with complex systems, there are

two main aspects of explanation, *intelligibility* and *empirical adequacy* that have to be given weight. Afterwards, I discuss how this plays out in the special case of mechanistic explanation that is prominent in biology. Molecular biologists want to achieve intelligibility by understanding how mechanisms work, but also by showing how biological phenomena connect to the basic properties of matter that are studied by physics and chemistry. The question arises, therefore, to what extent the explanatory project of molecular biology can be considered a 'reductionist' project, and whether systems biology might provide a non-reductionist alternative. My discussion will show that valid objections against explanatory reductionism are not those that are typically put forward by systems biologists. Instead, their objections seem to mostly target particular research strategies of molecular biology. For this reason, I sketch a framework for the analysis and comparison of strategies of scientific discovery. I will introduce complexity and heuristics as fundamental concepts of this framework and thereby prepare the stage for the following chapters.

1.2 Scientific Explanation

We have discussed scientific activity as an instance of human problem solving that should be analyzed in terms of means and ends. The explanation of phenomena is certainly one prominent end that biologists strive for, but it is not the only one. Molecular biology, in particular, due to its close links to the medical sciences, is also involved, for example, in the development of new tools for diagnosis and therapy. In the context of this thesis, however, I want to focus almost exclusively on scientific explanation. I believe that it is one of the guiding ideas behind the project of molecular biology that prediction and control will be achieved *via* an understanding of the mechanisms of life. Therefore, I am interested in the debate around systems biology insofar as it is concerned with the question of how to go about understanding and explaining biological phenomena.

1.2.1 General Conceptions of Scientific Explanation

Scientific explanation is a relationship between something that has to be explained, the *explanandum*, and something that does the explaining, the *explanans*. Philosophers of

science have been debating for many decades about the structure of scientific explanations and about the right criteria to distinguish good from bad explanations (see e.g. Salmon 1989, Psillos 2002).

The more recent discussion about scientific explanation starts with the Deductive-Nomological (DN) model of explanation (Hempel 1965). Its main idea was to frame explanations as sound deductive arguments in which the explanandum, a sentence describing the phenomenon or event to be explained, logically follows from a set of premises among which must be at least one 'law of nature.' According to the DN-model, explanatory force derives from the subsumption of the explanandum under generalizations that describe strong regularities in nature. This model has been criticized on many occasions. One of the main weaknesses is due to the fact that its proponents have been unable to provide a satisfying account of what distinguishes real laws of nature from accidental generalizations. Other problems are that the DN-model proves insensitive towards certain strong intuitions we normally have about explanations, such as asymmetry (effects should not explain their causes), and explanatory irrelevance (taking birth control pills should not explain why men do not get pregnant).

Important attempts to find a more adequate conception of explanation have come mainly from two different directions. Michel Friedman and Philip Kitcher have proposed accounts of explanation as *unification* that stress the importance of scientific understanding. Like the DN-model, unificationist accounts still conceive of explanations as arguments, but they restrict the allowed set of arguments by invoking an additional criterion of economy (Friedman 1974, Kitcher 1981). The basic idea is that scientific understanding consists in explaining a wide range of phenomena on the basis of only a few basic laws or argument patterns:

> [S]cience increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more. (Friedman 1974, 15)

Newton's theory increased our understanding of the world since it allowed us to subsume the movements of celestial and terrestrial bodies under the same set of principles. Explanations are better to the extent that they allow us to derive the explanandum from a

theory that unifies the phenomena better than another one.

Philosophers like Wesley Salmon (1984), by contrast, argued that a conception of explanation should capture the intuitive idea that 'causes explain their effects.' Accordingly, the facts to be included in an explanation should be only those that refer to the causal history of the explanandum phenomenon. The criteria for what counts as a good scientific explanation are not given by epistemic criteria, such as their derivability from a unified theory, but the extent to which they show how the phenomenon fits into the actual causal structure of the world.

Both causal and unificationist accounts have their own difficulties that I do not want to address in detail. What I want to highlight is that they point to two different conceptions of the general aim of scientific inquiry. Wesley Salmon referred to these broad categories as 'epistemic' and 'ontic' conceptions of explanation. On the one hand, the aim of science is to organize our knowledge about the world and to provide understanding of its complexity by reducing it to some restricted set of principles. On the other hand, scientific effort is directed at uncovering the causal patterns in the world *as they actually are*, that is, irrespective of whether they are simple or complex. To be sure, these aims are not necessarily mutually exclusive. As Herbert Simon writes:

> The central task of a natural science is to make the wonderful commonplace: to show that complexity, correctly viewed, is only a mask for simplicity; to find pattern hidden in apparent chaos. (Simon 1996, 1)

However, this quote precisely captures the aspect of faith that is a necessary part of scientific inquiry. Simplicity is a desideratum of the human intellect, but the world is complex. The only way in which we can expect to gain scientific understanding is by believing that the world is only *apparently* complex, that its real structure can be made intelligible.

There is no question that most phenomena in the realm of biology are extremely complex, and most people do not believe that this complexity is readily explained in terms of a small set of principles. Nevertheless, we will see that most biologists expect the knowledge they uncover to be simple in some sense—even though they diverge in their ideas about what kind of simplicity is to be expected. Eventually, I want to show that the main conflict between systems biologists and molecular biologists can be explained in terms of competing ideas about biological simplicity. These ideas translate into different pro-

posed research strategies. Before I come to this point, however, I will have to say more about explanation in biology.

1.2.2 Mechanistic Explanation in Biology

Molecular biology studies complex systems such as bacteria, fruit flies, or yeast cells. The explanandum is most commonly the behavior or capacity of a living system or of a part of it. Molecular biologists are interested in explaining, for instance, how the genetic material of an organism is replicated and distributed, how proteins are formed and how they act in different contexts, or how an undifferentiated egg can give rise to a complex multicellular organism. They try to explain these behaviors by describing underlying molecular mechanisms.

Branching off from the general discussion about scientific explanation, many philosophers of science have recently focused on the concept of mechanism and on how scientists explain phenomena in terms of mechanisms (e.g. Glennan 1996, Machamer et al. 2000, Bechtel and Abrahamsen 2005, Bechtel 2006, Craver 2007). Even though the detailed accounts differ between authors, the general idea is that a mechanism is a complex system of causally interacting parts that produces a phenomenon. As a representative example, I will mention the following characterization given by William Bechtel:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel 2006, 26)

Thus, a mechanistic explanation cites facts about the relevant parts (= components) of a structure, about what these components do, and about their organization. Organization refers to the way in which the components are situated relative to each other. Information about organization partly lies in the spatial configuration of the components, but often the more important aspect is 'functional organization' which consists of the relevant causal interactions and the temporal order of events.

For example, the description of the mechanism underlying a particular type of signal transduction includes as component parts the extracellular signaling molecule (ligand),

the surface receptor, and all the downstream messengers involved in the signaling chain. It includes facts about the structure of these molecules and the ways in which they interact. Finally, it cites the location of these components and the temporal order in which the signaling cascade occurs. Taken together, these facts explain the transduction of the signal.

Interestingly, there is no agreement about whether mechanistic accounts of explanation fall into the category of ontic or of epistemic explanation. Some people do not think that intelligibility is a necessary property of explanation and thus prefer an ontic stance:

> Some phenomena might be so complex that they overwhelm our limited cognitive systems .... It would be wrong to say that the phenomena produced by such complex mechanisms have no explanation. The explanations exist even if we cannot represent them cognitively. (Craver 2007, 34)

Whereas others stress the fact that explanation is a cognitive operation, performed by human beings:

> The important insight is that mechanisms are real systems in nature, and hence one does not have to face questions comparable to those faced by nomological accounts of explanation about the ontological status of laws. But it is crucial to note that offering an explanation is still an epistemic activity and that the mechanism in nature does not directly perform the explanatory work. (Bechtel and Abrahamsen 2005, 424–425)

The kinds of explanations provided by molecular biologists thus potentially exhibit both of the discussed aspects of scientific explanation. By describing a mechanism, they map out the causal processes that are responsible for a phenomenon and thus fulfill Wesley Salmon's requirement that explanation must reveal how the explanandum fits into the causal structure of the world. On the other hand, these descriptions normally reduce the apparent complexity of a phenomenon by making intelligible how it is produced by the organized interaction of a set of component parts.

What is often neglected in discussions of mechanistic explanation is the fact that at least some scientists see the main goal of molecular biology in creating a link between the biological realm and the more fundamental disciplines of chemistry and physics. Even

though philosophers usually construe mechanistic explanations as connecting different levels of organization, such as the cellular and the molecular level, they do not imply that reference to some fundamental or otherwise privileged level must be involved. However, molecular biologists do not strive for *any* kind of mechanistic explanation, they look for explanations at, or at least involving, the molecular level. Molecular biology can thus be considered to promote a strong unificationist ideal: it provides intelligibility partly by showing how biological phenomena can be understood in terms of a relatively small set of principles and interactions from chemistry and physics. Due to this aspect, molecular biology has often been perceived as a *reductionist* enterprise.

Indeed, the most common accusation from the side of systems biologists depicts the project of molecular biology as 'overly reductionistic' and thereby ignoring the real complexity of living systems. On top of that, many people are worried about an 'imperialistic' tendency of molecular biology to invade other disciplines, with the long-term aim of showing that everything can be explained in terms of genes and molecules—perhaps even consciousness and mental states (van Regenmortel 2004). These worries are not plucked out of thin air when considering that what came to be called molecular biology was initially conceived within the Rockefeller Foundation's 'Science of Man' agenda. As the historian of science Lily Kay points out:

> Within that agenda, the new biology (originally named "psychobiology") was erected on the bedrock of the physical sciences in order to rigorously explain and eventually control the fundamental mechanisms governing human behavior, placing a particularly strong emphasis on heredity. (Kay 1993, 8)

In order to evaluate such claims, and to see what role the advent of systems biology might play in this context, we need to define better what is meant by 'reductionism' in different contexts. My aim in the following section is to show that both traditional molecular biology and systems biology can be considered as following an ideal of reductive mechanistic explanation, unless 'reduction' is understood in a very narrow sense. The main differences instead can be detected in research strategies that rely on diverging ideas of the organization of living systems.

1.3 Reductionism in Biology

Reductionism became a widely discussed topic in the philosophy of biology after molecular biology had provided an explanation of the principles of heredity in molecular terms. Initially, the main issue was whether variants of *theory reduction* could be applied in this context. The term goes back to Ernest Nagel (1961), and it refers to the derivation of a higher-level theory from the laws of a more fundamental theory. A standard example of theory reduction is the explanation of classical thermodynamics in terms of the principles and concepts of statistical mechanics. Theory reduction requires the connection of the theoretical vocabularies of the two theories via 'bridge principles'. In the case of thermodynamics, for instance, reduction was achieved by translating higher level terms like 'temperature', into a language that speaks only about molecules and their properties.

In biology, the debate initially revolved around the question of whether the laws of classical genetics could be successfully reduced to the principles of molecular genetics. Even though people like Kenneth Schaffner (1969) tried to adapt the model of theory reduction to the biological context, the difficulties involved in translating Mendelian concepts like 'dominance' into molecular terms eventually lead to an "anti-reductionistic consensus" (Waters 1990). One obstacle to reduction was seen in the problem of 'multiple realization': the concepts of classical genetics can be instantiated by a wide variety of different molecular mechanisms, and no single molecular principle seems to be able to explain the observed higher-level regularities. Another obvious difficulty is that the knowledge produced by molecular biology is not organized into a small body of laws, which seems to be a requirement for successful theory reduction (for an overview of the debate, see e.g. Brigandt and Love 2012).

Many people, however, subsequently pointed out that the conception of theory reduction is not adequate to illuminate the issue of reductionism in biology. Sahotra Sarkar (1992), for example, proposed that one should distinguish between three broad categories of reductionism: *theory* reductionism, *explanatory* reductionism, and *constitutive* reductionism. Theory reductionism, as mentioned, views reduction as a relation between theories. Explanatory (or *epistemological*) reductionism holds that the reduced entity is explained by the reducing entity, irrespective of whether these entities are framed as theories, laws, mechanisms, or even individual observations. Finally, constitutive (or *ontolog-*

ical) reductionism merely asserts that upper-level systems are composed of lower-level parts and conform to their principles, which does *not* imply that upper-level phenomena must be explained in terms of lower-level principles. 'Physicalism' is a form of constitutive reductionism which for most philosophers and biologists is uncontroversial. It asserts that all biological properties *supervene* on physical properties. This is simply to say that all biological facts are 'fixed' by the physical facts, and that there can be no change in a biological property without a corresponding change in an underlying physical property. Biologists nowadays do not believe in vitalistic 'life-forces' anymore.

What is at issue in the more recent debates about biology, therefore, is usually some form of explanatory reductionism. Models of explanatory reduction most commonly start from the idea of causal explanation and aim at capturing the explanation of a higher level feature in terms of the interaction of its parts. Such models can thus be entirely consistent with explanations in molecular biology. They do not require a full theory of molecular biology with genuine explanatory laws. In some models of reduction, the explanatory force of molecular explanations may nevertheless derive from laws, but these are not biological but more fundamental physical or chemical laws (e.g. Weber 2005, Rosenberg 2006). Multiple realizability is not necessarily a problem for conceptions of explanatory reductionism since it is not required that one and the same mechanism must explain all instances of a higher level regularity. For instance, even though different molecular factors and processes underlie Mendel's law of segregation in different biological species, each of these instances can be separately explained with reference to its underlying mechanism.

Given this clarification, what could be the motivation for criticizing explanatory reductionism in molecular biology? The philosophical literature on this topic is vast, and I will restrict myself to discussing only those issues that are of relevance for my general argument.

One aspect that is often mentioned by critics of explanatory reductionism is the importance of context and organization in biological explanations (e.g. Gilbert and Sarkar 2000). A molecular feature or mechanism does not always play the same causal role, but can be involved in the production of different phenomena depending on its context. Complete explanations must therefore include information about the larger system in which a mechanism is embedded. In many cases, the context objection is directed against

forms of *genetic* reductionism that assign a privileged explanatory role to genes.

A different set of objections regards the status of higher level properties. Even though explanatory reductionism does not have to deny that causes at levels higher than the molecular level exist, e.g. a ball breaking the glass of a window, it is committed to the idea that these higher level causes have explanatory force *only in virtue of underlying lower level causes*. Different objections have been raised against this idea. Some people have argued for the existence of *emergent* properties at the systemic level, that is, of properties that cannot be explained or predicted in terms of the properties of the parts. According to strong forms of emergentism, there are systemic properties that are *in principle* irreducible to the properties of the components, while according to others, irreducibility only holds *in practice* due to the complexity of the world and our limited cognitive powers (for an overview, see e.g. Stephan 1999). Other scholars argue that especially the organizational features of biological systems are irreducible to the lower level (e.g. Mitchell 2003, Craver and Bechtel 2006). In general, this class of arguments is related to the issue of 'downward causation', i.e. to the question whether entities at higher levels can exert causal influences on lower level entities.

Still another way of arguing for the autonomy of higher levels is with reference to the irrelevance of molecular details in many higher level processes. This objection is closely related to the argument from multiple realizability discussed above. Recall the example of the ball breaking the window. We know that the ball and the window are both made up of molecules, and we assume that the events that cause the deformation and eventual disruption of the window can be spelled out in terms of molecular interactions. However, we prefer the explanation in terms of the ball and the window to the one in terms of gazillions of molecules—not simply because it is more manageable, but mainly because it seems to describe the process at the *right* level. It does not matter whether the ball is made of rubber or of leather (as long as it has sufficient momentum), and similarly it does not matter whether some minuscule detail in the constitution of the window had been different. Even if we are able to explain a particular instance of the breaking of a window by giving a complete report of the underlying molecular processes, we thereby seem to fail to capture what different window breaking events have in common. As Ingo Brigandt argues:

> An important aim of scientific explanation is to discover the most salient and relevant causal features, and entities above the molecular level can have a stronger causal influence and be less indispensable than causal connections on the lowest level. (Brigandt 2006)

If we believe in physicalism, as most people do, the entities in question are constituted by molecular entities, and the relevant causal features supervene on properties at the molecular level. However, even though every change at the higher level implies a corresponding change at the molecular level, the converse is not true. Many higher level processes are robust to a wide variety of changes at the molecular level, and introducing this kind of molecular detail does not seem to add any explanatory power.

I will now discuss whether the three types of arguments put forward against reductionism are valid, whether they apply in the context of molecular biology, and whether it makes sense to understand systems biology as an alternative, non-reductionistic explanatory project. My conclusion will be that most of these arguments do not provide objections to explanatory reductionism in molecular biology per se, but instead are better understood in a methodological context, that is, as criticisms of particular research strategies in molecular biology.

First, consider the context objection. It is not clear why explanatory reductionism should in general not be able to accommodate cases in which systemic context is relevant. As Brigandt and Love (2012) argue, "models of explanatory reduction can take the organismal context for granted without being committed to reducing it molecularly. Science can avail itself of causes as difference makers relative to a given causal context." Marcel Weber, who devises a model that he calls 'physical reductionism', admits that molecular explanations make use of higher level terms in order to specify e.g. the cellular context of a mechanism, but he argues:

> These concepts are *descriptive* rather than explanatory. They serve to identify the kind of system that is to be explained. The terms that do real explanatory work are all physical and chemical terms…; they refer either to molecular species…, to species of macromolecular aggregates…, or to purely physical entities…. (Weber 2005, 27).

In cases where contextual elements cannot be considered causally inert in this way, because the mechanism produces different effects in different relevant contexts, there is still the option to extend the mechanism by reducing the relevant parts of the context to the molecular level (Delehanty 2005). For this reason, compelling arguments that appeal to the role of context must at the same time provide evidence for the existence of irreducible higher level properties, which will be discussed below.

When looking at historical examples of research in molecular biology, one certainly finds many instances in which the importance of context was initially underestimated. As Phillip Sharp, one of the co-discoverers of RNA splicing, recalls:

> Fifty years ago ... everyone assumed that the structure of a gene was a contiguous string of base pairs, from which information was transferred for synthesis of a protein. (Sharp 2005, 279)

Thus the biologists did not take into account the possibility that the cellular context of the protein synthesis machinery could specifically modify the 'content' of the transferred information. However, such episodes do not show that explanatory reductionism as such is misguided. At best, they provide evidence for the methodological claim that the mechanistic models proposed by molecular biologists are not sophisticated enough. Indeed, systems biologists often point out limitations of this kind in the practice of traditional molecular biology, and they aim at developing approaches that take more of the systemic context into account. However, it is important to notice that this by itself does not imply that systems biology moves beyond explanatory reductionism.

Let us now move to the kinds of arguments that invoke the existence of irreducible higher level properties. In his dissertation, Pierre-Alain Braillard (2008, Chapter 2) has argued that the conceptions of emergence, whether taken in a weak or strong sense, do not provide the right criterion to distinguish systems biology from the traditional approach of molecular biology. I will briefly recapitulate the main line of his argument. According to 'weak' conceptions of emergence, a property of a system is emergent if none of its parts has this property (Stephan 1999).[1] For example, wetness is a weakly emergent

[1] A different idea of 'weak emergence' has been proposed by Bedau (1997), which is defined as underivability except by simulation. This conception might provide a useful criterion to distinguish systems approaches from molecular biology (Bedau, personal communication). However, it should be noted that, like other weak conceptions of emergence, it is consistent with explanatory reductionism.

property of water since it does not make sense to attribute it to a single molecule of H_2O. However, we can explain this property in terms of the chemical properties of individual molecules and their interactions; weak conceptions of emergence can thus be entirely consistent with explanatory reductionism. To hold that molecular biology deals only with non-emergent properties then implies that its typical project is to explain the properties of living systems by directly attributing them to the properties of some of their (molecular) parts. But we have seen that molecular biologists explain in terms of mechanisms, and in a mechanism the parts act *together* to produce a particular behavior. Therefore, explanations in terms of mechanisms are almost always explanations of emergent properties taken in the weak sense. Strong forms of emergentism are mainly discussed in the philosophy of mind, especially in contexts in which the status of mental states or phenomenal qualities is at issue. Supporters of strong emergence claim that there are higher level properties that are both irreducible and causally active. Even though the relevance of these discussions for the narrower biological context discussed here is not obvious, one should mention that powerful arguments have been put forward against the coherence of such views (Kim 1999). In general, it seems that systems biologists do not put forward arguments that rely on the existence of strongly emergent properties (cf. Gatherer 2010). Arguments that rely on irreducibility in practice, on the other hand, seem to make the criteria for what counts as a good explanation dependent on our current cognitive and computational powers. According to Alex Rosenberg (2006), the claim that a particular phenomenon cannot for practical reasons be reduced to the molecular level may well be true, but could also turn out to be false:

> For all we know, there are limits to the complexity and diversity of the natural realm, and what is more important, technological advance in information storage and processing may substantially enhance our capacity to understand macromolecular processes and their combinations. Consider how much of an advance bioinformatics has made in the time since the early 1980s when sequencing ten base pairs a week was an accomplishment. By the early years of the twenty-first century, computational biology was able by a computational algorithm to identify all the genes on a chromosome from the brute nucleotide-sequence data. It would be a mistake to underestimate the power

of the human mind and its prostheses. (Rosenberg 2006, 14-15)

Anti-reductionist arguments that invoke irreducibility in practice, therefore, must rely on a metaphysical position regarding the degree of complexity or "disorder" (Dupré 1993) of the world. Since the correctness of such positions cannot be proven in any strict sense, it ultimately seems to depend on whether one opts for an optimistic or pessimistic reading of the historical record of previous reductionistic research projects.[2] At any rate, as we will see most practicing systems biologists do not hold anti-reductionistic positions of this kind. Finally, if explanations are taken to go beyond the scope of reductionism because they include information about the way in which a system is organized, then arguably even in molecular biology most explanations have to be considered non-reductive. As we have seen earlier, an account of how the parts in a mechanism are organized is one of the key features of the typical mechanistic models put forward by molecular biology. Therefore, if systems biologists accuse molecular biologists of ignoring the complex forms of organization present in biological systems, they cannot mean that organization is completely absent in the accounts of traditional molecular biology. Again, it seems that the only way to make sense of such allegations is as stating that the organizational schemes devised by molecular biologists are not sophisticated enough. This corroborates the idea that the purportedly anti-reductionistic voices from the side of systems biology are better understood as pertaining to methodology rather than to explanation.

The argument referring to the irrelevance of molecular detail seems, at least in my view, to have some bite against explanatory reductionism, as far as the autonomy of higher level disciplines is concerned. However, if the point is that some causal processes at higher levels are more 'salient' because they are robust to fluctuations at the molecular level, then this robustness itself asks for an explanation. In this regard, systems biologists appear often to be *more* reductionistic than molecular biologists since they want to provide explanations for higher level robustness in terms of detailed molecular models. It is perhaps true that molecular biologists sometimes do 'not see the forest for the trees' when they assign causal relevance only to the molecular level, thereby ignoring robust higher level regularities. But this does not necessarily point to a defect of explanatory reductionism. Rather than being anti-reductionistic in any strong sense, most approaches

[2] In the end, this seems to be the main disagreement between Rosenberg and Dupré. For an illustration, see <http://www.philostv.com/john-dupr-and-alex-rosenberg/>.

in systems biology might instead better be understood as attempts to show that it is possible to see both the forest *and* the trees. Whether such attempts will turn out to be successful, is a different question. In Chapter 4 I will come back to the questions related to this argument and show how the investigation of robustness challenges some of the underlying assumptions of current conceptions of mechanistic explanation.

As this brief discussion suggests, and as I want to substantiate throughout this thesis, the epistemic *aims* of systems biology are not necessarily in conflict with explanatory reductionism. My discussion so far, however, indicates that the main differences lie in the epistemic *strategies* proposed to reach these aims. In the next chapter I will argue that the strategies of traditional molecular biology can be understood as one specific way of coping with biological complexity. Whether one wants to call this project 'reductive' in the end seems to be a matter of terminology. The true difference of systems biology can be assessed by finding a way to compare epistemic strategies. By way of preparation, I will now turn to a general discussion of research strategies in biology and of the problem of complexity in particular.

1.4 The Complexity of Discovery

The term 'discovery' in the philosophy of science usually refers to the generation of hypotheses, explanations and theories. Marcel Weber points out that, at least as far as theories are concerned, this is a misnomer because theories are not discovered but constructed by the human mind (Weber 2005, 51). Moreover, speaking of 'discovery' suggests that the development of new concepts and ideas is guided by ingenious intuition, helped perhaps by luck and accident, and eludes rational reconstruction. The logical empiricist tradition, invoking the logical distinction between the 'context of discovery' and the 'context of justification,' considered this whole aspect of science to be immune to philosophical analysis and thereby delegated it to scholars of psychology, sociology, and history.

Starting in the 1950s, philosophers of science began to move beyond questions of logics in the narrow sense and increasingly discussed the reasoning strategies employed by scientists in the process of 'discovery' (for a review, see (Schaffner 1993, Chapter 2). The underlying idea was that, even if this process is usually not guided by truth-preserving,

deductive reasoning, one might still find that scientists follow rational strategies when generating theories. In particular, if scientific activity is conceived as a special case of human problem solving (Langley et al. 1987), then the process of discovery consists in the generation and testing of *possible* solutions to a given scientific problem. Herbert Simon described such problem solving in analogy with the search through a maze:

> The process usually involves a great deal of trial and error. Various paths are tried; some are abandoned, others are pushed further. Before a solution is found, a great many paths of the maze may be explored. The more difficult and novel the problem, the greater is likely to be the amount of trial and error required to find a solution. At the same time, the trial and error is not completely random or blind; it is, in fact, rather highly selective... Problem solving requires *selective* trial and error. (Simon 1962, 472)

Thus, what makes human problem solving powerful are efficient strategies to restrict the set of possible solutions. Simon calls these strategies 'heuristics,' after a term introduced by William Whewell in the 19th century and later readopted by the mathematician George Pólya. According to Pólya,

> [h]euristic reasoning is encountered in all fields, theoretical or practical. Rigorous, precise, properly so-called logical reasoning is found in its pure form only in mathematics. (Pólya 1941, 450)

This statement suggests that the notion of heuristics is extremely broad since it covers basically all types of reasoning that are "fallible" in some sense. However, I want to use a more specific characterization of heuristic reasoning by restricting myself to the particular model of human problem solving developed by Simon and others (Langley et al. 1987). According to this model, scientific discovery is a *search* through a *problem space*. This space is determined by the structure of the research problem and by concepts and parameters specifying possible solutions. In scientific contexts these parameters and concepts usually stem from background theories and beliefs (Resnik 1997). If the problem space is small, one might consider to simply use random search to test all candidate solutions. But in more complex scenarios this is not an efficient strategy. Heuristics are rules of thumb that facilitate the discovery process by restricting or directing the search

through the problem space. They make the search selective, thereby raising its efficiency over blind trial-and-error search.

An illustration may be given by the game of chess. Even though the concept of a 'best move' in chess is in principle well-defined, going through all possible scenarios to find this move is impossible in practice, even for modern computers. Therefore, chess computers (and good chess players) use heuristic strategies to speed up the search, most notably by restricting it to a small number of possible moves ahead and by evaluating these moves according to various factors that are thought to influence the strength of a position. These strategies cannot guarantee that the eventually selected move is the best possible, which is why even the most powerful chess computers and the grand masters sometimes lose.

Large problem spaces make the application of heuristics unavoidable also in scientific contexts. William Wimsatt observes that heuristic strategies are applied whenever "the complexity of the systems we are studying exceeds our powers of analysis" (Wimsatt 2007b, 75) and discusses four important general characteristics of their use:

First, heuristics are *not truth-preserving*. Differently from algorithms, they do not guarantee that the result is a correct solution to the given problem.

Second, heuristics are *cost-effective* with respect to more reliable procedures. This is because they only take into account a restricted subset of the space of possible solutions.

Third, if heuristics produce errors, these errors are *systematically biased*. In order to be efficient, a heuristic procedure makes assumptions about the structure of a problem and the form of its solution that are not directly backed by available knowledge. It usually fails when these assumptions are not justified.

Fourth, heuristics can be understood as effectively *transforming the initial problem* into a related but different problem that is easier to solve.

All this suggests that heuristics are useful, but must be handled with care. There is a risk to reach erroneous conclusions if a bias is introduced that goes unnoticed. On the other hand, the fact that such bias is always systematic gives rise to the hope that one can detect errors and subsequently refine the strategy. In the following chapters we will see how these characteristics of heuristic strategies come into play in concrete examples of biological research. Different approaches can be shown to rely on different families of heuristics that correspond to alternative sets of assumptions about the complexity and

organization of living systems. Before turning to this part, however, I want to make the link between complexity and heuristics more explicit. Problems of complexity seem to lie at the heart of the debates around systems biology, and we need a better understanding of the concept of complexity in order to determine the 'heuristic distance' between alternative approaches.

1.4.1 Two Concepts of Complexity

Complexity is a buzzword—not only in science, but in virtually all areas of societal discourse. The term is heavily used in the discussions around systems biology; on many occasions as a rhetorical tool to argue for the inferiority of one particular approach or to justify another. It seems, however, that there is no general agreement about a precise definition of complexity in biology, and it is not always certain whether different people mean the same thing when talking about complexity.

As a first step of clarification, it is useful to notice that there are two different senses in which complexity enters into the scientific realm: in an ontological and in an epistemological sense. First, complexity is studied as an interesting property of systems. Scientists investigate the ways in which systems are complex or show complex behavior, and they try to understand how complexity can evolve or emerge in a system. I will refer to this property as 'intrinsic complexity'. On the other hand, complexity is used with reference to the difficulty of certain scientific tasks, which means that what is considered complex then is not necessarily the system itself, but a given problem regarding the understanding, prediction or control of its behavior. With Hans-Jörg Rheinberger, I will refer to this second kind as 'epistemic complexity' (Rheinberger 1997a).

This distinction seems relatively obvious. However, when talking about biological complexity both scientists and philosophers often tacitly shift between these two meanings of complexity. The following quote from Warren Weaver's 1948 article *Science and Complexity* may serve as an example:

> The significant problems of living organisms are seldom those in which one can rigidly maintain constant all but two variables. Living things are more likely to present situations in which a half-dozen, or even several dozen quantities are all varying simultaneously, and in *subtly interconnected* ways. Often

> they present situations in which the essentially important quantities are either non-quantitative, or have at any rate *eluded identification or measurement* up to the moment. Thus biological and medical problems often involve the consideration of *a most complexly organized whole*. (Weaver 1948, 536, emphasis added)

He calls these problems instances of 'organized complexity.' However, complexity here refers, on the one hand, to the "subtly interconnected ways" in which the system is organized and, therefore, to an intrinsic property of the system. On the other hand, Weaver invokes the fact that our information about the system is limited, which suggests that he understands complexity at the same time in an epistemic sense. It is very intuitive that those systems which are complex are also those which are hard to study, but the link is maybe not as obvious as it may seem and does not justify a conflation of the two concepts. If complexity is understood as an intrinsic property of a system, it should not depend on the state of knowledge of the investigator and the currently available tools of analysis. When talking about epistemic complexity, on the other hand, we precisely must take into account the investigator's particular cognitive limitations and her access to information. Is it possible to say something more precise about the two concepts?

1.4.2 Intrinsic Complexity

It seems that neither scientists nor philosophers have come up with a concise definition that would capture all the possible ways in which we would want to refer to systems as complex. This becomes especially clear when people try to make comparisons between the complexity of different systems. An example is the debate on complexity trends in evolution, in which the need for a precise definition of complexity is particularly pressing. Ever since Darwin, evolutionary biologists have discussed whether evolution shows a trend of increasing complexity. It is clear that the collection of empirical evidence regarding this question requires some way of quantifying or at least ordering complexity. But even in this context, as one of the involved biologists soberly remarks, "[m]ost agree… that nobody knows precisely what is meant by the word 'complexity' when referring to a biological organism" (Adami 2002, 1085).

There is a discipline, that is sometimes called 'complexity science,' that seems to pre-

cisely deal with the intrinsic complexity of systems. Here, the word refers to a property exhibited by a particular class of dynamical systems. These systems are characterized by features like chaos, nonlinearity, or self-organization. Starting in the 1960s and 1970s, in the course of what some refer to as the 'complex systems revolution,' systems with such features came to be widely studied by mathematicians, physicists and other theoretical scientists (e.g. Hooker 2011). However, it seems that the sense of complexity suggested by such studies is both too narrow and not precise enough to capture complexity in biology. It is too narrow because the mentioned hallmarks of complexity science are not necessarily relevant in all the contexts in which biologists want to speak of complexity. The computational scientist Tjeerd Olde Scheper, for instance, remarks that "[o]ne of the mysteries surrounding the phenomenon of chaos is that it can rarely be found in biological systems" (Olde Scheper 2008, 145). On the other hand, complexity science does not seem to provide a general account of what complexity is and how to measure it, either. Complexity science, therefore, does not give us a general framework that could be readily applied to biological disciplines. By this I do not mean to say that the properties investigated in complexity theory are unrelated to what biologists mean by 'complexity'—there are certainly many systems that biologists consider complex precisely because they possess some of the features studied by complexity scientists. Moreover, there is no doubt that the ideas and mathematical tools developed within complexity theory provide useful tools for the analysis of biological models, especially in systems biology. Yet, it does not provide the kind of general characterization of complexity that we would be interested in.

The inability to provide a general definition may lead to the impression that complexity is a rather mysterious property and will, once properly understood, provide the key to deep metaphysical riddles, such as the difference between the living and the non-living. Undeniably, there has been a certain hype about complexity, notably about the concept of chaos, and it has left some people with the impression that, if we just captured the 'phenomenon' of complexity in the right way, this would provide us with a "fundamentally different idea of how to understand reality" (Hayes 1992 as quoted in Kellert 2008, 11-12). But instead of driving us towards some kind of mysticism, the plurality of ideas about complexity should rather direct our attention to the features that our different ways of *representing* complexity have in common. A very obvious idea about complex systems

is that they are *difficult to represent*, be it their structure or their behavior. I take it to be uncontroversial that a system is minimally complex if we can give a short and simple description of all relevant aspects of its structure and behavior. The more complex a system, the more difficult it will be for us to describe its structure or behavior. Even though characterized with respect to possible representations of a system, this idea of intrinsic complexity should *not* be confused with what I have earlier called 'epistemic complexity' and which I will discuss in more detail in Section 1.4.3. A helpful clarification is provided by the following quote from the introduction of a book called *Complexity* whose authors set out to give a comprehensive discussion of the ways in which "complexity manifests itself in nature" (Badii and Politi 1997, xi):

> [T]he concept of complexity is closely related to that of understanding, in so far as the latter is based upon the accuracy of model descriptions of the system obtained using a condensed information about it. Hence, a "theory of complexity" could be viewed as a theory of modeling, encompassing various reduction schemes (elimination or aggregation of variables, separation of weak from strong couplings, averaging over subsystems), evaluating their efficiency and, possibly, suggesting novel representations of natural phenomena... *[A] system is not complex by some abstract criterion but because it is intrinsically hard to model, no matter which mathematical means are used.* (Badii and Politi 1997, 6, emphasis added)

According to this view, complexity should be assessed with regard to our representations of structure or behavior. A system is complex to the extent that it resists 'condensing' the amount of information that is needed for describing it. The last part of the quote suggests that this conception, even though subject-dependent in an obvious sense, is not necessarily 'subjective' in the sense of depending on the contingent capacities of a particular cognitive agent. Incidentally, focusing on the representation of systems provides a connection to the study of the complexity of syntactic structures (e.g. Kolmogorov [1963] 1998). This connection suggests that one might find a way to express the complexity of a system in terms of some measure of computational complexity. However, the authors of *Complexity* notice that,

> the limited domain of applicability of all existing complexity measures strongly

> suggest that there cannot be a unique indicator of complexity, in the same way as entropy characterizes disorder, but that one needs a set of tools from various disciplines (e.g. probability and information theory, computer science, statistical mechanics). As a result, complexity is seen through an open-ended sequence of models and may be expressed by numbers or, possibly, by functions. Indeed, it would be contradictory if the 'complexity function', which must be able to appraise so many diverse objects, were not itself complex! (Badii and Politi 1997, 10–12)

Fortunately, for the purpose of my analysis I am not in need of the 'universal complexity function.' Instead, the idea that intrinsic complexity corresponds to the extent to which the description of a system can be condensed will be a sufficient guide. Provided that we have found the optimal representation of a system, the resistance to condensation can be expressed roughly as the number of degrees of freedom required. As I will show, this captures most of the intuitions that people share about the complexity of biological systems.

Consider Warren Weaver's distinction between 'disorganized' and 'organized' complexity in the article mentioned earlier. The former applies to systems in which the number of variables is very large, but each variable individually shows "helter-skelter" behavior like the molecules in a gas (Weaver 1948, 538). When dealing with such systems, one can often apply statistical methods in order to find a compact description of average behavior. What Weaver is suggesting, therefore, is that disorganized complexity in many cases is just simplicity in disguise. Sometimes, as in the case of statistics, the introduction of a new analytical method can reveal that a system is not as complex as had previously been thought. Biological systems, by contrast, are *organized*. As a consequence, their description amounts to "dealing simultaneously with a *sizable number of factors which are interrelated into an organic whole*" (Weaver 1948, 539, emphasis in original). The effective number of variables needed for the description of the behavior of a system may thus serve as a good measure of complexity. A similar view can be found in Herbert Simon's reasoning about complexity. He argues that,

> [h]ow complex or simple a structure is depends critically upon the way in which we describe it. Most of the complex structures found in the world are

> enormously redundant, and we can use this redundancy to simplify their description. But to use it, to achieve the simplification, we must find the right representation. (Simon 1962, 481)

In summary, complexity can be understood as an intrinsic property of systems, even if it reveals itself only in our representations. Yet, there is not necessarily one particular feature in the world that makes our descriptions 'long.' Nonlinearity and chaos, but also organization or sheer size can contribute to the complexity of a system.

One further complication must be addressed. The preceding discussion suggests that the complexity of a system can be assessed by choosing some suboptimal representation as a starting point and then condense this representation as much as possible. But in practice our representations are always partial from the beginning and constructed from a particular theoretical perspective. We do not have, at least up to now, one universal and consistent way of describing reality. In particular, there is not necessarily one privileged decomposition of a system into parts, nor even a unique way of determining what the system's behavior is. As Stuart Kauffman observes,

> not only are multiple views about what a system is doing possible, but also any system may be decomposed into parts in indefinitely many ways, and for any such part, it too can be seen as doing indefinitely many things. (Kauffman 1970, 259)

Different theoretical perspectives on a system might therefore yield different descriptions of behavior based on different decompositions. Wimsatt, drawing on Kauffman's ideas, argues that, as long as we do not have one exhaustive and unifying theory, each theoretical perspective taken by itself can only give an impoverished view of the real objects (Wimsatt 1972, reprinted in Wimsatt 2007b, Chapter 9). In particular, it seems that when we call a system 'complex' from a given theoretical perspective, we can in effect only judge the complexity of our particular representation:

> Short of waiting for the ultimate all encompassing reduction to an all-embracing theory, one can only talk about the internal complexity of our different perspectives or 'views' of an object. Nor could one avoid this conclusion by taking the complexity of the object as some aggregate of the complexities of the

> different views of the object, since part of its complexity would be located at the interfaces of these views—in those laws, correlations and conceptual changes that would be necessary to relate them—and not in the views themselves. (Wimsatt 1972, 68–69)

This suggests that, in order to get an idea of the actual complexity of a system, we must also consider how different theoretical perspectives relate to each other. In this regard, Wimsatt offers two concepts—'descriptive complexity' and 'interactional complexity'—that may be seen as proxies for the possibly inaccessible 'actual' complexity of a system. A system is descriptively complex to the extent that different theoretical perspectives pick out decompositions into parts that do not spatially coincide. Scientists belonging to different biological disciplines decompose an organism, like a fruit fly, differently into parts, e.g. according to cell types, developmental fields, or physiological systems. In simpler systems the decompositions according to different perspectives will tend to coincide more.

In order to understand Wimsatt's slightly more complicated concept of interactional complexity, we have to consider the system in a state space representation that describes causal interactions of state variables within a system. Each theoretical perspective picks out different properties of a system and therefore works with a different set of variables. Depending on the desired level of predictive accuracy, one can neglect causal links below a certain threshold of interaction strength and thereby obtain a decomposition into subsystems with strong internal bonds.[3] A system is interactionally complex if many of these subsystems partly fall into different perspectives. One will neglect important causal factors and not be able to predict its behavior with precision, unless one considers it from more than one perspective.

Wimsatt's analysis highlights one further important aspect. When assessing the complexity of a system, not only must different theoretical perspectives be taken into account, but also our desired level of precision. A system that appears to behave in a very simple fashion when represented in a relatively crude way, such as a mass of water flowing through a tube, becomes extremely complex as soon as we aim for a more precise description in which factors like viscosity and turbulence cannot be neglected anymore.

We have to accept that each theoretical perspective gives us an only crude approxima-

[3] Here Wimsatt is directly inspired by Herbert Simon's concept of 'near-decomposability' (Simon 1962) which will be discussed in more detail in the next chapter.

tion of the intrinsic complexity of a system, at best a lower bound. Moreover, it depends on what exactly we consider to be the relevant behavior of the system and on our desired level of precision. Yet, the difficulties we have in dealing with certain systems compared to others must at least to some extent be based on features of these systems themselves. For this reason, I reject a position of 'anthropocentric pluralism', as defended by Meunier (2011), according to which complexity is exclusively grounded in the variety of human interests. It would be odd to say that scientists are struggling with certain problems only because of the variety of their interests (and it would probably sound like an insult to patients who suffer from an incurable type of cancer). I will now turn to a clarification of the relationship between intrinsic complexity and epistemic complexity.

1.4.3 Epistemic Complexity and Heuristics

The concept of intrinsic complexity has to be distinguished from the complexity of solving a difficult scientific problem. Biologists typically face the problem of describing and explaining the behavior of a system while being in a situation of incomplete information and limited experimental access. As a result, there are a great number of possible ways in which the system *could* be organized that are consistent with their current state of knowledge. Ideally, in order to determine the actual structure of the system, scientists must find ways to eliminate all of these possibilities except for one. This task can be understood as a problem of *search* since it amounts to finding the right element in a set of possible solutions to a problem. It seems, therefore, that biologists, like many other empirical scientists, have to solve two problems at once: identifying the structure of a system and explaining its behavior. Mathematicians and theoretical physicists know that it can already be quite hard to represent and predict the behavior of an abstract system whose structure is completely defined and known, but due to the additional problem of finding the causal structure, the tasks biologists face are potentially much more difficult.

There are thus two factors that can contribute to the epistemic complexity of a scientific task: the intrinsic complexity of the system under study and the complexity of the search for its actual structure. But these factors are not independent. We have characterized the intrinsic complexity of a system roughly as the number of independent variables needed to describe a system, whereas the complexity of the search for this description is

related to our state of knowledge about the system. If the intrinsic complexity of a system is very low, then already a few observations might be sufficient to fix the unknown parameters and to solve the epistemic problem. By contrast, if the intrinsic complexity is high, and our description of the system cannot be substantially 'condensed,' then the information needed to make determinate statements would have to be close to complete. We see, therefore, that the epistemic complexity of the scientific task increases, other things equal, with the intrinsic complexity of the system under study.

We have mentioned that in the face of complexity scientists have to resort to heuristic strategies. We can now be more precise about this. The goal of applying heuristics is to *tentatively reduce the epistemic complexity of a particular scientific task*. Some of these strategies are very general and applied across a wide variety of scientific fields, whereas others are very specific to particular areas of research. Concrete examples and the way they work in practice will be presented in the following chapters. In the reminder of this chapter, I want to discuss some general features of heuristics in science. First of all, we can introduce an important distinction. Some heuristics primarily serve to attack the problem of search, whereas others are targeted at tackling the intrinsic complexity of a system. I will discuss both of these in turn.

Heuristics of Search

To get a very rough idea of the complexity of search in molecular biology, consider a system of which we initially can say only that it contains *atoms*, a lot of them.[4] Obviously, there is a huge number of ways in which these atoms could be organized into molecules inside this system. Next, there is a huge number of ways in which the molecules could be organized into larger structures. On top of that, there are many possible ways in which the individual parts may behave and act, and in which the particular causal interaction of two components could be described. In short, there is a combinatorial explosion of possible ways in which the system *could be organized*. If we only consider the number of possible pairwise interactions between individual components, for example, we see that it roughly increases exponentially with the number of parts in a system. Already for systems of moderate size this amounts to an astronomical number of possible interactions.

[4] This is not intended to give a representation of how biologists initially frame the problem.

To be sure, existing background knowledge and available experimental evidence helps to exclude a large number of these possibilities. Chemistry, for instance, provides a lot of information regarding the kinds of assemblages of atoms that are energetically possible. But in most cases this alone is not sufficient to generate a manageable search space. During the process of discovery of a particular mechanism, biologists often say that the mechanism is still 'poorly understood.' What they mean is that they only have a very rough idea of its structure and organization, that there are still many remaining possibilities for how the system could be organized, and that decisive information as to its real structure is still missing. Scientists are thus confronted with a problem of search through a large set of possibilities, and the complexity of this search problem can in principle be estimated by measuring the time it would take a computer to go through and check all of these possibilities.

We can now better characterize what constitutes the 'maze' of the discovering scientist in domains like biology. The problem space is given by the possible ways in which a system could be organized according to reliable background knowledge and experimental evidence. High epistemic complexity means that it is impossible to to go through all candidate solutions by random search. In such situations it becomes rational to apply heuristic strategies in order to make progress towards the solution. These strategies work by making certain additional restricting assumptions about the solution to the problem, thus reducing the number of possibilities that must be considered. An obvious heuristic move is to start with the simplest conceivable organizational scheme. More sophisticated, but still very general heuristics of search are the strategies of *decomposition* and *localization* (Bechtel and Richardson 1993) which will be discussed in detail in the next chapter.

Strategies of search are often guided by metaphors and analogies. The ideas of 'the mind as a computer' and the 'animal as a machine,' for example, are metaphors that can serve as powerful heuristics. According to the interactionist view on metaphor, initially proposed by Max Black, a metaphor is "an instrument for drawing implications grounded in perceived analogies of structure between two subjects belonging to different domains" (Black 1993, 31). Metaphors thus often bring together two different ways of representing complex systems such that one of them 'inherits' some of the structural features of the other. If the system of comparison is more familiar and better understood than the system

under study, one obtains concrete suggestions about possible ways of organization and the relevant level or levels of analysis.

Tackling Intrinsic Complexity: the Role of Idealization

Even if the structure and internal organization of a system are more or less known, it can be difficult to understand its behavior. Only in very few situations there is a straightforward algorithm that can be applied to 'solve' the system. In most cases one has to approach understanding via intermediate steps that transform the initial problem into a more manageable one. Here, heuristic strategies often work by introducing *idealizations*. Since idealization raises several relevant issues, I will discuss it in some detail. Michael Weisberg (2007) has recently distinguished between three kinds of idealization: *Galilean idealization*, *minimalist idealization*, and *multiple model idealization*. These different practices are used to reach particular scientific aims that Weisberg calls "representational ideals" (Weisberg 2007, 639).

Galilean idealization, after McMullin (1985), introduces distortions into theories in order to make them computationally tractable. An example of this kind of idealization, used by Galileo himself, is to neglect the influence of a medium of resistance in the description of the motion of massive bodies. Galilean idealization leads to a simplified representation of the target system and thereby provides a first step towards the solution of the initial problem. The distortion can be removed as soon as advances in computational power and mathematical methods allow for a more complete account of the phenomenon of interest.

Minimalist idealization, by contrast, consists in studying a model which includes only the "core causal factors which give rise to a phenomenon" (Weisberg 2007, 642). Such an idealization is introduced when factors that are known to be causally irrelevant are explicitly neglected. It is thus very close to the operation of *abstraction* which consists in deliberately omitting causal detail. Differently from Galilean idealization, however, this kind of idealization is not just an intermediate step on the way towards a better understanding, but rather promotes the goal of finding the most concise description of a system after it is already understood.

Multiple model idealization, finally, is the practice of building several incompatible

models for a phenomenon. It is used when there is no expectation of arriving at a single 'best model.' This practice is pursued mostly in disciplines dealing with extremely complex phenomena, such as ecology or meteorology. The need for multiple models is justified by the existence of *tradeoffs* between different scientific goals. In complex domains there is often not one single model that can satisfy all goals simultaneously. This position was initially formulated by the ecologist Richard Levins:

> The multiplicity of models is imposed by the contradictory demands of a complex, heterogeneous nature and a mind that can only cope with few variables at a time; by the contradictory desiderata of generality, realism, and precision; by the need to understand and also to control; even by the opposing esthetic standards which emphasize the stark simplicity and power of a general theorem as against the richness and the diversity of living nature. These conflicts are irreconcilable. Therefore, the alternative approaches even of contending schools are part of a larger mixed strategy. But the conflict is about method, not nature, for the individual models, while they are essential for understanding reality, should not be confused with that reality itself. (Levins 1966, 431)

Thus the existence of multiple models for the same phenomenon is not necessarily a symptom of disagreement about the right solution to the problem, but may instead reflect a pragmatic reaction to the perceived complexity of nature. As Levins suggests, multiple model idealizations are generally not expected to be replaced by one correct model. Instead, our understanding often derives from the combined account of several models that are individually incorrect. In Levins's own famous words, "our truth is the intersection of independent lies" (Levins 1966, 423).

Galilean idealization and multiple model idealization are used in domains with high intrinsic complexity. They are applied when the structure and organization of the system under study is already known in some detail and the remaining problem is to generate a model that fulfills a particular scientific goal. Both of these strategies can be seen as reducing the epistemic complexity of the task by creating a model that deliberately *underestimates* the intrinsic complexity of the system. In the case of Galilean idealization, this serves as an intermediate step towards a more accurate representation. The way in which idealized models can serve as heuristic tools is discussed at length in Wim-

satt (1987, reprinted as Wimsatt 2007b, Chapter 6), and we will come back to specific instances of this practice in later chapters. Multiple model idealization, on the other hand, is used when one has given up the goal of doing justice to the intrinsic complexity of the system and settles for the optimization of other scientific goals, such as predictive accuracy or generality. In molecular biology, differently from ecology or meteorology perhaps, there seem to be many scholars who believe that it will eventually be possible to find the right models, and thus they assume that the intrinsic complexity of the systems they are studying lies within our powers of analysis.

Minimalist idealization, by contrast, is not so much a strategy for solving scientific problems, but rather a way of rendering the solution of this process. Here, all irrelevant causal detail is stripped away in order to arrive at a representation that is true to the system's intrinsic complexity and successfully explains its behavior. It should be noted, however, that such instances of successful explanation often serve as exemplars for the solution of further problems in the same or related domains and thereby acquire heuristic character.

As I want to show later on, the roles of intrinsic complexity and idealization gain importance in systems biology approaches. In traditional molecular biology, by contrast, the emphasis is on heuristics of search. But obviously, the distinction between the two kinds of heuristics is not as clear-cut, nor is their assignment to different scientific approaches.

1.5 Conclusion

In this chapter I have introduced some of the basic notions that I want to apply in my analysis and motivated the particular approach I want to follow to analyze the relationship of systems biology and traditional molecular biology. I consider scientific activity in biology to be a rational process, directed towards achieving explanation and understanding. Explanation, after Salmon, has an ontic and an epistemic aspect. On the one hand, scientists try to figure out actual causal structures and thus strive for realism in their representations of the world: events are explained by other events. On the other hand, scientific explanation is supposed to provide intelligibility by reducing complex phenomena to simpler principles. In the mechanistic explanations given in most areas of biology, both of

these aspects can be found. Biologists certainly try to figure out causal structures, but at the same time they want to grasp how these structures work as mechanisms that produce particular behaviors or fulfill certain functions within larger systems.

One of the main accusations of systems biologists is that molecular biology ignores the complexity of living systems by being 'reductionistic.' Allegedly, molecular biologists want to eventually reduce all living phenomena to physics and chemistry. My brief discussion of the general philosophical debate on reductionism suggests that the mechanistic approach of molecular biology can be considered to follow a model of explanatory reductionism. However, this model is so general that it is likely to accommodate most of systems biology as well. Systems biologists do not put into question the knowledge that has been accumulated by molecular biologists and the explanations it provides. It rather criticizes the way in which molecular biologists attempt to approach the problems that are as yet unsolved. I have argued, therefore, that the relevant differences between molecular biology and systems biology are to be found at the methodological level. In particular, at the level of strategies of discovery.

I have presented a general framework for the intended analysis of different research strategies. In this framework, scientific discovery is conceived as a special case of human problem solving that relies on heuristics. Heuristics are fallible but efficient strategies to deal with the epistemic complexity of scientific tasks. It is important to pay attention to the distinction between the epistemic complexity of a task and the intrinsic complexity of the studied system. The epistemic complexity of a task is both due to this intrinsic complexity and the complexity of the search of identifying the structure and organization of the system. These two aspects broadly correspond to the different functions that heuristic strategies can fulfill in scientific discovery.

2

RESEARCH STRATEGIES OF MOLECULAR BIOLOGY

Summary

My aim in this chapter is to characterize the problem solving approach of molecular biology in terms of a particular set of heuristic strategies. I start by discussing some very basic strategies of discovering mechanisms, drawing in particular on the work of Bechtel and Richardson (1993) and Darden (2006). Afterwards, I present two case studies of mechanistic discovery in molecular biology in order to identify further, more specific heuristics. In particular, this analysis will illuminate the habit of molecular biologists to look for mechanistic accounts that are relatively simple and apparently not in need of quantitative reasoning. I will show that this habit relies on a particular idea of biological organization and complexity.

2.1 Introduction

In order to answer the question of whether systems biology is *new* or *different*, we must somehow characterize the traditional approach of molecular biology. However, this task is not straightforward since, as Sahotra Sarkar has observed,

> perhaps the only guideline for demarcating the boundaries of molecular biology is that research is guided by an exploration of interactions at the molecular or sub-molecular level. However, if this characterization is pushed to its extreme, there is a problem. Since all of biology seems to be using molecular techniques, is there any "non-molecular biology" left? (Sarkar 1996, 7)

Thus it seems difficult to characterize molecular biology by pointing to a particular domain of biological phenomena or by describing the specific problems that it addresses. A similar point has been made by the historian and philosopher of biology Richard Burian:

> On my rather traditional account, disciplines are organized and institutionalized bodies of research focused around a core group of questions. Molecular biology, taken widely, is extremely well organized and institutionalized; nonetheless, on my account it is not a discipline, because it does not center on a focal group of questions. Molecular biology, after all, studies, among many other things, the structure and behavior of proteins, but also of polysaccharides, lipids, lysosomes, ribosomes, membranes, muscle fibrils, etc., etc. Molecular biology is thus a technique-based field that impinges on, or includes, a number of disciplines, many interdisciplinary investigations, and many investigations whose disciplinary location, if any, is uncertain. (Burian 1993, 387–388)

If molecular biology is such a heterogeneous and open endeavor itself, how can we arrive at a characterization that allows us to compare it in philosophical terms with another seemingly heterogeneous and open endeavor such as systems biology?

Perhaps one can find the right starting point for such a comparison by considering that the set of 'techniques' that make up a field like molecular biology do not have to be restricted to the material realm. Molecular biology might not pursue a well-defined set of problems, but it might nevertheless have a preferred set of cognitive strategies to deal with biological problems. As I have indicated in Chapter 1, the key to the comparison of the different approaches in contemporary molecular biology lies in the concept of complexity. If I am right, then we need to understand how traditional molecular biology conceives of the organization of biological systems and what methods it proposes to reduce epistemic complexity.

It is not obvious from the start that one should find such a set of general heuristics that would allow for a characterization of research in molecular biology. It is at least conceivable that the particular strategies used by molecular biologists are as diverse as the problems that they are applied to. However, when we look at typical examples of discovery and explanation in molecular biology, we see certain common traits that suggest a shared idea of basic strategies. In the recent philosophy of the life sciences, the reasoning strategies involved in scientific discovery have prominently been discussed in terms of the search and refinement of *mechanisms*. It has been argued that this provides the right tools to describe the production of scientific hypotheses and to understand scientific change neither as a sequence of refutations nor as an irrational replacement of paradigms. Instead, a focus on mechanisms promises to capture science as an "error-correcting process" (Darden 2006, 2) that allows researchers to revise and adapt their initially sketchy hypotheses in the light of new findings.

I will discuss the strategies of mechanistic discovery that have been proposed by Bechtel and Richardson (1993) and Darden (2006) and show that they allow us to partly characterize research in molecular biology. However, they do not fully do justice to one aspect that seems especially relevant for a comparison with systems biology, namely the fact that mechanistic accounts in molecular biology are mostly *qualitative*. How is it possible that a science that attempts to explain complex phenomena in terms of molecular properties could largely do without any kind of formalization and quantitative reasoning? One might think that molecular biology up to recently was simply not developed enough to become a quantitative science, and that the advent of systems biology marks a step of maturation. But I am not too happy with this interpretation. It seems that in the early days of molecular biology it was expected that biology would necessarily turn into a quantitative science. Warren Weaver, who actually coined the term 'molecular biology' (Kay 1993), wrote in 1948:

> As never before, the quantitative experimental methods and the mathematical analytical methods of the physical sciences are being applied to the biological, the medical, and even the social sciences .... It is tempting to forecast that the great advances that science can and must achieve in the next fifty years will be largely contributed to by voluntary mixed teams, somewhat sim-

> ilar to the operations analysis groups of war days, their activities made effective by the use of large, flexible, and highspeed computing machines. (Weaver 1948, 541–542)

It is not clear then why the subsequent developments of molecular biology in the 1950s, which are usually perceived as spectacular successes, should have led to the insight that molecular biology was not a mature science after all. It seems more plausible that these successes suggested a research program that could be efficient in the absence of 'mathematical analytical methods.'

The history of molecular biology is often described in terms of two distinct phases (e.g. Rheinberger 2007). The first phase was centered around the characterization of DNA structure in the 1950s and can be understood as the result of a cooperation of different disciplines, such as biophysics, biochemistry and genetics. Early molecular biology was, however, not simply a continuation of these disciplines, but formed an "active assemblage in its own right" (Rheinberger 2007, 219). Sahotra Sarkar, in a similar spirit, has referred to Watson and Crick's description of the double helix structure of DNA as a "confluent model" (Sarkar 2005, 22) that channeled important insights from different sources, and motivated a new research program. Conceptually, this phase was dominated by the notion of 'genetic information,' and the period following the deciphering of the genetic code can be understood as a stage of 'normal science' (Morange 1998, Chapter 15) in which biological problems were framed in terms of the informational vision of life. With my analysis of the discovery of the mechanism of protein synthesis in Section 2.3, I want to suggest that particular aspects of this informational vision can illuminate the more specific heuristic strategies applied in molecular biology.

The beginning of the second phase of molecular biology, around the 1970s, has been described as a transition towards molecular biotechnology. New techniques of genetic engineering allowed molecular biologists to overcome the limitations of test tube assays and to directly intervene into the intracellular mechanisms (Morange 1998, Chapter 16). These new possibilities marked a deep shift in the general development of molecular biology. The results obtained by means of the new tools led to important refinements at the conceptual level. However, as I want to show in Section 2.4, important aspects of the classical molecular vision were retained and have continued to guide molecular research

up to the present.

I will start in the next section by discussing accounts of mechanistic discovery proposed by Bechtel and Richardson (1993) and by Darden (2006). The former propose very general heuristics that are applied across a wide range of disciplines, whereas the latter also discusses strategies that seem to be relatively specific to molecular biology. My analysis of two case studies will, however, go beyond their accounts in order to give a more complete picture of the set of heuristics used in molecular biology.

2.2 Discovering Mechanisms

In their pioneering work on mechanistic explanation William Bechtel and Robert Richardson (1993, 2nd ed. 2010) develop a picture of theory development and change that heavily relies on an analysis of concrete historical case studies taken from biochemistry and cognitive neuroscience. They frame this analysis within the general picture of human problem-solving discussed in Chapter 1. The reasoning strategies they discuss inevitably entail a risk of failure, but such failures often represent starting points for the revision of initial proposals. Therefore, the framework of heuristics, when understood in this way, provides the means to reconstruct and to learn from instances of both progress and failure (Bechtel and Richardson 1993, Chapter 2).

In spite of this explicit focus on heuristics, most of the later discussions in the philosophy of the life sciences have not pursued this aspect further, but instead picked up almost exclusively on Bechtel and Richardson's conception of mechanism and its import for scientific explanation. This is unfortunate since the methodological considerations they elaborate are visionary, to say the least, and seem particularly relevant for an adequate understanding of the current developments in molecular and systems biology.

The title of the book, *Discovering Complexity*, already hints at the general idea of understanding scientific discovery as a process of progressive revision, leading to increasingly complex representations of the structures that underlie the studied phenomena. Throughout, two basic strategies of discovering mechanisms, *decomposition* and *localization*, are discussed which are argued to have guided, and to continue to guide, much of the activities of scientists, notably in the life sciences. These strategies are *mechanistic*

in the sense that they can be understood in analogy with the way in which we attempt to explain the working of an engineered machine. This is thus an instance of how metaphor is turned into heuristics:

> A machine is a composite of interrelated parts, each performing its own functions, that are combined in such a way that each contributes to producing a behavior of the system. A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization. (Bechtel and Richardson 1993, 17)

An overarching theme of mechanistic reasoning, therefore, is the idea that understanding the behavior of a complex system consists in determining what the parts of the system are and what they do. The authors introduce a broad distinction between two classes of strategies for isolating the components of a system, to which they refer as *analytic* and *synthetic*, respectively. Analytic strategies try to identify components of the system physically and then perform experiments by intervening on these components in order to assess their contribution to the overall behavior. Biologists know that the systems they study are composed of physical parts, some of which can be distinguished and intervened on by experimental means. Such experiments might provide clues to the way the system produces its behavior. Synthetic (or functionalist) strategies, by contrast, start from a conjecture about the way in which a behavior might be produced by a set of hypothetical component operations. Many models in cognitive science and artificial intelligence, for instance, propose hypothetical models of how particular cognitive tasks are achieved. These models can then be tested by comparing their performance to the actual behavior of the system.

Both kinds of strategies are 'heuristics of search' in the sense discussed in Chapter 1, and consequently both are prone to errors. Analytic strategies proceed by privileging those parts of a system that are readily accessible to experimentation. The observation of a strong effect might lead to the conclusion that the part intervened on is responsible for a behavior; but it might merely be involved in providing necessary background conditions. On the other hand, the hypothetical organizational schemes devised by synthetic strategies usually draw on resources beyond the system's observed behavior. In the

absence of further empirical constraints, they might merely produce a model that represents one *possible* way of how the behavior is brought about, which is not necessarily the *actual* one.

Therefore, Bechtel and Richardson argue, successful discovery usually requires an interplay of both analytic and synthetic strategies. The combined strategies of decomposition and localization can be interpreted as one particular way in which this interplay may take place.

2.2.1 Decomposition and Localization

The heuristic strategy of decomposition is characterized as follows:

> *Decomposition* allows the subdivision of the explanatory task so that the task becomes manageable and the system intelligible. Decomposition assumes that one activity of a whole system is the product of a set of subordinate functions performed in the system. (Bechtel and Richardson 1993, 23, emphasis in original)

One starts with a complex problem, the explanation of a phenomenon, and arrives at a reduction of (epistemic) complexity by conceptually subdividing it into more manageable subtasks. This is possible because the activity of the system is conceived as the product of component functions.

To give a toy example, let us assume that our task is to explain how a telephone works and that we initially don't know anything about its inner structure. We observe that the telephone allows people separated by large distances to talk to each other. Assuming that this complex activity is the product of several sub-operations, we look for a functional decomposition. The telephone must be able to convert the human voice into a transmittable signal and, on the other hand, convert incoming signals into acoustic signals that resemble the messages that have been sent at the other end. Furthermore, the signals must be transmitted in some way from one person to the other. Thus we have identified three subordinate activities through which the system may perform its overall function that we could call *conversion*, *re-conversion*, and *transmission*.[1] The assumption that each

[1] For the sake of simplicity, I restrict myself to the Bell-style setup of two connected apparatuses and ignore the additional complexities brought about by the existence of a telephone network and the possibility of connecting to particular people by dialing.

of these activities is performed by a different part of the system, and thus can be explained independently, considerably simplifies the initial task.

Decomposition, according to the terminology introduced earlier, is a synthetic strategy since it proposes a hypothesis about how the overall phenomenon is produced. The second strategy of *localization* is supposed to ground such hypotheses in the physical structure of the system:

> *Localization* is the identification of the different activities proposed in a task decomposition with the behavior or capacities of specific components. (Bechtel and Richardson 1993, 24, emphasis in original)

Localization thus presupposes decomposition into subordinate functions and consists in finding a mapping between this *functional* decomposition and a *structural* decomposition of the system. This structural decomposition may already be available due to background knowledge. Alternatively, it can be the result of a specific line of experimental research inspired by the initial proposal of functional decomposition. In the case of the telephone, when looking for structural parts that are responsible for the identified suboperations, we may eventually find that three particular components, that we could call *microphone*, *earphone*, and *wire*, correspond to *conversion*, *re-conversion*, and *transmission*.

Of course the caricatural example of the telephone omits several features that are relevant in the discovery of biological mechanisms. First of all, we have taken for granted what the boundaries of the system are that performs the complex activity. In the practice of scientific discovery, however, it is often far from obvious where and at what level one should look for a mechanism. Bechtel and Richardson suggest that one of the first stages in the discovery of a mechanism consists in the search for the *locus of control*. An example where initially there was controversy is the search for the locus of control for the phenomenon of respiration. During the 19th century competing proposals were put forward according to which respiration occurred either in the lungs, in tissues, or in the blood. Eventually, this controversy was resolved in favor of the cells found in biological tissues. Identifying the locus of control goes along with the segmentation of a system from its environment. This system is established as the site at which to look for the factors that produce and control the behavior. The external context is considered to only provide

background conditions and to not properly exert control on the mechanism (Bechtel and Richardson 1993, Chapter 3). The search for a locus of control by itself is a heuristic strategy that can introduce bias by assuming some degree of context-independence. Wimsatt considers this a reductionist research strategy:

> [T]he focus of the reductionist will lead him to order his list of 'economic' priorities so as to simplify first and more severely in his description, observation, control, modeling, and analysis of the environment than in the system he is studying. (Wimsatt 2007b, 81)

Note that 'reductionism' here is understood in a *methodological* sense, and could be characterized as implying the application of research strategies that are somehow directed 'inward,' that is, towards the analysis of smaller segments of reality.

Once a locus of control has been identified, there are still many possible ways of decomposing an activity into subordinate activities, and likewise many different ways of structurally decomposing a system. In their case studies, Bechtel and Richardson (1993) observe that scientists often start with the simplest assumption of *direct localization*, which means that they look for one specific component that by itself is responsible for the activity.

> Direct localization assumes that there are a number of components in the system, that these components function independently, and that any complexity in the behavior of the system is the effect of isolable subsystems. (Bechtel and Richardson 1993, 64)

Consider a modern personal computer that allows the user to do a variety of things, such as writing a text, listening to music, or watching a movie. Each of these activities is enabled by a different program, and each of the programs works independently from the others. In the same way, as Bechtel and Richardson discuss, direct localization was used by some early investigators of brain function to assign different cognitive tasks to different parts of the human brain. The same idea underlies the decomposition of the physiological system into 'organs,' each of which is responsible for one of the activities of the whole system.

Obviously, direct localization does not yet explain an activity. All it does is to "locate an underlying system within a complex system" (Bechtel and Richardson 1993, 65). It merely 'relocates' the locus of control without explaining *how* the activity is produced. Thus, if direct localization is successful, it has to be considered as the starting point for a lower level analysis of the identified subsystem. Failure of direct localization, by contrast, directs the attention towards more complex forms of localization at the initial level of analysis:

> Simple [i.e. direct] localization differentiates tasks performed by a system, localizing each in a structural or functional component. Complex localization requires a decomposition of systemic tasks into subtasks, localizing each of these in a distinct component. (Bechtel and Richardson 1993, 125)

Strategies of complex localization give up the assumption of independence and consider that different components interact and together produce the overall activity of the system. In the case of the telephone, it is important that the components are arranged in a particular way, otherwise they would not be able to perform the overall task of 'allowing communication at a distance.' In other words, the organization of the components becomes crucial.

As before, scientists often start with simple assumptions, for instance, that the organization is 'linear,' which means that the overall task is performed as a chain of suboperations in which the product of the activity of one component serves as the input to the next operation. Simple organizational schemes allow researchers to study the activity of each component in isolation and to understand the behavior of the whole system by 'simulating' the chain of events in their minds. Later in this chapter I want to show that traditional molecular biology has a strong bias towards such simple forms of organization.

2.2.2 Assumptions and Limits of Decomposition

As Bechtel and Richardson stress, decomposition and localization can be directly successful only if certain assumptions about the system under study are met. A closer look at these assumptions reveals the heuristic character of these strategies. First, for the system to be amenable to a structural decomposition into parts, we must assume that the system

is constituted by more or less stable subsystems (which may themselves consist of further parts); that is, we imply a *hierarchical* organization into levels. In addition, we assume that among the identified components of the system each has an intrinsic function and performs this functions in relative independence from the others. Put differently, we assume that the system is composed of functional *modules* that we can study independently to understand their role in the systemic context.

Discussing the strategies involved in the investigation of hierarchical systems, Herbert Simon (1962) introduced the concept of *near decomposability* as the structural counterpart of modular organization. In his terminology a *decomposable* system is one in which the interactions *between* the parts are negligible when compared with the forces acting *within* the parts. When confronted with such a system, we can treat the parts as if they were independent of each other. As an example of a decomposable system Simon mentions the case of a rare gas in which the intermolecular forces are many orders of magnitude smaller than the chemical bonds holding together the individual molecules. Elaborating on this idea, Simon then characterizes *nearly decomposable* systems as those where the interactions among the subsystems are not negligible, but weak. He goes on to specify two main properties of nearly decomposable systems:

> (*a*) in a nearly decomposable system, the short-run behavior of each of the component subsystems is approximately independent of the short-run behavior of the other components; (*b*) in the long run, the behavior of any one of the components depends in only an aggregate way on the behavior of the other components. (Simon 1962, 474)

This means that the behavior of the systems can—in approximation—be described in terms of aggregate variables, or modules, and the organization of the system itself can be accounted for by referring to the intrinsic properties of these modules together with a set of input-output relationships between them.

The assumption of near decomposability allows for a tremendous reduction of epistemic complexity. It allows us to neglect the behavior of the inner parts of the modules, thereby considerably reducing the required dimensionality of our models. However, by making this assumption, we restrict ourselves to a very specific class of systems of relatively low intrinsic complexity, and we cannot take for granted that the systems in nature

will fall into that class. Imagine that we constructed a virtual network of interacting nodes by assigning interactions of varying strength to pairs of nodes at random. The probability of ending up with a nearly decomposable system in this way is extremely small; in other words, within the space of possible systems of this general sort, nearly decomposable systems are very rare.[2] Why should we expect to find them in nature?

In order to argue for the utility of decomposition and localization as general research strategies, one has to give reasons why the strength of this underlying assumption is unproblematic. One can either argue that it is justified simply because nature *is that way*, hierarchically organized, and allows for a description in terms of nearly decomposable systems. Alternatively, one can maintain that the assumption of near decomposability represents a good *first approximation* even in situations where it is invalid, and that the strategies of decomposition and localization should be regarded as a useful guide towards adequate mechanistic explanations also of more complex systems.

Simon himself appears to be going for the first option and presents an evolutionary argument for the predominance of hierarchical modular systems in nature. His claim is that "complex systems will evolve from simple systems much more rapidly if there are stable intermediate forms. The resulting complex forms will be hierarchic" (Simon 1962, 473). He illustrates this claim with a parable: Two watchmakers, Hora and Tempus, are both building watches out of 1000 components while being constantly disturbed by incoming telephone calls. Tempus builds his watches in one go, whereas Hora first constructs subassemblies of 10 parts, assembles those into larger subassemblies, and so on. Tempus's watches are stable only when fully completed, so every time he is interrupted he will have to start from scratch. Hora, by contrast, only loses the subassembly he was currently working on. Simon shows with explicit calculations that Tempus, on average, needs much more time to assemble a watch completely, even though the assembly of Hora's watches needs more steps. Like the watchmakers, nature is building complex structures in the face of permanent perturbations. Therefore, Simon thinks that "the lesson for biological evolution is quite clear and direct. The time required for the evolution of a complex form from

[2]Simon reasons in a similar way, using a representation in terms of 'nearly decomposable matrices.' These are matrices that can be arranged so that all large elements lie in square sub-matrices along the main diagonal. He observes that this is a "rather strong property for a matrix to possess, and the matrices that have this property will describe very special dynamic systems—vanishingly few systems out of all those that are thinkable" (Simon 1962, 475)

simple elements depends critically on the numbers and distribution of potential intermediate stable forms" (Simon 1962, 471). Evolution is thought to be much more efficient if it builds systems from stable intermediates, which is why we can expect hierarchical order and modularity to be ubiquitous in nature. On a more cautious note, however, Simon reflects upon the double task of a heuristic strategy to both capture the actual structure of a system and to simplify our description of it:

> The fact ... that many complex systems have a nearly decomposable, hierarchic structure is a major facilitating factor enabling us to understand, to describe, and even to "see" such systems and their parts. Or perhaps the proposition should be put the other way round. If there are important systems in the world that are complex without being hierarchic, they may to a considerable extent escape our observation and our understanding. Analysis of their behavior would involve such detailed knowledge and calculation of the interactions of their elementary parts that it would be beyond our capacities of memory or computation. (Simon 1962, 477)

Thus, Simon seems to argue that our strategy to understand complex systems by decomposing them might be of no help when dealing with different classes of systems. This concern becomes all the more pressing when we consider that Simon's argument for the ubiquity of modularity has been challenged by other authors and may not have the intended general scope with regards to biological evolution. According to the watchmaker argument, we can expect hierarchical organization in living systems because it is much more costly to evolve an integrated structure. However, once a structure of a certain size has evolved, it is not obvious that its organizational features will be maintained. Modularity can have fitness decreasing effects as well; therefore, which type of organization prevails and whether it will be maintained depends on the particular conditions in which a system evolves. Wimsatt (1972, reprinted in 2007b, Chapter 9) argues that the subassemblies of a system will over time tend to become *more* integrated since "the optima and conditions of stability for a system of aggregated parts are in general different ... from the optima and conditions of stability for its parts taken in isolation" (Wimsatt 1972, 76). A further issue regards the extent to which we can expect decompositions according to structural criteria to coincide with functional decompositions. The study of large networks in biology

suggests that,

> functional modules do not in general coincide with structural ones in biological systems ... and ... functionality in metabolic and gene regulatory networks is not localized at particular components of the system but delocalized or distributed over entire subnetworks. (Krohs 2009, 269)

This already gives a hint at the importance of this discussion for systems biology, and it will be taken up again in the next chapter.[3]

Bechtel and Richardson are aware of these concerns and concede that for many systems the assumption of near decomposability will not be justified:

> [A] wide variety of organizations may be revealed by beginning with an assumption of near decomposability. The resulting models may not retain the integrity of the components, but may describe what we have termed an integrated system. In such a system nature is at best minimally decomposable. If organization becomes even more dominant in explaining the behavior of the system ..., we reach a point where decomposition and localization in any recognizable form have to be surrendered. (Bechtel and Richardson 1993, 199)

They describe a continuum of systems with simply decomposable systems at on end and fully integrated systems at the other. They maintain that decomposition and localization are useful strategies, even for systems that are only minimally decomposable. Sophisticated forms of organization might be unveiled by starting with the approximation of near-decomposability and adjusting our models subsequently. In this process, initial failures of localization may provide important hints towards more accurate accounts, so the heuristics can work as powerful tools for the detection of errors. At a certain level of integration, however, altogether different strategies may be needed:

> There are other systems, yet farther out on the continuum, in which localization and decomposition appear to be hopeless, or even misguided. The hallmark of these cases is that, given a principled structural analysis, the activities of the parts seem to be different in kind from, and so far simpler than, those performed by the whole. (Bechtel and Richardson 1993, 202)

[3] For a more comprehensive discussion of the concept of modularity in general see Callebaut (2005).

But even if the localization of functions fails due to complex organization, the behavior of the system is nevertheless produced by the activities occurring within it. For this reason, a general mechanistic perspective might still be useful, and increasingly powerful computational methods might allow us to build models of such systems as well.

Traditional molecular biology, as I want to show, remains firmly within the scope of decomposition and localization strategies. Giving up these strategies might, therefore, be one way in which approaches in systems biology deviate from the traditional model. However, they are but the most general in the set of heuristics used by molecular biologists. As Bechtel and Richardson show, they have been applied in a wide variety of scientific fields. I will now turn to the discussion of some more discipline-specific strategies of molecular biology.

2.2.3 Pruning the Hypothesis Tree: The Role of Constraints

Heuristic strategies can be understood as psychological, or cognitive, constraints on the space of possible explanatory accounts of a phenomenon. Towards the end of their investigation, Bechtel and Richardson discuss the role of further kinds of constraints, among which they mention *phenomenological*, *operational*, and *physical* constraints. Phenomenological constraints exist because the way in which a phenomenon is characterized is often suggestive of particular explanatory models and potentially excludes certain possibilities. Operational constraints are determined by the available experimental procedures and material systems. They force us to build our theoretical models on the basis of the kinds of observations that we can obtain. Finally, physical constraints are given by the background knowledge about the physical realization of the lower level components. This background knowledge does obviously not only come from physics, it also encompasses firmly established and sufficiently general insights from biochemistry or from molecular biology itself.

Many of the ideas in Lindley Darden's (2006) collection of articles can be seen as investigating the role of these different kinds of constraints in concrete examples. In collaboration with Carl Craver she discusses specific cases studies from molecular biology and neuroscience, and together they identify some of the more specific strategies of mechanistic reasoning involved in these disciplines. I will later take her discussion of the dis-

covery of the mechanism of protein synthesis as an entry point for my own analysis, in which I try to go further and identify some additional features that will be relevant for my discussion of systems biology.

One of the starting points of Darden and Craver's reasoning is the insight that the detailed analysis of mechanisms can reveal constraints on their discovery (Darden 2006, Chapter 2; published earlier as Craver and Darden 2001). Discovery is conceived as a gradual and piecemeal process in which initially rough sketches are elaborated into increasingly detailed mechanistic accounts. Bechtel and Richardson's ideas on discovery are essentially accepted, but the authors hold that

> the contribution remains incomplete without a careful look at the products of this discovery process. Thinking carefully about mechanisms and especially their organization highlights a broad variety of constraints on their discovery in addition to those that come from localizing and decomposing. (Darden 2006, 49)

Mechanisms are characterized as collections of *entities* and *activities* (Darden 2006, Chapter 1; published earlier as Machamer et al. 2000). Entities are essentially the relevant structural components, while activities are the behaviors in which these entities can engage in the context of the mechanism. A given scientific field at a given time has a "store" of established entities and activities out of which (accounts of) mechanisms can be assembled (Craver and Darden 2001). Established entities figuring in explanations in molecular biology, for instance, are macromolecules, ions, cellular structures etc.; examples of activities are covalent bonding, lock-and-key binding in enzymatic reactions, or conformational changes. The background knowledge about these entities and activities already imposes constraints on the kinds of mechanisms in which they appear. However, the key concept for Darden is *productive continuity*, which captures the idea that when trying to understand a mechanism, scientists often look for intermediate steps in a chain that connects an input event to an observed output event. When studying a signal transduction pathway, for instance, molecular biologists often start with knowledge about the extracellular signaling molecule (ligand) and about the cellular reaction it triggers, and they ultimately want to identify all the intermediate molecules and reactions involved. An adequate account of the mechanism must show how each stage of the process produces the

next without leaving any gaps. Productive continuity is explicitly related to ideas about the temporal asymmetry of causality, and the flows of energy, matter, and information in biological processes (Darden 2006, Chapter 3).

The goal of eliminating gaps, or, more generally 'black boxes,' guides the process of discovery. Biologists start with incomplete models and gradually add more detail until they reach productive continuity. Productive continuity is relevant for the explanation of a phenomenon since it makes the working of the underlying mechanism *intelligible*. The existence of gaps or black boxes simply means that there is some part of the mechanism that we do not yet fully grasp. In this context Darden and Craver specifically identify the strategies of *schema instantiation* and *forward/backward chaining*. In the former one first proposes a relatively abstract description of a mechanism, that is, a *mechanism schema*, and then searches for components that fit the placeholders in this description. This strategy is maybe best considered as a special case of Bechtel and Richardson's localization strategy. Mechanism schemas are often derived by abstraction from existing mechanistic accounts of better understood phenomena; thus analogical reasoning plays an important role in this strategy. In forward/backward chaining, by contrast, one starts from already known, or hypothesized, components and then attempts to work forward or backward, taking advantage of constraints that the components impose on the possible ways of filling the gaps. Consider again the example of signal transduction. If molecular biologists find out, for instance, that the receptor that the ligand binds to belongs to the class of receptor tyrosine kinases, this tells them that the next step in the chain must be a protein that can bind to the specific sites that are created as a result of receptor activation. In this way by chaining forward, or backward, through the process they are often able to figure out the whole cascade.

It should be obvious that these additional strategies are restricted to the realm of well-behaved, nearly decomposable systems. Moreover, Darden and Craver exclusively discuss cases of linear (i.e. sequential) organization. Their emphasis on productive continuity suggests that they take such sequential organization to be a predominant feature of the systems they consider, and they do not propose any strategies for cases with more complex organization. In general, they do not discuss the heuristic character of their strategies, and their analysis of constraints gives the impression that these play an exclusively

beneficial role in the discovery process:

> Constraints determine the shape of the space of hypothesized mechanisms. Most simplistically, this space can be understood as a tree with terminal nodes representing possible mechanism schemata for the phenomenon to be explained. The addition of constraints prunes the tree or changes the weights on different branches. (Darden 2006, 48)

Scientists, in this picture, proceed by pruning the tree of hypotheses until, in the most favorable case, only one account, the *actual* mechanism, is retained. In the same context Darden also speaks of a process of "iterative refinement" (Darden 2006, 272). Even though she acknowledges that scientists are often mistaken about particular elements of a mechanism, or about what particular further constraints these elements impose, she does not discuss the systematic errors that might be introduced whenever strategies are applied that are heuristic in nature. The metaphor of pruning the tree suggests instead that scientists are in possession of a determinate algorithm that will eventually lead them to the right solution: Cut branches away in a particular order until you end up with the right one. In practice, however, scientists must often get rid of a lot of branches in the beginning before they can even get an overview of the tree's structure.

But even if in Darden's account an explicit discussion of the limits of heuristic reasoning strategies is missing, her description still captures some important aspects of discovery in molecular biology. The fact that these strategies are fallible and rely on certain assumptions about the organization of the underlying system will be important when it comes to alternative proposals from the side of systems biology. In the reminder of this chapter, I want to discuss concrete examples to illustrate the specific reasoning strategies of molecular biology and illuminate (though in a very crude way) the historical context in which they have to be understood.

2.3 Example: The Mechanism of Protein Synthesis

In this section I want to discuss an episode from the early period of molecular genetics around the middle of the 20th century, at the center of which, of course, lies the discovery of the structure of DNA by Watson and Crick in 1953. More specifically, I want to look

at the discovery of the mechanism of protein synthesis. The role of this example within the larger context of my project is twofold. First, it serves to illustrate how the strategies introduced in the previous sections are actually applied in practice—and to thereby also show the adequacy of the general framework of heuristics. It has to be noted, however, that precisely due to its impact on subsequent research, one cannot take this episode as representative for how discovery proceeds in molecular biology in general. I think, however, that the main features of a particular 'molecular vision of life' (Kay 1993) crystallized around this period, and these can be illustrated well by looking at some of the steps of this particular episode of discovery. Therefore, the main reason to discuss this example is to motivate a characterization of molecular biology as a specific research program. The early work in molecular genetics not only created important and fundamental knowledge on which generations of biologists could build subsequently, it also suggested a particular perspective on the organization of living systems that has expressed itself in very specific research strategies up to this day. This is partly a historical claim, and to properly substantiate such a claim here is both too large a job and beyond my expertise. However, I think that it is made sufficiently plausible by observing how these strategies have continued to shape discovery in molecular biology throughout the second half of the twentieth century. For this reason, I will look at a more recent episode of discovery in the next section.

The discovery of protein synthesis has been described as a process in which a central problem was attacked by two "local traditions" (Burian 1993) from different starting points and eventually culminated in the convergence of the two approaches (Rheinberger 1997b, Morange 1998, Darden 2006). The biochemist Paul Zamecnik, one of the prominent figures involved in this process, has described it with the metaphor of building a tunnel by digging from two sides (Zamecnik 1962, 47). However, only with hindsight one can say that both groups of researchers were actually working on the same problem, and it is important to take into account the different perspectives from which they started. The group of early molecular biologists, inspired by the discovery of the double helix structure of DNA and fascinated by the idea of a genetic 'code,' framed the problem in terms of information transfer: How can a sequence of nucleotides in a string of DNA determine the assembly of a chain of amino acids in a particular protein? Biochemists like Paul Zamecnik, by contrast, approached protein synthesis as a chemical process that involves a

number of catalytic reactions with particular energy requirements and energy barriers.

The heuristics of decomposition and localization are clearly exemplified in this episode and expressed in the eventually accepted scheme in which a set of activities, self-replication, transcription, and translation, is localized in a set of macromolecular components: DNA, messenger RNA, and polypeptide. Initially, the finding that protein synthesis does not occur directly at the site of DNA in the nucleus—it was shown to occur in cell-free systems not containing DNA—revealed a failure of direct localization and motivated the search for a more complex organizational scheme consisting of a sequential process of intermediate steps. We will see, however, that early molecular biologists and biochemists conceived of this scheme in very different ways.

2.3.1 Early Molecular Biology and the Coding Problem

Early molecular biologists, like James Watson and Francis Crick, tried to understand protein synthesis by focusing on the role of genes. Their problem was, therefore, to understand how the order of nucleotide bases in a sequence of DNA is related to the structure of a protein.

George Gamow, an astrophysicist of Russian origin who was mostly ignorant of biology, suggested to Watson and Crick that their problem could be solved without performing any experiment, that is, without having to open the black box of possible chemical reactions that might figure as intermediate steps between DNA and protein (Morange 1998, Chapter 12). Thus they started from a 'coding hypothesis,' according to which the amino acid sequence of a particular protein is determined by a sequence of DNA. Crick summarized the situation in the following way:

> While the indirect evidence in favor of some relationship of this type is very suggestive, the direct evidence is fragmentary in the extreme, and nothing whatever is known about the actual mechanisms involved. It is possible, however, to consider the problem in an abstract way as that of translating from one language to another; that is, from the 4-letter language of the nucleic acids to the 20-letter language of the protein, without any detailed consideration of the chemical processes involved. This approach is often referred to as the coding problem. (Crick 1959, 35)

Lily Kay (2000) suggested that the attempts at solving the coding problem must be understood as a transformation of the biological problem of protein synthesis into a problem of information theory. The mathematical theory of information, developed mainly by Claude Shannon, was highly influential at the time, and together with cybernetics was one of the main theoretical resources for early molecular biology. But the early molecular biologists were in the end not able to solve the coding problem by relying on these theoretical tools alone. Crick eventually admitted that it was impossible, due to the lack of empirical constraints, to draw any definite conclusions on the nature of the code and to decide between different proposed coding schemes. As I discuss below, the coding problem could not be solved until the biochemical black box was finally opened. In spite of this failure of the theoretical approach, the framing of the problem in informational terms became entrenched, and information as a metaphor, or a "metaphor of a metaphor" (Kay 2000), has continued to guide the ideas of molecular biologists. The idea of a biological process as a 'flow of information' made biologists focus on very particular organizational schemes and, even though the concepts of information theory did not enable them to sufficiently constrain the explanatory problem by itself, it nevertheless carried important heuristic value. The molecular biologists' way of thinking in terms of large macromolecules and their informational content eventually played a significant role in 'cracking the code' by experimental means.

2.3.2 Localizing Energy: the Biochemical Perspective

The Biochemists started from a different store of entities and activities, focusing on the chemical structure and reaction schemes of smaller molecules that were found to be involved in the process, notably on peptides and covalent bonding reactions.

Their attempts of decomposing the problem were influenced by background knowledge that consisted mainly of findings on the structure of proteins. The model of the primary structure of proteins as a sequence of amino acids linked by peptide bonds went back to the ideas of Emil Fischer and Franz Hofmeister in the beginning of the 20th century. Another important resource was the work of Frederick Sanger (1941) who had shown in his analysis of the sequence of insulin that the order of amino acids in proteins does not follow a simple pattern, but that the amino acids instead are arranged in an irregular se-

quence which is exactly reproduced in each molecule. Biochemists, therefore, assumed that the amino acids are added one by one to the growing polypeptide chain via peptide bonding reactions. Working out the details of this reaction scheme implied a first task decomposition of the overall process of protein synthesis. Separating this step from the rest of the mechanism of protein synthesis confronted them with a much more manageable problem.

In her discussion of the same example, Darden (2006) describes this strategy as an instance of backward chaining since the biochemists were beginning with the output of the mechanism, looking for the steps in the process leading up to the end product. The reaction of peptide bonding was known to be an *endergonic* process, which means that it absorbs energy in the form of work and cannot occur spontaneously. For this reason biochemists had to figure out what makes the reaction energetically possible. Zamecnik's cell-free rat liver system was used in the search for an active intermediate with the hypothesized function of energizing the reaction. This intermediate turned out to be aminoacyl-adenylate, an activated amino acid carrying an energy rich adenine monophosphate group. In addition, Zamecnik's system allowed them to identify macromolecular complexes in the cytoplasm composed of RNA and protein, at the time called 'microsomes', that were recognized as the sites of polypeptide synthesis.

This story nicely illustrates the application of some of the reasoning strategies discussed earlier, as well as the role of constraints. Aside from making use of backward chaining, we have seen how biochemists sought to decompose the initial problem into more manageable chunks. A quote by Zamecnik, looking back in 1962 on the already accomplished work, illustrates the explicit way in which the strategies of decomposition and localization figured in his reasoning:

> As one contemplates the way in which protein molecules such as insulin or myoglobin are constructed, it appears that a number of steps must be involved in the process, and three separate questions may be posed for experimental attack. (Zamecnik 1962, 47)

Thus the particular characterization of the phenomenon, via known structural properties of the end product and available background knowledge, imposed important constraints on the task decomposition—phenomenological constraints in Bechtel and Richardson's

terminology. First, the fact that proteins were constituted by amino acid chains implied an activity that would provide the required energy to create peptide bonds. Second, the observation that the synthesized polypeptide chain had an irregular but nonrandom sequence suggested the existence of a process directing the incorporation of specific amino acids. Third, there had to be some way in which the genetic information travelled from the DNA in the nucleus to the site of protein synthesis.

Thus the three questions raised by Zamecnik concerned the localization of these activities in specific structural components of the system. We have seen that the first of them was to a large extent answered in his own lab. Attacking the other two questions, however, required the concerted effort of both biochemists and early molecular biologists.

2.3.3 Localizing Information: transfer RNA and messenger RNA

One of the two remaining problems was the question of how RNA might be involved in the assembly of the specific sequence in the growing polypeptide chain. While biochemists had not really considered the functional role of the RNA found in the microsomes, molecular biologists almost immediately grasped its potential as a carrier of genetic information.

The initial strategy to understand the possible role of RNA, pursued among others by James Watson, was to assume that RNA determined the specific sequence of proteins in a manner analogous to the way in which one of the strands in the DNA double helix determines the complementary sequence of the other. Thus, RNA molecules, due to their capacity to build weak hydrogen bonds with other molecules, were thought to provide a scaffold for the controlled incorporation of amino acids into the growing polypeptide chain. The data were not sufficient, however, to determine the structure of the RNA complex. Francis Crick, who had become skeptical about the idea of a structural template, instead proposed a different idea. According to his 'adaptor-hypothesis,' each amino acid is initially attached to a small molecule that can specifically bind to a coding template of RNA and thereby determine the future location of the amino acid in the polypeptide chain. The integration of molecular biology and biochemical reasoning finally took place when the people in Zamecnik's lab used Crick's hypothesis to interpret their discovery of small and soluble RNA (S-RNA) molecules that were different from the RNA found in the

microsomes. Strikingly, the S-RNA molecules were found to be covalently bound to amino acids. Moreover, it was subsequently revealed that there were 20 specific enzymes, each corresponding to one amino acid, that were responsible for catalyzing both the reaction to build the activated aminoacyl-adenylate and the subsequent fixation to soluble RNA.

According to the adaptor model, the instructions to build a protein did not depend on the specific three-dimensional structure of RNA, but was simply 'read' from its sequence. This suggested a powerful way to study the specific relationship between RNA and the sequence of amino acids. Thus the *in vitro* systems that had been devised for the biochemical study of protein synthesis could subsequently be turned into a tool to break the genetic code. Heinrich Matthaei and Marshal Nirenberg loaded their system with a synthetic poly-U nucleic acid; that is, with a chain of RNA consisting of only the nucleobase uracil (UUU…). They observed that the thereby synthesized polypeptide was made up of repeated instances of the amino acid phenylalanine: the first codon was deciphered. This opened the door to a further set of experiments in which the precise relation between the 4 nucleotides in RNA and the 20 amino acids in proteins was systematically determined.

The way in which the 'coding problem' was finally solved is often taken as evidence for the superiority of experimental approaches over purely theoretical speculations. However, as Michel Morange observes,

> Matthaei and Nirenberg had dared to take the idea of a genetic code to its logical conclusion and to try to determine this code experimentally, without worrying about the precise nature of the RNA involved in protein synthesis … . They also rejected the idea, deeply rooted in the biochemists' view of the world but rarely openly expressed, that the form of RNA molecules played an essential role in protein synthesis. (Morange 1998, 136–137)

The informational perspective had thus left its distinct traces in experimental practice. It had not been possible to crack the code by paper and pencil, but it was nevertheless possible to express the organizational scheme in purely informational terms, without reference to specific biochemical detail.

Now that the link between RNA and protein had been largely understood, the remaining task, according to Zamecnik's list, was to figure out how genetic information is transmitted from DNA to the RNA templates. This relationship was eventually unraveled by

the French researchers François Jacob and Jacques Monod. When they began their investigations, the general opinion was that the RNA found in the microsomal particles (today known as ribosomes) served as the natural template for protein synthesis. Experimenting on sexually reproducing bacteria, Jacob and Monod investigated the transfer of a chromosomal fragment carrying a gene for the enzyme β-galactosidase. They observed that as soon as the fragment entered a bacterium that previously had lacked the gene, the enzyme instantly started being synthesized at maximum rate. This was a puzzling result because no microsomal particles were transmitted in the process that could trigger the synthesis, and their assembly from scratch was thought to take much longer. Jacob and Monod, therefore, postulated the existence of an additional, short-lived intermediate form of RNA whose role it was to carry the information from the genes to the microsomal particles, a prediction that was eventually confirmed by the discovery of mRNA.

We have seen the extent to which the general heuristic strategies were applied in the discovery of the mechanism for protein synthesis. I quoted Paul Zamecnik, who explicitly discussed the decomposition of the process into different steps, and I described the efforts of localizing the corresponding activities that were undertaken by different groups of researchers. There is a further respect in which the problem was decomposed. As discussed, it was known from Sanger's work that proteins, in spite of their intricate three-dimensional configurations, were composed of linear chains of amino acids. In the beginning it was *not* known, however, that the assembly of the polypeptide chain and the folding into the three dimensional structure were actually two separate steps in the overall process. Francis Crick, in an article written before decisive progress on the coding problem had been made, emphasized that treating those steps as independent processes was a simplifying assumption:

> Our basic handicap at the moment is that we have no easy and precise technique with which to study how proteins are folded, whereas we can at least make some experimental approach to amino acid sequences. For this reason, if for no other, I shall ignore folding in what follows and concentrate on the determination of sequences. It is as well to realize, however, that the idea that the two processes can be considered separately is in itself an assumption. (Crick 1958, 144)

The black-boxing of the protein folding process was thus admittedly a heuristic move to make a complex problem more tractable and to bring it within the reach of available experimental methods. The assumption in the end turned out to be correct. The discovery of the mechanism of protein synthesis, however, was not merely a sequence of ingenious guesses. Many of the initially proposed ways to attack the problem turned out to be misguided, but they often revealed important clues about more adequate ways of structuring the epistemic task. It is important to highlight that considerable amounts of selective trial and error were involved in the discovery of the actual mechanism, but the errors are usually remembered less well.

2.3.4 From Specificity to Information

I discuss the example of protein synthesis not only because it illustrates the successful application of general heuristic strategies. One further aim is to show that this episode (among others, to be sure) has shaped the way in which scientific problems were framed in the subsequent development of molecular biology. By looking at the discovery of this fundamental mechanism, we can catch a glimpse of the origin of some of the more specific heuristics that have become part of the methodological toolkit of molecular biology.

As pointed out earlier, Darden's concept of productive continuity presupposes that living systems can be understood in terms of relatively simple functional schemes that essentially depict them as consisting of sequential processes. The emphasis, therefore, lies not so much on organization, but rather on the specific activities connecting the steps in such a process. However, before the intermediate steps were discovered, it was not obvious at all why the mechanism should be organized in such a simple way. Before Zamecnik's group produced their results, for instance, many biochemists had assumed that protein synthesis must be understood as the reversal of proteolysis, the breakdown of proteins. The role of genes in directing or controlling this process was far from clear. The specific conception of productive continuity, in terms of a "flow of information" (Crick 1958), had certainly not been guiding the process of discovery from the beginning. The idea that genes influenced the conformation of proteins directly by means of their three dimensional structure was gradually replaced by the conception of an information transfer that was in a certain sense *independent* of the underlying biochemical reactions. Michel

Morange describes this development as the "break between form and information" that "made it possible … for molecular biology to come of age" (Morange 1998, 149). It motivated a general conception of living processes in terms of linear information flow.

This is maybe best illustrated by comparing Crick's adaptor hypothesis with the earlier idea that proteins were formed by using RNA as a structural template. As I discussed, Crick's guess was that there would be a set of 'adaptor molecules' providing specific connections between each amino acid and its corresponding nucleotide codon. The link between RNA triplet and amino acid was contingent from a biochemical point of view. The group of Seymour Benzer tested Crick's hypothesis in an elegant study (Chapeville et al. 1962). It was known that the S-RNA molecules found by Zamecnik's group consisted of an amino acid covalently bound to a sequence of RNA. Thus, they created 'artificial' adaptor molecules by transforming the amino acid (cysteine into alanine) of one species of S-RNA while leaving its RNA part intact. With the same setup that Matthaei and Nirenberg had previously used to crack the code, they showed that the system now incorporated alanine into the polypeptide chain when stimulated with a template normally coding for cysteine. The position of amino acids in protein molecules thus depended solely on the sequence of the coding template, not on the biochemical properties of the amino acid itself. This showed that the genetic code was *arbitrary* from a biochemical point of view, and it also explained why purely theoretical approaches to the coding problem were doomed to failure.

However, the role of biochemistry in this development is ambivalent. On one hand, as has been pointed out by historians, the research pathways leading to the solution of the coding problem turned out to be "far more biochemical and far less theory-driven than … anticipated" (Burian 1993, 401–402). The constraints that allowed the researchers to work out the details of the mechanism were to a large extent derived from the results of specific biochemical experiments, and not so much from information theoretical reasoning. On the other hand, those very results turned out to limit the perceived importance of biochemistry, as a theoretical resource, for the subsequent development of molecular biology. The complex metabolic reaction schemes that biochemists had studied in the early twentieth century, such as the integrated system responsible for fermentation (discussed in Bechtel and Richardson 1993, Chapter 7), did not seem to provide the right exemplars

to illuminate the information transmitting mechanisms of molecular genetics. Instead, the role of biochemistry was largely reduced to the study of the specific reactions occurring in individual steps within such processes. The concept of biochemical specificity certainly continued to play a crucial role in molecular biology, and it supplied substantial heuristic power by reducing the expected number of important interactions occurring in a biological system. Yet, from an informational perspective, the main role of biochemistry was now to explain how a signal was transmitted from one component to the next. It had no bearing on the general route of the signal and its significance for the rest of the system. Jacques Monod captured the independence of the informational pathways from the chemical nature of the underlying signals with his concept of "gratuité" (gratuity): "Physiologically useful or 'rational', this relation is chemically arbitrary—'gratuitous', one might say" (Monod 1971, 77).

There is another way to appreciate how the informational perspective detached the organizational schemes of molecular biology from the theoretical framework of biochemistry. This connects the discussion with the question that Weaver's quote raised in the beginning of this chapter. Quantitative aspects, concentrations, kinetic parameters etc., had always been of crucial importance in many applications of biochemistry. Take as an example the well-known Michaelis-Menten model of enzyme kinetics. It describes the process in which an enzyme converts a substrate by forming an intermediate complex. Even though a qualitative account of how one single molecule of substrate binds to one molecule of enzyme, and how the former is subsequently converted, may partly illuminate the process, it completely neglects the kinetic process that has to be described at the population level.[4] In order to understand how the presence of the substrate affects the amount of product, one has to apply subtle mathematical methods, and in order make predictions, one needs precise quantitative measurements of the required kinetic parameters (Gunawardena 2012b). Hence, it is an essential feature of the model that it describes the dynamics of *populations*, or concentrations, of molecules.

The explanatory schemes of molecular biology, by contrast, typically do without any quantitative features. To understand the relevant aspects of the mechanism of protein

[4] If I speak of large sets of molecules 'populations,' it is mainly for lack of a better term. I do not want to suggest any strong analogies to the populations of organisms dealt with in ecology and evolutionary theory. My usage is thus closer to the idea of a 'statistical population.'

synthesis, for example, we do not have to know how many molecules are turned over or what the initial concentrations are. We can explain the process by restricting our description to the level of individual molecules, and this is precisely what is done in the typical cartoons of molecular biology. Biochemical complexity is expected only in the specific interactions of individual macromolecules. A fundamental assumption underlying the mechanistic schemes of molecular biology, therefore, is that *the individual molecule is sufficient to represent the population.* Before coming back to this point in more detail, however, let me discuss a more recent example from research in molecular biology.

2.4 Example: The Spindle Assembly Checkpoint

As Sahotra Sarkar observes, "research in molecular biology has always favoured research programs that attempted to push forward its frontiers" (Sarkar 1996, 8). If we want to understand what distinguishes molecular biology from systems biology, we have to take into account the dynamic nature of scientific fields and disciplines, and we cannot assume that a characterization of molecular biology based on a research episode from the middle of the twentieth century will be sufficient for our purposes. For this reason, I will now turn to a more recent example: the study of the spindle assembly checkpoint. Important progress regarding the mode of operation of this important cell cycle control mechanism has been made during the last two decades.

There are several reasons why the investigation of the spindle assembly checkpoint provides a good case study in the general context of my project. First, this mechanism relates to the fundamental principles of cellular regulation, and its basic properties are conserved in eukaryotic species from yeast to human. Thus the discussion can both reveal how scientists conceive of individual molecular mechanisms and also illuminate their perspective on how these mechanisms hang together in the systemic context of the cell as a whole. Moreover, the discovery of this mechanism has brought together different experimental approaches, such as structural and cytological studies, *in vitro* biochemistry, and genetic techniques, that have been applied across a variety of different organisms. This work can, therefore, be taken as representative of a wide area of research in molecular biology. The most important reason for this choice, however, is the fact that the spin-

dle checkpoint mechanism has recently also been approached from a systems biology perspective. The following discussion thus provides a natural entry point into a deeper analysis of the differences between traditional molecular biology and systems biology.

2.4.1 Discovery of the Spindle Assembly Checkpoint

All living things are composed of cells, and all cells arise by the division of preexisting cells. In eukaryotes, each daughter cell contains a set of chromosomes like that of the mother cell. The eukaryotic cell cycle consists of a series of distinct stages that are highly regulated in order to guarantee the correct duplication of the hereditary material and subsequent division of cells (e.g. Morgan 2007). The concept of a 'cell cycle checkpoint' is based on the idea that these processes do not simply unfold independently,[5] but that there are additional control mechanisms that ensure that later events are dependent on the completion of earlier events. For example, cells do not enter the mitotic phase of the cycle, in which division takes place, if DNA synthesis has not been completed (Hartwell and Weinert 1989).

Another crucial step of the cell cycle, within mitosis, is the transition from *metaphase* to *anaphase*, in which the duplicated chromosomes have to be distributed correctly to mother and daughter cells. The spindle assembly checkpoint monitors the fidelity of chromosome transmission in this process. After the chromosomes have been replicated, they are condensed into joined pairs, the so-called *sister chromatids*, held together by a protein complex called *cohesin*. In metaphase all chromatid pairs are aligned in the central region of the cell, while in anaphase they are separated and pulled apart toward opposite poles. Both alignment and segregation of chromosomes are carried out by the *mitotic spindle*, a sub-cellular structure that mainly consists of long cylindrical polymers called *microtubules*. The microtubules emanate from opposite poles of the cells and can attach to the individual chromatids.

Once the sister chromatids in one pair are attached to microtubules coming from opposite directions (bipolar attachment), they move together towards the equatorial plane of the cell because the dynamic properties of the microtubules create tension. Anaphase

[5] In early embryonic divisions of the frog *Xenopus* the events of the cell cycle apparently occur independently of one another, that is, without extrinsic feedback control mechanisms, and are driven only by the cyclic activation of particular protein complexes. For details see Murray and Kirschner (1989).

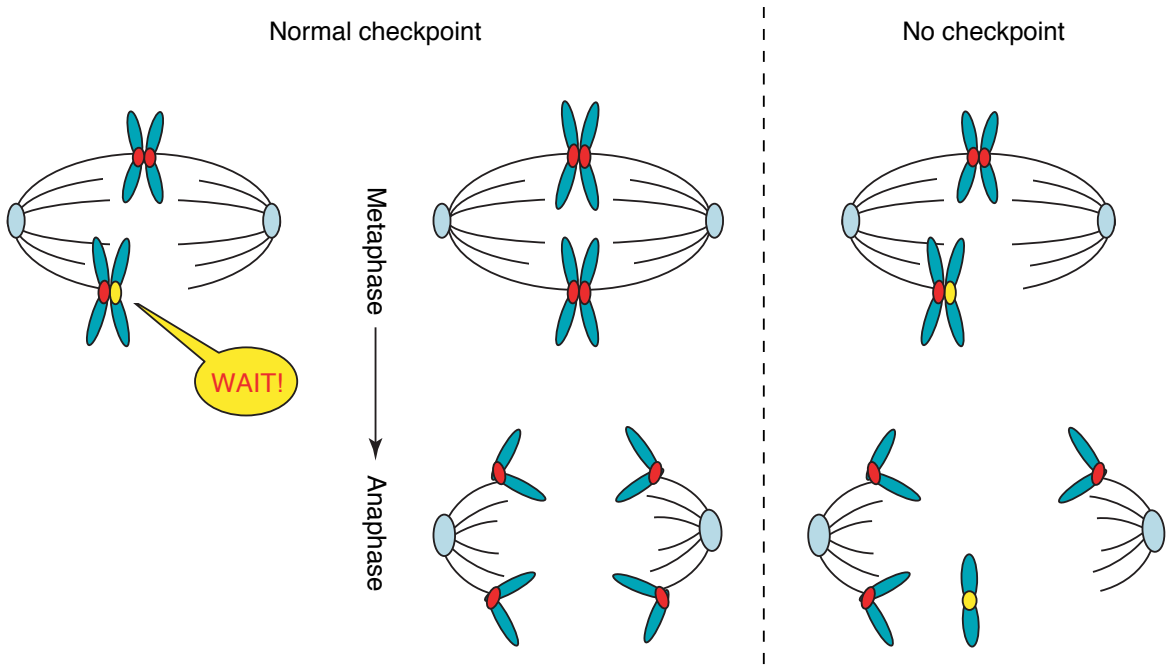


Figure 2.1: Basic behavior of the spindle assembly checkpoint. For further explanation see text. *Source:* Karess (2005).

starts when the link between the sisters is severed and the chromatids are pulled in opposite directions. The spindle assembly checkpoint is a surveillance mechanism that delays the initiation of chromatid separation until all chromosomes are properly attached to the spindle. Thereby it prevents the premature start of anaphase which might lead to infidelity in the distribution of chromosomes and, as a result, to a decrease in fitness or death, or to genetic disease in the case of multicellular organisms (Figure 2.1).

Evidence for the idea that the transition from metaphase to anaphase involves active regulation has existed for a long time. J. G. Carlson, based on cytological examination of mitosis in grasshopper neuroblasts, suggested in 1956 that this transition was regulated by chromosomes. He observed that right before metaphase individual chromosomes occasionally moved away from the others, eventually returning to the plane of alignment. He reports that,

> by watching this movement, one can predict exactly when anaphase will begin, for it never starts until all the chromosomes are in the equatorial plate, and it always starts as soon as the last one has reached it. (Carlson 1956, cited in Lew and Burke 2003, 252-253)

Systematic experiments in support of Carlson's observations were performed by Zirkle (1970), who managed to selectively destroy metaphase spindles with ultraviolet radiation.

Using this technique, he was able to artificially delay the beginning of anaphase. He concluded that the arrival of the last chromosome on the metaphase plate acts as a trigger for anaphase onset (as cited in Lew and Burke 2003, 253). However, substantial steps beyond this observational evidence were not made until two decades later. In 1989, Hartwell and Weinert had proposed the idea of a 'checkpoint mechanism' as a general schema to explain the regulation of cell cycle events. They argued that the dependence of a later stage on the completion of an earlier one might be either due to 'substrate-product order' or to a 'checkpoint mechanism.' In case of the former, a series of events is ordered by a "principle intrinsic to the components themselves" (Hartwell and Weinert 1989, 630). This kind of principle is found, for example, in the formation of the bacteriophage T4:

> All structural proteins are synthesized at the same time, and unassembled proteins remain unassociated until the partially assembled structure becomes ready for their addition. (Hartwell and Weinert 1989, 630)

Hence, some structural units act as necessary substrates for the assemblage and addition of others, which enforces a temporal order (Figure 2.2, **A**). In a process regulated by a checkpoint mechanism, by contrast, a later stage is *actively inhibited* until the earlier stage has been completed (Figure 2.2, **B** and **C**).[6] Hartwell and Weinert conclude that in order to establish the presence of a checkpoint mechanism, it is not sufficient to observe that one event only occurs after another one has been completed. But they describe a way in which one might distinguish experimentally between the two scenarios:

> The existence of a control mechanism is suggested when one finds chemicals, mutants, or other conditions that relieve a dependent relationship; that is, conditions that permit a late event to occur even when an early, normally prerequisite event, is prevented. (Hartwell and Weinert 1989, 630)

Thus, in contrast to the case of substrate-product order, it should in principle be possible to remove a dependency that is due to a checkpoint mechanism by interfering with the control (Figure 2.2, **D**).

[6] Hartwell and Weinert mention that control might also work by activation, for instance if the completion of DNA synthesis produces an activator of mitosis. However, they write: "Since it is difficult to distinguish control by activation from substrate-order by an empirical test, we will concentrate our discussion of control mechanisms on those that act by inhibition" (Hartwell and Weinert 1989, 630).

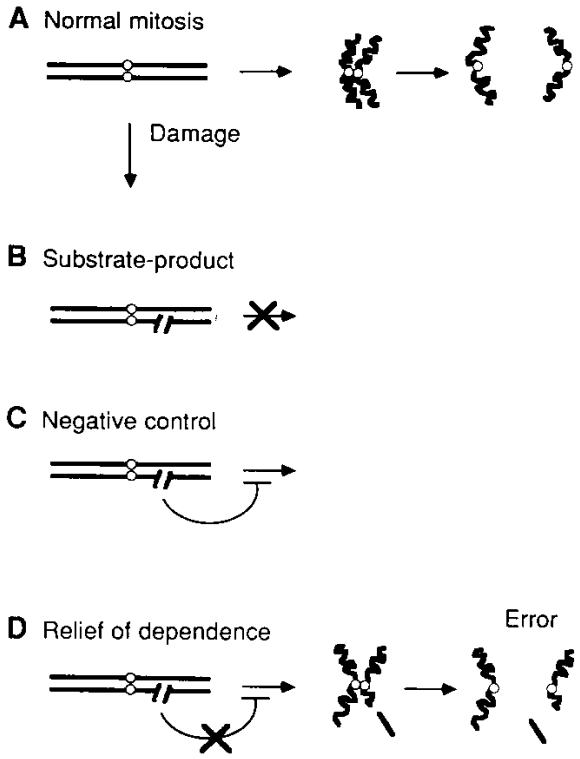


Figure 2.2: Different types of process ordering, illustrated with the example of DNA damage control. (**A**) In normal mitosis the chromosomes are condensed and segregated after successful replication. (**B**) Substrate-product order would exist if mitosis is blocked because the damaged chromosome is an inadequate substrate for chromosome condensation. (**C**) Checkpoint control: DNA damage creates a signal that inhibits chromosome condensation. (**D**) In case of checkpoint control: if the negative inhibition is removed, the damaged chromosome can pass through mitosis. *Source:* Hartwell and Weinert (1989).

The observed delay of chromosomal segregation in the metaphase-anaphase transition strongly indicated the existence of a checkpoint mechanism, and Hartwell and Weinert's article suggested an experimental strategy. By making use of mutagenic screening techniques in the budding yeast *Saccharomyces cerevisiae*, two independent studies subsequently identified several genes whose disruption caused cells to resume the cell cycle, even in conditions in which normal cells would arrest in metaphase. Hoyt et al. (1991) found three genes that, when mutated, enabled yeast cells to build a new bud, and thus to enter a new cell cycle, even when spindle assembly was prevented by the microtubule inhibitor benzimidazole. They called these genes Bub1, Bub2, and Bub3 (for **B**udding **U**ninhibited in **B**enzimidazole). In a very similar experiment, Li and Murray (1991) identified three further genes, Mad1, Mad2, and Mad3 (for **M**itotic **A**rrest **D**eficient), as being implicated in the purported control mechanism.

Independently from these genetic investigations, there were attempts to uncover how exactly the putative checkpoint mechanism instantiates Hartwell and Weinert's general

scheme. In order to delay anaphase, the system must somehow 'sense' that the attachment process has not been completed. McIntosh (1991) put forward the hypothesis that the chromosomes that are not correctly attached produce a signal that delays anaphase:

> An effective system for sensing when all the chromosomes are appropriately associated with the spindle is more likely to assess the chromosomes that are not attached than those which have already done so. Clearly, it is easier to distinguish one or more chromosomes unattached from none unattached than it is to discriminate between 45 attached and 46. Furthermore, many aneuploid cell lines show perfectly normal mitosis with the chromosomes they do have, suggesting that an absolute count of the chromosomes attached to the spindle would be insufficient information from which to make the decision to proceed with anaphase. It would seem that the cell has a way to detect unattached chromosomes. Presumably, these chromosomes emit a signal that tells the rest of the cell to delay anaphase onset. (McIntosh 1991, 617)

The group around the cancer researcher Conly Rieder set out to localize the source of this signal. Their sophisticated microscopic studies revealed that *kinetochores* play an important role in the regulation of anaphase onset. Kinetochores are large protein complexes that assemble in the central region of the chromosomes and provide 'docking stations' for microtubules. By means of a laser micro-beam, Rieder et al. were able to selectively destroy specific areas on chromosomes in living rat kangaroo cells. Their results showed that chromosomes whose kinetochores are destroyed by the laser are no longer able to delay anaphase. From this they concluded that "molecules in or near the unattached kinetochore ... inhibit the metaphase-anaphase transition" (Rieder et al. 1995, 941). Their observations also revealed the striking result that one single unattached kinetochore is sufficient to keep the whole cell arrested.

There was thus evidence for some of the molecular players of the putative mechanism, and there were independent results pointing to the kinetochores as the location involved in checkpoint signaling. The different strands of research were brought together for the first time when the behavior of homologs of the Mad2 protein was investigated in human cells and frog cell extracts (Li and Benezra 1996, Chen et al. 1996). Both studies showed that the respective proteins localize at the kinetochores after chromosome condensation,

but disappear when the attachment of microtubules is completed. In the following years, most of the other proteins identified in the initial yeast screens could be shown to localize at the kinetochores as well.

Subsequent research efforts were directed at understanding how the concerted action of the identified checkpoint proteins can delay the onset of anaphase. It was known that the progression from metaphase to anaphase in a regular cell cycle depends on the enzymatic activity of a large protein complex called *anaphase promoting complex/cyclosome* (APC/C). This complex is responsible for the destruction of the cohesin rings that hold the sister chromatids together and for the degradation of *cyclins*, which are proteins that drive the progression through the cell cycle and whose loss triggers the program of mitotic exit. However, in order to perform these activities, the APC/C needs to be activated by binding to another protein called Cdc20.

It was, therefore, reasonable to assume that the putative checkpoint mechanism would delay the onset of anaphase by inhibiting the activity of APC/C, either directly or by interfering with its activator Cdc20. By means of different techniques for the detection of protein-protein interactions, such as the yeast two-hybrid system and co-immunoprecipitation, Hwang et al. (1998) could show that the checkpoint proteins Mad1, Mad2, and Mad3 indeed all interact with Cdc20. Further studies suggested that especially the interaction between Mad2 and Cdc20 was crucial for the activity of the checkpoint. Howell et al. (2000) proposed a model according to which the unattached kinetochore serves as a catalytic site for the assembly of a Mad2-Cdc20 complex that sequesters Cdc20 and thereby prevents it from activating APC/C. By the turn of the century the basic scheme of the mechanism was largely agreed upon, and a review in the journal *Cell* summarized:

> The basic plan of the signaling cascade is now well established. Central to the spindle checkpoint is the kinetochore. Prior to spindle attachment, kinetochores generate a diffusible 'wait anaphase' signal, which inhibits the anaphase promoting complex/cyclosome .... As a kinetochore binds microtubules ... its wait signal generator is silenced and the inhibition of anaphase is released. (Shah and Cleveland 2000, 997)

However, important questions remained, notably concerning the recruitment of the checkpoint proteins to the kinetochores, the exact mode of APC/C inhibition, and the deac-

tivation of the signal after the attachment. The "second decade of checkpoint studies" (Musacchio 2011) has given at least partial answers to some of these questions.

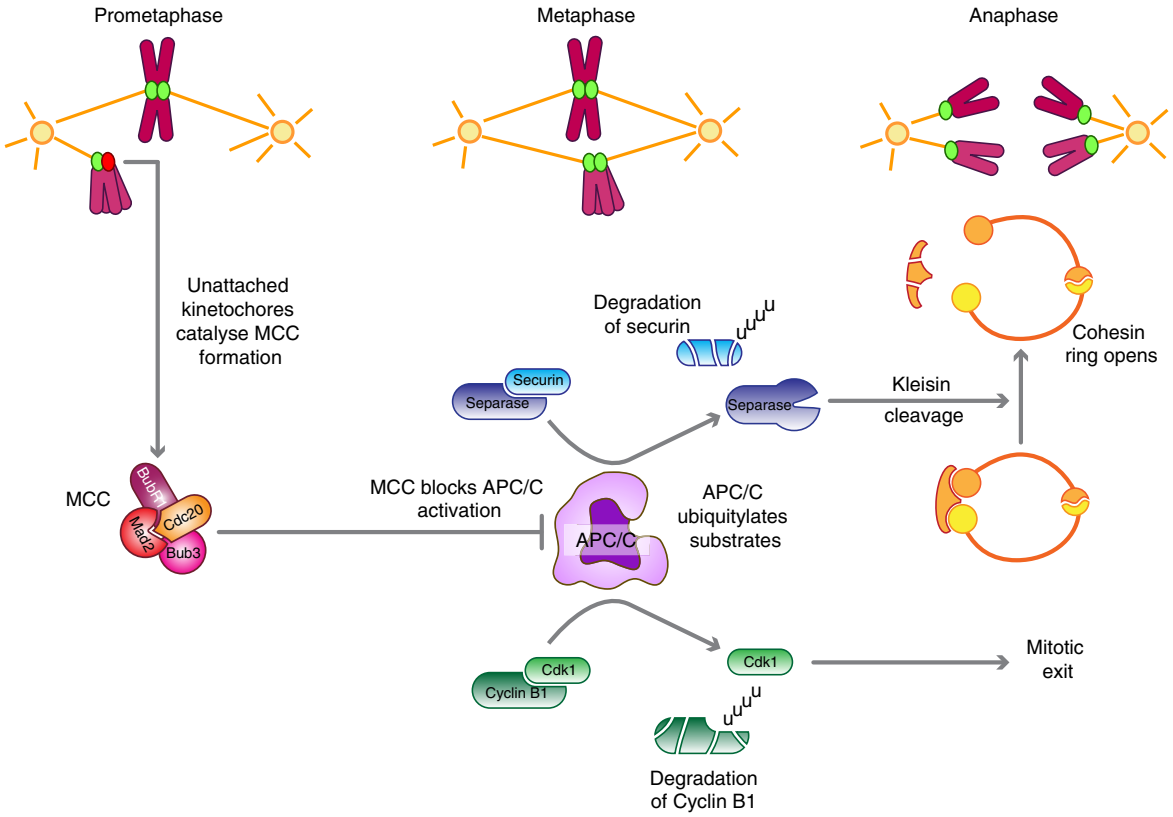


Figure 2.3: Representation of the spindle assembly checkpoint mechanism adopted from a recent review article. The figure also represents some of the downstream effects of APC/C activation. For further explanation see text. *Source:* Lara-Gonzalez et al. (2012).

It was shown, for example, that other checkpoint proteins apart from Mad2 are involved in forming an inhibiting complex with Cdc20. Sudakin et al. (2001) found in the human HeLa-cell line a factor that they called *mitotic checkpoint complex* (MCC). This complex contained the homologs of the Cdc20, Mad2, Mad3, and Bub3 proteins and was shown to bind to the APC/C, thereby preventing its enzymatic activity. Against expectations, however, they also reported that this complex was not only generated at the kinetochores and present also in cells before they enter mitosis, a finding that has led to competing hypotheses about how kinetochores exactly contribute to the inhibition of the APC/C.

By means of a sophisticated photobleaching technique, Shah et al. (2004) were able to monitor the turnover of proteins at the kinetochores. Their results suggested that Bub1, Mad1, and a portion of Mad2 are stably bound to unattached kinetochores, in line with the idea of a catalytic platform. Subsequently, De Antoni et al. (2005) proposed a 'template

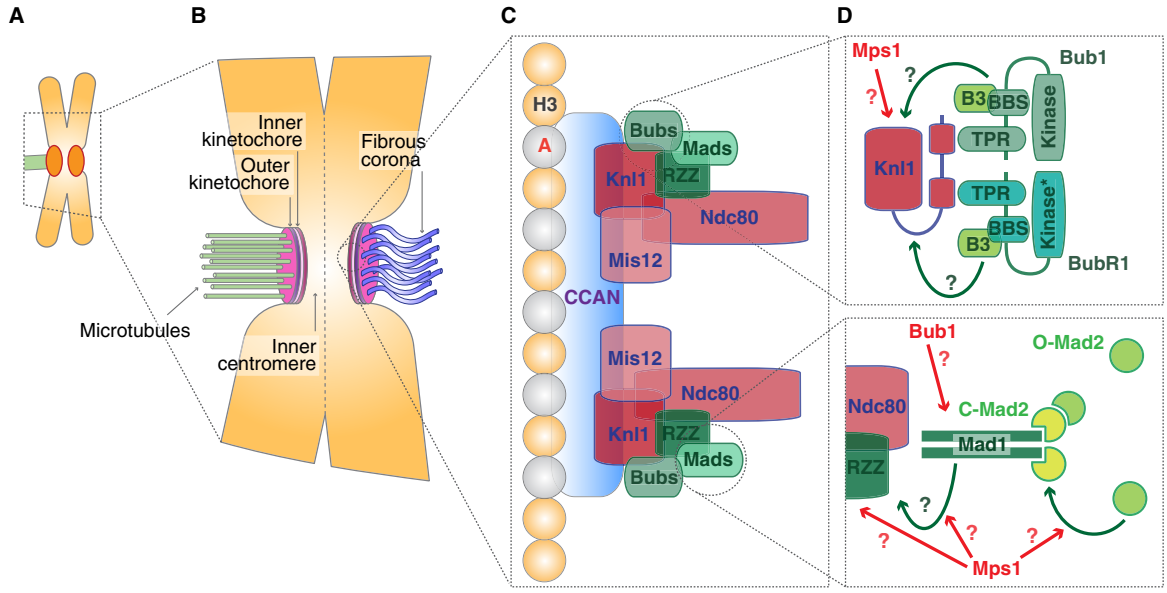


Figure 2.4: Figure adopted from a recent review article that illustrates, with increasing zoom levels, the amount of detail that has been accumulated about the SAC components. The two panels on the right (D) represent hypothetical models for the recruitment of checkpoint proteins. Question marks highlight interactions that have not been established, yet. *Source:* Lara-Gonzalez et al. (2012).

model,' according to which the binding of free Mad2 to Cdc20 is catalyzed by a Mad1-Mad2 complex that is stably bound to the kinetochores. This model is by now widely accepted (Lara-Gonzalez et al. 2012).

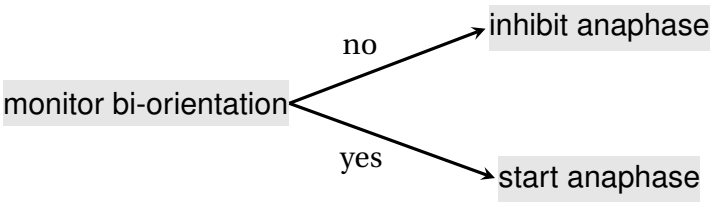
In recent years, an important strand of research has consisted in studying the function of the kinetochores and their role in recruiting checkpoint proteins and monitoring the attachment of microtubules. Here, biochemical studies have led to increasingly complex models of kinetochore structure, but many open questions remain (Figure 2.4). Progress in this context is expected to arise from even more detailed structural investigations. As a recent review suggests:

> [T]he relationship between kinetochores and checkpoint control remains poorly understood. Crucial advances in this area in the third decade of checkpoint studies (2011–2020) are likely to be brought about by the characterization of the mechanism of kinetochore recruitment, activation and inactivation of checkpoint proteins, which remains elusive for the majority of checkpoint components. (Musacchio 2011, 3595)

2.4.2 Analysis of the Example

Just as in the example of protein synthesis, the basic strategies of decomposition and localization can be seen at work in the discovery of the spindle assembly checkpoint mechanism. The initial explanandum phenomenon consisted in the observed capacity of the cell to delay the onset of anaphase until all chromosomes are properly bi-oriented. Hartwell and Weinert's general analysis of cell cycle regulation provided a possible decomposition of the cell's capacity to delay anaphase in terms of an extrinsic control mechanism. Although this move did not yet define the locus of control of the phenomenon, it nevertheless suggested that one could focus on a separate set of components and would not have to take into account the cell cycle machinery as a whole. Furthermore, the idea of a checkpoint immediately suggested an experimental test and a strategy of search for the underlying components.

The initially proposed schema can be represented in the following way:



From this description it should be obvious that the sub-operations are not independent, but produce the phenomenon only when properly connected. Thus the organization of the proposed mechanism, already in this crude form, goes beyond what Bechtel and Richardson call 'direct localization.' This is also the reason why the strategy to localize components of the mechanism by screening for 'checkpoint-deficient' yeast mutants could not discriminate between different sub-operations of the mechanism. Indeed, it turned out that some of the genes identified in the initial screens are involved in monitoring (Bub1, Mad1), while others are (mainly) involved in inhibition (Mad2, Mad3, Bub3).[7] Note, moreover, that this scheme is ambiguous. It is not obvious, for example, whether 'allow anaphase' is an additional component operation, or whether it simply consists in the removal of the inhibitory activity. Nevertheless, the proposed decomposition allowed re-

[7] It turns out that Bub2 is not involved in the spindle assembly checkpoint, but in a different checkpoint mechanism that ensures the correct positioning of the spindle before division.

searchers to focus on particular aspects of the problem, thus reducing its epistemic complexity.

The example also fits very well with the strategies discussed by Darden (and Craver). Notably, it can be interpreted as a clear case of 'schema instantiation.' Researchers started with an abstract 'mechanism schema' and subsequently filled in increasing amount of detail. Initially, the component operations were simply 'black boxes,' and an important part of the work consisted in opening these boxes by accumulating molecular knowledge.

The efforts to localize the components of the mechanism can be described, just as in the case of protein synthesis, as proceeding from opposite directions, thereby instantiating Darden's strategy of forward/backward chaining. Forward chaining can be found in the investigations of Rieder et al. who managed to localize the monitoring function to the unattached kinetochores. The fact that single molecular complexes, that are very small compared to the volume of the whole cell, are able to keep the cell cycle arrested, led to the idea that the kinetochores produce a diffusible signal that carries out its inhibitory function in the cytosol. Thus, this strand of research was reasoning forward, trying to fill in the molecular details into the scheme 'from left to right.' Other groups, by contrast, such as Hwang et al., started from the other end of the chain and focused on the possible interaction of the checkpoint mechanism with the cell cycle machinery. This endeavor revealed the interaction of Mad2 and Cdc20 and subsequently led to the discovery of the mitotic checkpoint complex. As we have seen, the gap between the two strands of research was rather small since most of the molecules involved in inhibition were found to localize at the kinetochores.

Thus, once the "basic plan of the signaling cascade" (Shah and Cleveland 2000) was established, the researchers could focus on a set of manageable explanatory sub-tasks corresponding to the individual sub-operations of the mechanism. Among these were the following questions: How does the kinetochore detect whether microtubules are attached in the correct way? How is the inhibitory complex (MCC) assembled? How does it inhibit the enzymatic activity of the anaphase promoting complex (APC/C)? How is the inhibition removed when all chromosomes are correctly attached? The explanation of these sub-operations required them to go to lower levels and, in particular, to investigate the inner structure of macromolecular complexes. This problem decomposition was furthermore

reflected in a division of labor into different research projects. Most publications, apart from review articles, focused on contributing to only one of the mentioned subtasks.

Apart from the use of the basic heuristic strategies, there is a further parallel between this example and the discovery of the mechanism of protein synthesis. This parallel lies in the use of a particular informational perspective. We said earlier that explanations of the basic mechanisms of molecular genetics, such as the one discussed in Section 2.3, relied on a conceptual detachment of the organization of informational pathways from the underlying biochemical processes, and we made reference to Jacques Monod's notion of 'gratuity' in this context. Biochemistry was invoked to explain how one step in the informational chain leads to the next, but the kinetic aspects of biochemical reactions were not considered relevant for the explanation of these mechanisms. Something very similar can be observed in this example. Biochemistry plays an important role in describing the structure of macromolecular complexes and individual reactions. However, these investigations are used exclusively to fill in the black boxes corresponding to the single elements in the signaling cascade. This cascade itself is represented in purely qualitative terms and does not rely on any detailed kinetic information about the occurring biochemical reactions. Quantitative information about the dynamical nature of certain reactions is occasionally mentioned, but it serves to draw conclusions about the qualitative nature of the process. For example, Shah et al. (2004) measure the turnover of various checkpoint proteins at the kinetochores. The quantitative information arising from these measurements, however, is subsequently used to draw a qualitative distinction between those proteins that are stably bound to the kinetochores and those that participate in the diffusible signal.

The final point in the previous section was the observation that when describing the mechanism of protein synthesis, biologists could restrict themselves to the description of events occurring at the level of individual molecules. In the present example, the same observation can be made. The overall process can again be explained by describing it in terms of individual molecules. For instance, the total checkpoint signal is simply taken to be the aggregate of the signals produced at the individual kinetochores. Similarly, the signal produced by one individual kinetochore is taken to be the aggregate of the activity of the individual molecules. The spindle assembly checkpoint mechanism, as it has been

described in this section, once more reflects the assumption that there are no non-trivial population effects involved in the processes of molecular biology.

2.5 Conclusion

I started this chapter by describing some very basic heuristics of mechanistic science, drawing on work by Bechtel and Richardson (1993) and Darden (2006). Next, I discussed two case studies of discovery in molecular biology that exemplify these basic heuristics, but also reveal additional and more specific strategies that belong to the perspective of traditional molecular biology.

Decomposition and localization, as described by Bechtel and Richardson, are fundamental strategies that scientists adopt to understand the behavior of complex systems. The goal is to describe the behavior in terms of sub-operations that are produced by specific components of the system (that might be complex systems themselves). Since the complexity of an epistemic task is related to the number of degrees of freedom of the system under study, as described in Chapter 1, we can understand how the strategies of decomposition and localization reduce this complexity: proper subsystems, of necessity, have less parts than the system in which they are embedded. The strategies thus seem to be well-suited for the discovery of mechanisms with a relatively small number of parts that can be described without taking into account the whole complexity of the systemic context. To be sure, this does not imply that the scientists assume that these mechanisms work in isolation: we have seen, for example, that the spindle assembly checkpoint mechanism is conceived as tightly integrated into the general machinery of the cell cycle. However, in this case, just as in the case of the mechanism of protein synthesis, it is possible to represent the mechanism as receiving an *input* from the environment and as generating an *output* that can in turn serve as an input for another part of the system. For instance, the dynamics of microtubule formation and movement are not included into the description of the spindle assembly checkpoint, but their attachment serves as an input to trigger a particular response in the mechanism. Similarly, the output of the checkpoint mechanism is the inhibition (or the release) of the activity of the APC/C, yet the downstream effects of this complex do not have to be included in the description. In general, it seems

that molecular biologists think of the mechanisms they study as quasi-independent modules, communicating via input-output relationships, and, therefore, as subsystems of a nearly decomposable system in the terminology introduced by Herbert Simon (1962) as discussed in Section 2.2.2. Decomposition and localization might be called 'reductionist' strategies in that they aim at the explanation of a behavior in terms of the parts of a system, while weighting less the interactions of the system with entities at the same level of description.

The strategies discussed by Darden (2006) can be understood as more specific heuristics that are applied within the more general framework of decomposition and localization. Especially the strategies of forward/backward chaining point to the prevalence of *sequential* organization in the mechanistic accounts of molecular biology. Apart from *direct localization*, which locates an activity in an individual part of the system, sequential organization is probably the simplest form of organization. The epistemic complexity is further reduced if one thinks of a process as a linear chain of events because one can zoom in once more and focus on the individual links of the chain. In a recent article, Bechtel has discussed the assumption of sequential organization in biological mechanisms:

> The assumption of sequential order reflects the practices of many scientists, who attempt to envisage sequentially the qualitative changes occurring in the mechanisms they investigate. More fundamentally, this reflects the sequential nature of human mental processes. We perceive successive states of the world, and in imagination we redeploy perceptual processes ... and so imagine changes sequentially. (Bechtel 2011, 536)

Bechtel highlights the features of human reasoning that incline scientists to search for particular schemes of organization. He also mentions that sequential organization allows them to describe mechanisms qualitatively. While this is partly true, I think that additional assumptions must be taken into account in order to understand why molecular biologists restrict themselves to qualitative mechanistic accounts. After all, many processes in the realm of physics or engineering are studied with quantitative methods despite being sequentially organized (think for example of the investigation of a sequence of electronic elements such as resistors and coils). My discussion of the two case studies suggests that there are two further assumptions that underlie to the idea that qualitative

descriptions suffice for the purpose of explaining biological mechanisms. Both of these rely, even though differently, on the idea of mechanisms in molecular biology as processes of information transfer.

The first assumption, for which I have employed Monod's notion of 'gratuity,' is that the transmission of information in biological systems is basically unrestricted by chemical principles. Molecular biologists invoke biochemistry for the explanation of individual signaling reactions, phosphorylation, inhibition, etc., but not for the dynamics of the overall processes. The idea of gratuity relies on the assumption that evolutionary processes, even though having to work with chemical 'bricks,' had the freedom to 'engineer' physiological systems in a largely unconstrained way:

> [T]he very gratuitousness of these systems, giving molecular evolution a practically limitless field for exploration and experiment, enabled it to elaborate the huge network of cybernetic interconnections which makes each organism an autonomous functional unit, whose performances appear to transcend the laws of chemistry if not to ignore them altogether. (Monod 1971, 78)

The idea of gratuity is what actually enables biologists to investigate the individual links in a sequence independently from one another: there is no dependency of the single steps in the process on the overall organization of the system. Compare this to the example of two resistors arranged in a series circuit. According to Ohm's law, the voltage drop across one of the resistors depends on the resistance of the other resistor. Gratuity in this context would mean that we could understand what the resistor does by studying it in isolation. Note, that this particular 'cybernetic' vision of living systems is conceptually independent from the idea that all biological processes are controlled by genes. The spindle assembly checkpoint mechanism provides an example of control in which genes are not directly involved. Molecular biologists, even though they work within an informational perspective, are not necessarily genetic reductionists.

The second assumption that I have highlighted in both examples is the fact that population effects are largely disregarded in the accounts of molecular biologists. In the typical descriptions of a mechanism, molecular biologists content themselves with describing what happens to individual molecules, even though they are aware that the actual molecular players are in most cases populations of molecules. This habit relies on the

tacit assumption that there is a simple relationship between the activity of the individual molecule and the activity of the population. For instance, if a molecule of *A* inhibits the activity of molecule *B* by binding to it, the expectation is that the activity of a population of *A*s simply inhibits the activity of a population of *B*s. In particular, it is assumed that the activity at the population level can be described qualitatively if the interaction at the molecule is discrete. However, the effect of a population on another can in general not simply be equated with the effect of an individual member, which can already be seen by looking at a simple example from ecology, such as the Lotka-Volterra model. This model describes the dynamics of two interacting species, one a predator and one its prey. While at the level of individual members 'predation' implies one organism hunting and killing another, the interactions at the population level can be more complex. The prey population is not simply killed but depleted at a certain rate depending on the size of the predator population. Moreover, the complex behavior showed by the model, such as the occurrence of oscillations, can only be explained when explicitly describing the process at the population level.

Taken together, the set of heuristics that I have outlined here, from the more general (decomposition and localization, assumption of sequential organization) to the more specific (gratuity, neglect of—or disregard for—possible population effects), implies a particular perspective on the organization and complexity of living systems. They facilitate the process of discovery and at the same time suggest that we can actually achieve an understanding of biological phenomena. Epistemic complexity is reduced to a large extent by restricting the set of expected causal structures; the main focus is, therefore, on what I have described in Chapter 1 as the problem of search.

The heuristics discussed in this chapter do not exhaust the repertoire of epistemic strategies employed by molecular biologists, and they should certainly not be taken as literally representing the beliefs of individual scientists. What I have tried to find are some general features, shared by most work in traditional molecular biology that imply a certain idea about the intrinsic complexity of living systems. The search for mechanisms with simple organization, that can be described in qualitative terms, rests on this idea, and does not simply express an aversion of molecular biologists to the use of mathematical tools.

As I have mentioned, the label of 'methodological reductionism' might be used in a meaningful way for describing the strategies of decomposition and localization. It is not obvious, however, whether it makes any sense to call the more specific heuristics that I have discussed reductionist, unless reductionism is taken in such a broad sense that it applies to basically any approach of making complex phenomena intelligible (that is, to most of science). Instead of forcing molecular biology into the pigeonhole of reductionism, it seems more promising to address the particular combination of heuristic strategies that make up its 'epistemic toolkit.' In this way the question of what (if anything) is new or different in systems biology can be addressed in a more meaningful way.

3

STRATEGIES OF SYSTEMS BIOLOGY

Summary

In this chapter I discuss a number of case studies from recent work in systems biology. My focus is on the role of mathematical tools in biological discovery, that is, in the process of developing and revising mechanistic models. I will show how systems biology replaces some of the heuristic strategies of molecular biology. The examples are ordered roughly according to increasing size of the models. At first I discuss models of 'small' mechanisms which continues the discussion of the spindle assembly checkpoint mechanism started in Chapter 2. These models allow systems biologists to relax the more specific assumptions of the traditional approach. Afterwards, I deal with the study of large networks that raise more fundamental issues about modularity and decomposability. Finally, I discuss a recent example of whole-cell modeling which proposes a particular solution to the problem of integrating different models.

3.1 Introduction

After having characterized the approach of traditional molecular biology, I will now turn to the question of whether systems biology provides additional, or alternative strategies to the general project of biological discovery. As I mentioned before, systems biology is

not one homogeneous endeavor, but rather a large collection of different approaches that have their historical roots in various traditions of theoretical biology or other theoretical fields studying complex systems. However, a key feature of all approaches in systems biology is the use of mathematical tools. It is often claimed that this renewed interest in mathematical methods is due to the accumulation of quantitative data by means of modern experimentation:

> With the availability of quantitative data on the transcriptome and proteome level, there is an increasing interest in formal mathematical models of gene expression and regulation. (Wolkenhauer 2001, 258)

However, being quantitative for a science is not an aim in itself, and I have argued in Chapter 2 that molecular biology could for a long time do without quantitative methods because of a particular view on the organization of living systems—not necessarily because of a lack of quantitative data.

In a recent article, Rasmus Winther has identified four main functions of mathematical modeling in biology (Winther 2012): *unification* of both models and data, *model fitting* to data, *mechanism identification*, and *prediction*. Even though we will find instances of all of these functions in the discussed examples, I argue that *mechanism identification* is central to systems biology in its current form. In other words, what I want to show in this chapter is that one of the main roles of mathematical models in systems biology is to facilitate the discovery of mechanisms. Modeling is used as an additional tool to restrict the set of possible causal structures underlying a particular phenomenon. In spite of increasing amounts of data at the level of gene expression at the RNA and protein level, of concentrations of metabolites, of epigenetic modifications etc., most areas in molecular biology are lacking knowledge about the underlying causal structures. At the same time many of the measurements, despite being quantitative, often lack in both precision and accuracy. This explains why instead of trying to understand the behavior of systems whose structure is largely known, a big part of systems biology consists in figuring out this causal structure in the first place. The following quote is taken from an article about modeling of complex signaling networks:

> We believe that modeling these important biological systems cannot wait until all the rates are reliably measured, or even until all the various players and

> interactions are discovered. Indeed, the most important role of modeling is to identify missing pieces of the puzzle. It is as useful to falsify models—identifying which features of the observed behavior cannot be explained by the experimentalists' current interaction network—as it is to successfully reproduce known results. (Brown et al. 2004, 185)

The examples I discuss in the following are supposed to substantiate this idea of mathematical models as having a productive role in the discovery of mechanisms. I discuss a number of examples in order to do justice to the heterogeneity of the field, but I am aware that I will have left out whole strands of research in systems biology, and I can therefore make no strong claim of generalizability.

3.2 Mathematical Models of Small Mechanisms

In this section, I want to discuss the practice of building relatively small mathematical models of individual mechanisms that are also studied by traditional molecular biology. Building simple mathematical models to understand and explain particular phenomena is not an invention of systems biology, but there is a longstanding tradition of modeling in biology that goes back at least to the models of predator-prey interactions that were studied, for instance, by Lotka and Volterra in the first decades of the twentieth century. Another famous historical example that has received a lot of attention from philosophers of science (e.g. Weber 2005, Craver 2007) is Hodgkin and Huxley's model of neural rhythms (Hodgkin and Huxley 1952). Looking at these instances of success, one might get the impression that modeling is mainly used in order to understand or to explain phenomena that are somehow unexpected or puzzling, such as chaotic or oscillatory behaviors in a system. I want to suggest, however, that an important, and perhaps the most important role, of mathematical models of mechanisms in systems biology lies in their potential to facilitate biological discovery. I discuss two basic strategies, 'thin' and 'thick modeling,' that can be used at different stages of the process.

3.2.1 Example: Modeling the Spindle Assembly Checkpoint

In Chapter 2 I discussed the discovery of the spindle assembly checkpoint mechanism. I showed how this phenomenon was explained in qualitative terms, mainly by elucidating the structure of large macromolecular complexes and simple biochemical reactions that instantiated an abstract 'checkpoint mechanism schema.' From my review of the experimental literature it appeared that the basic scheme had been worked out and that the remaining work to be done would consist in the further specification of molecular details. However, throughout the last decade there has also been a significant number of articles approaching the spindle assembly checkpoint from a 'systems perspective' (Doncic et al. 2005, Mogilner et al. 2006, Sear and Howard 2006, Ibrahim et al. 2008, Mistry et al. 2008, Simonetta et al. 2009, Doncic et al. 2009, He et al. 2011, Dao Duc and Holcman 2012). The aim of this section is to take a closer look at two examples in order to illuminate what exactly a 'systems approach' might consist in, what its aims are, and what differences we can find with respect to the 'traditional approach' of molecular biology discussed in the previous chapter.

To begin the discussion, I will quote from a recent review that summarizes different modeling efforts and attempts to synthesize a coherent picture of a "quantitative systems view of the spindle assembly checkpoint" (Ciliberto and Shah 2009). In the beginning of the article, the authors explain the particular interest in the spindle assembly checkpoint as a target of computational modeling:

> The high fidelity and robustness of this process have made it a subject of intense study in both the experimental and computational realms. A significant number of checkpoint proteins have been identified but how they orchestrate the communication between local spindle attachment and global cytoplasmic signalling to delay segregation is not yet understood. Here, we propose a systems view of the spindle assembly checkpoint to focus attention on the key regulators of the dynamics of this pathway. These regulators in turn have been the subject of detailed cellular measurements and computational modelling to connect molecular function to the dynamics of spindle assembly checkpoint signalling. (Ciliberto and Shah 2009, 2162)

Thus, they maintain that the mechanism, in spite of the accumulated amount of molecular details, is not yet well understood. Differently from the reviews by molecular biologists that we have seen, however, they do not see the main problem in missing molecular data, but rather in a missing link between "molecular function" and "the dynamics of spindle assembly checkpoint signalling." What exactly does this mean?

In the introduction to their review, Ciliberto and Shah write:

> Given its role, it is not surprising, but yet striking, that the spindle assembly checkpoint can delay anaphase in response to a single uncaptured chromosome, exhibiting excellent sensitivity. Once this last chromosome attaches, the spindle assembly checkpoint disengages and rapidly promotes anaphase onset. High fidelity and speed are usually competing design constraints in manmade machines, and as such the underlying logic and quantitative mechanisms of the spindle assembly checkpoint are of interest to life scientists and physical scientists alike. (Ciliberto and Shah 2009, 2162)

This shows that the checkpoint mechanism is interesting for quantitative modeling because it solves a 'design problem' that would provide a challenge for human engineers. On the one hand, it has to work reliably because the fidelity of chromosome segregation is of crucial importance for the cell. It must, therefore, be sensitive to the signal produced by one single unattached kinetochore. On the other hand, it has been observed that anaphase onset occurs in a matter of minutes after the last chromosome attaches (e.g. Rieder et al. 1995, Howell et al. 2000). The observation of these competing constraints gives rise to the question of how the biological system solves this 'design problem,' which would represent a difficulty for human engineers. However, taking these constraints into account requires a quantitative and dynamic perspective on the system. To understand whether a proposed mechanism can produce reliable inhibition, even when coming from only one chromosome, one has to consider both the rate of the putative reaction that produces the inhibitory signal and the diffusion rate of the signal through the cytosol. Similarly, to understand whether the checkpoint can be relieved fast enough, one has to take into account the rate of disassembly of the inhibitory complex as well as the time it takes for the APC/C to carry out its activating function.

Ciliberto and Shah use the analogy of a washbasin to explain why these constraints

might be competing (Figure 3.1). The production of the inhibitor is represented by a faucet filling up the sink, while its dissociation corresponds to the outflow through the drain pipe. In the scenario represented in Figure 3.1 (**A**), the inhibitor is constantly flowing out, i.e. dissociated. This dissociation must be slow enough in order to allow for reliable inhibition while the checkpoint is active. In other words, a thin pipe is needed to guarantee that the the outflow does not exceed the inflow. As a result, it takes a long time to drain the sink: the silencing of the checkpoint is slow.

Figure 3.1 (**B**) proposes a possible solution to the design problem. This time, the dissociation rate is high, corresponding to a wide pipe, but the checkpoint produces an additional 'dissociation inhibitor' that plugs the pipe. As soon as the last kinetochore attaches to the spindle, both the faucet is closed and the plug is removed. In this way, the silencing of the checkpoint can be fast. Thus, with a slightly different checkpoint schema both constraints can be fulfilled. This analogy illustrates what a dynamic perspective on the

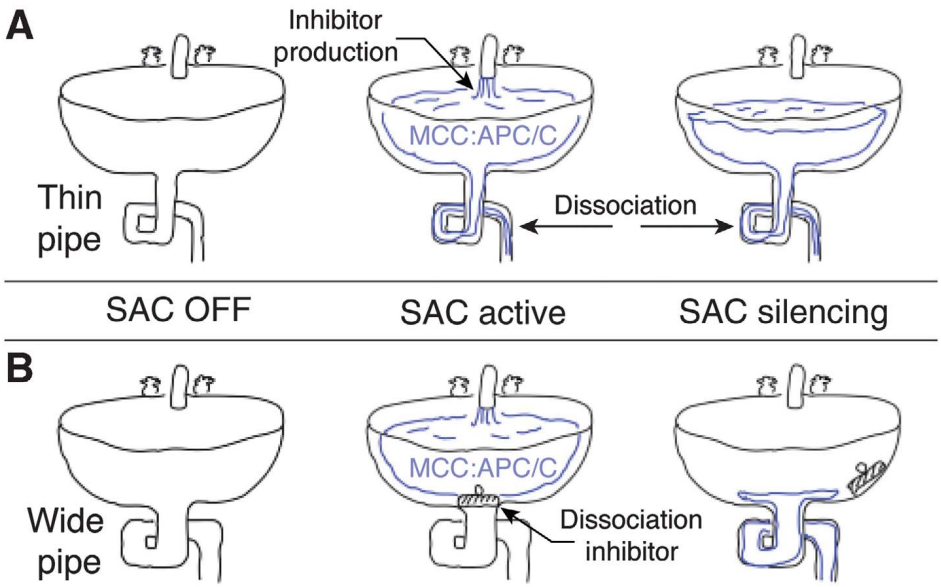


Figure 3.1: Two scenarios for the dynamic regulation of the spindle assembly checkpoint (SAC). For explanation see text. *Source:* Ciliberto and Shah (2009).

mechanism can contribute. The activities of the mechanism are not represented as if they were the actions of individual molecules, but in terms of changing quantities, which correspond to the concentrations or copy numbers of the different molecular species. At the same time, it becomes clear that such a dynamic vision must also take into account the 'systemic' nature of the mechanism. This is to say that the different steps in the process cannot be modeled independently since their dynamic features might depend on each

other. The scenario in Figure 3.1 (**A**) illustrates this kind of dependency. The slightly more complicated model in Figure in (**B**), by contrast, relieves the dependency by introducing an additional activity into the schema.

In order to deal with the dynamic nature of the mechanism, Ciliberto and Shah propose an approach that interprets its basic activities of terms of signaling *modules*:

> [T]hese activities, inhibition on the one hand and release of that inhibition on the other, must support the widespread observation of a single unattached kinetochore delaying the onset of anaphase. Moreover, the coupling of these activities and their relative dominance must be controlled entirely through kinetochore attachment to permit the rapid transition to anaphase on kinetochore attachment. Each of these activities: inhibitor generation, release from inhibition and kinetochore attachment are themselves complex signalling pathways involving a myriad of molecular components. A systems view of spindle assembly checkpoint signalling focuses our attention onto the communication between signalling modules that are likely to govern the quantitative dynamics of this pathway. (Ciliberto and Shah 2009, 2163)

As we have seen, the conceptualization of a mechanism in terms of functional modules is implicit also in the traditional approach of molecular biology. In Chapter 2 I discussed how molecular biologists decompose a mechanism into separate activities, which allows a reduction of epistemic complexity since each step in the process can be addressed independently. The molecular biology strategy requires, however, that the interaction between the modules is straightforward. Investigating each activity as an independent step in a sequential process, it ignores the ways in which the properties of different modules might depend on each other. The idea of the systems approach is to focus instead on the communication between the modules. Epistemic complexity is reduced as well, but this time by *black-boxing molecular detail within each module*. As the authors of the review explain, with their systems view they "modularize the complexity of the components into the key communicating elements" (Ciliberto and Shah 2009, 2162).

The motivation for using a coarse-grained perspective in terms of modules, however, must not necessarily lie in the belief that these modules represent the 'real' parts of the mechanism, or that the project of figuring out the molecular structure in detail is mis-

guided. Mainly, the strategy serves to make the task well-constrained as a modeling problem. Even though the research on the spindle assembly checkpoint "has amassed a substantial amount of quantitative data" (Ciliberto and Shah 2009, 2166), this does not automatically enable scientists to build useful quantitative models at the molecular level. The reason for this lies in what Jeremy Gunawardena has called "the parameter problem" (Gunawardena 2010).

The Parameter Problem

The essence of the parameter problem is captured by the famous expression, that has been attributed to John von Neumann: "with four parameters I can fit an elephant, and with five I can make him wiggle his trunk" (quoted in Dyson 2004, 297). In an ideal world every parameter of a model would be determined by independent measurement; in biological practice, however, most properties of interest cannot be directly measured, and even those measurements that can be obtained might have been made in conditions different from those that are relevant for the model. For instance, many measurements of biochemical properties are performed *in vitro*, but the corresponding *in vivo* values might differ substantially (e.g. Minton 2006). The result is that virtually every quantitative molecular model in biology must rely on a number of unknown parameters. One possible way out is to try to find the missing parameters by *fitting* them to experimental data. This essentially means that one simulates a proposed model with different combinations of parameters and picks out the set of parameters that best reproduces the observed behavior of the target system. Obviously, for large numbers of parameters one might have to try a lot of combinations, but nowadays there are computational algorithms that greatly facilitate this procedure.

However, the problem that von Neumann's statement raises is that models with a large number of free parameters are able to reproduce a wide variety of behaviors. This is maybe best understood as a generalization of the theorem that for every n datapoints one can find a polynomial of degree $n-1$ that goes through all the points. So the fact that we find a mathematical function that reproduces the data does not tell us anything about the world because it is a result of pure mathematics. If we are confident that the structure of our model matches the causal structure of the target system, then parameter

fitting is a way to obtain measurements that might otherwise be unavailable. However, as will become clear, mathematical modeling in systems biology is often used as a strategy to determine this causal structure in the first place. If a model has many free parameters, the fact that the model accounts for the data might largely be due to mathematical reasons, and not because the model matches the target system.

There are two strategies to cope with the parameter problem, corresponding to what Gunawardena calls "thick" and "thin" models (Gunawardena 2010, 26). Thick modeling is acceptable when enough empirical data of the right kind are available. Here one tries to bring the assumptions of the model as close to reality as possible and, therefore, accepts a large number of unknown parameters. Most commonly, one 'trains' the model with a part of the available data, and afterwards tries to reproduce or predict other data. Deviations between predicted and observed behavior can then be exploited to modify the structure of the model. Thin models, by contrast, include only what are assumed to be the essential causal features of the system. These models can be tested against a small set of observations and generic physical constraints.

I will now turn to examples of both strategies that will illuminate their roles in the discovery and explanation of biological mechanisms. More specifically, I will present two different ways of modeling the spindle assembly checkpoint: Doncic et al. (2005) is an example of thin modeling, whereas Doncic et al. (2009) may be classified as thick. The fact that both strategies have been followed by the same research group indicates that they are not mutually exclusive or competing.

A Biophysical Model: Doncic et al. (2005)

The first article by Doncic et al. presents a comparison of three different models of the spindle assembly checkpoint mechanism in budding yeast. These models are evaluated with respect to the two basic requirements that were above identified as competing design constraints. According to the first constraint, which goes back to the experiments by Rieder et al. (1995) mentioned in Chapter 2, one single unattached kinetochore must be able to maintain the inhibition of the APC-Cdc20 complex. Secondly, the checkpoint inhibition must be removed very quickly since cells have been observed to proceed to anaphase in a matter of a few minutes after the attachment of the last kinetochore. Con-

cluding from the analysis of their models, Doncic et al. confirm that these are in general competing constraints; in other words: "improving inhibition comes at the expense of activation time and vice versa" (Doncic et al. 2005, 6335). In the end the authors find that only one of the three proposed models passes the test of properly fulfilling the two requirements.

All three model variants are loosely based on molecular knowledge, but there is no strict identification of model components with specific proteins, and the main focus is on the role of physical constraints. The cell nucleus is modeled as a sphere with one single kinetochore located in the center as a subsphere with significantly smaller radius (see Figure 3.2). The molecular processes are characterized by a set of reaction-diffusion equa-

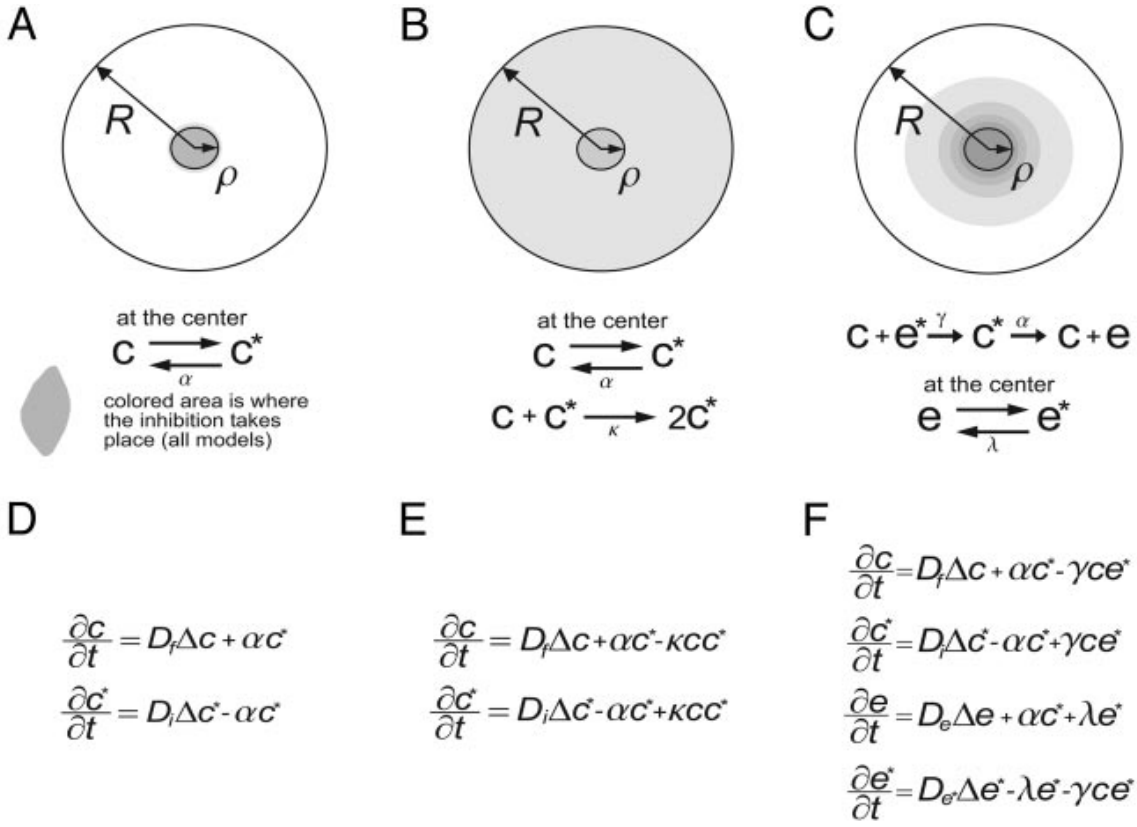


Figure 3.2: Three 'thin' models of the spindle assembly checkpoint. The panels in the upper row show the geometrical setup and basic reaction scheme of the 'direct inhibition model' (**A**), the 'self-propagating inhibition model' (**B**), and the 'emitted inhibition model' (**C**). Panels **D–F** show the corresponding sets of reaction-diffusion equations. *Source:* Doncic et al. (2005).

tions which describe both spatial and temporal changes of the molecular concentrations as well as the chemical interactions. Geometrical scale, reaction rates, and diffusion constants are chosen in agreement with known general properties of cellular systems. The at-

tachment process is simulated simply by setting all kinetochore dependent reaction rates to zero. In all three models it is assumed that the transition into anaphase is triggered by one component, called c, which diffuses through the nucleus. The checkpoint prevents this transition by inhibiting c through a mechanism that is specified differently in each model. Since the models lump the contributions of various molecular players into a small set of core activities, this is a clear case of thin modeling. As the authors explain:

> We did not simulate the full complexity of the network underlying the checkpoint but, rather, compared classes of mechanisms. Each class may be realized by a range of molecular machineries, but its essence can be summarized by a simple model, composed of just a few components. (Doncic et al. 2005, 6336)

In order to quantitatively evaluate the functional requirements, two read-outs are proposed. First, the inhibitory capacity of the checkpoint is measured by determining the steady-state fraction A_c of uninhibited c close to the nuclear boundary. Second, the rate at which the checkpoint is removed is quantified by T_b, the time it takes to get the fraction of uninhibited c above 90%. From observations in yeast, Doncic et al. infer that T_b should not be longer than 3 min., while the allowed level of inhibition is (arbitrarily) set to $A_c < 0.05$. These two constraints define the 'working range' of the mechanism, which means that a model is acceptable only if it satisfies both. In the following, I will give brief (qualitative) descriptions of the three proposed models. Further information is provided by Figure 3.2.

Direct Inhibition Model This model incorporates the simplest idea of checkpoint inhibition. It is assumed that the inhibitory activity of the checkpoint is locally restricted to the site of the kinetochore. If a c molecule hits the kinetochore, it is instantaneously transformed into the inhibited form c^*. Inhibited molecules spontaneously lose their inhibition at a constant rate α. Doncic et al. report that, for realistic parameters, the checkpoint performance is always outside the working range, which means that either the inhibition is not tight enough or that the removal of inhibition is too slow. Due to its simplicity, Doncic et al. are able to derive analytic solutions for the model behavior. In this way they find that T_b and A_c depend in opposite ways on α, the decay rate of inhibition. High val-

ues of α, corresponding to tight inhibition, always go along with very slow removal times. Conversely, if one chooses α small enough to fulfill the time constraint, the resulting inhibition is very poor. From the analysis of their first model the authors conclude: "Taken together, it appears that a system that relies solely on inhibition at the kinetochore itself cannot support good inhibition while maintaining rapid reactivation time" (Doncic et al. 2005, 6335).

Self-Propagating Inhibition Model The second model is based on the idea that the inhibitory signal is amplified through a positive feedback loop. The underlying biological interpretation is that the catalytic reaction at the kinetochore produces a molecular species that can itself catalyze an inhibiting reaction away from the kinetochore, (as had been suggested, for example, by De Antoni et al. 2005.) Mathematically speaking, this simply amounts to an extension of the first model with the additional feature that an inhibited c^* can bind to an uninhibited c at a rate κ and catalyze its inhibition anywhere in the nucleus. The authors find that this extended model easily gives rise to sufficient inhibition. However, due to the fact that the catalytic activity away from the kinetochores persists after the attachment, the system remains inhibited. More generally, the feedback loop becomes locked at high values of κ, producing a steady state that is independent of the attachment. In the limit of small values for κ, the model reduces to the direct inhibition model. The authors report that for intermediate values of κ and realistic parameters neither of the two checkpoint requirements are met. They conclude from this that the second model can be excluded as a candidate for the real mechanism as well.

Emitted Inhibition Model Finally, the authors consider a model with an additional molecular species e that represents an inhibitory complex. In order to carry out its inhibitory role, this complex must first be transformed into its activated form e^*. The activation reaction takes place at the kinetochore, and has the same form as the transformation $c \rightarrow c^*$ in the previous models. The activated complex then diffuses away from the kinetochore and transforms the c molecules into the inhibited form c^*. The authors report that the emitted inhibition model shows both sufficient inhibition and rapid reactivation times after the attachment. In addition, the model is consistent with the now widely accepted idea that the kinetochores catalyze the assembly of a diffusible 'mitotic checkpoint com-

plex' (MCC) that inhibits the APC/C away from the kinetochores.

A Reverse Engineering Approach: Doncic et al. (2009)

In this more recent work Doncic et al. adopted a fundamentally different strategy to which they refer as a 'reverse engineering' approach. Instead of focusing on the essential properties of the mechanism and analyzing the role of physical constraints, they attempted to capture the spindle assembly checkpoint network in full detail. They began by formulating a very general model that includes basically all of the molecular players known to be involved, but leaving open which are the relevant interactions between them. Specific parameter choices could then be used to generate different network topologies from this general scheme. Each choice can be interpreted as a hypothesis about the actual mechanism. The key idea was to compare the predicted behavior from a particular instance of the model to the phenotypic behavior of real cells under a set of different conditions. The instance of the model that best matched the observations was shown to reflect some of the features of the real checkpoint network.

The general model incorporates a biological picture according to which the core proteins of the checkpoint interact at the kinetochores and promote diffusible factors that can inhibit Cdc20. The model is composed of five 'checkpoint factors' (Mad1, Mad2, Mad3, Bub1, Bub3), that correspond to the proteins that were identified in the first checkpoint related yeast screens (see Chapter 2), plus two 'outside factors,' Mps1 and Ipl1, which correspond to proteins that are assumed to promote the kinetochore association of the checkpoint factors. In the most general form of the model, each factor can potentially activate each of the checkpoint factors. This corresponds to a network with five nodes for the checkpoint factors, each having four possible edges representing the connections to the other factors plus one additional edge connecting to an outside factor. Each edge can be assigned a direction (determining who activates who), and a weight between 0 and 1 indicating the strength of the interaction (Figure 3.3). Checkpoint function is controlled via the inhibition of the APC:Cdc20 complex, either by sequestration or by degradation of Cdc20. However, only a subset of the checkpoint factors, namely Mad2, Mad3, and Bub3, and the possible complexes formed between them, can contribute to inhibitory activities. Each of the inhibitory elements can contribute to sequestration and degradation accord-

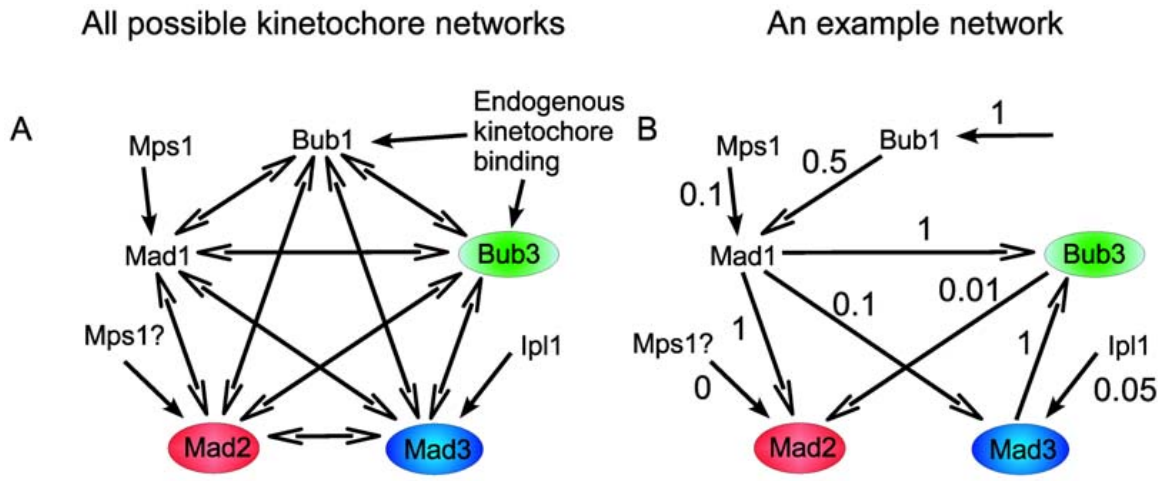


Figure 3.3: The network underlying Doncic et al.'s reverse engineering approach. **A** shows the general network with all possible interactions, while **B** corresponds to one particular instantiation. *Source:* Doncic et al. (2009).

ing to a given weight. The total inhibition rate is then given by summing the contributions from all the elements.

The output of the model, the level of active APC:Cdc20 during an active checkpoint, must somehow be connected to a measurable phenotype. One feature that can be used to assess correct functioning of the spindle assembly checkpoint, and that can be easily quantified in yeast cells, is the chromosome missegregation rate (CMR). If the checkpoint does not function properly, chromosomes will be separated even in the presence of imperfect attachments. This leads to cells with either too few or too many chromosomes. In order to make the connection with their model, Doncic et al. assumed that the CMR is proportional to the level of active APC:Cdc20. The underlying idea is that the more of it there is, the more likely it is that anaphase will be initiated prematurely, leading to errors in chromosome segregation. This move allowed them to predict the impact of deletions of network components on the CMR for any network topology. To identify the topology that comes closest to the actual checkpoint, they searched for the best overall fit with a set of mutant yeast strains in which key checkpoint proteins were deleted.

The actual reverse engineering part of the procedure consisted of an experimental and a computational step. First, in a series of experiments, the CMR of wild type yeast cells and of mutants deleted of either of the genes Bub1, Bub3, Mad1, Mad2, Mad3 were determined. The second step consisted of a computational screen through thirty million possible networks, comparing the output of each to the measured CMRs of the mutants.

The networks differed in their edge configurations and corresponding activation values, as well as in the types of sequestering and degrading complexes. The screen identified a total of 105 networks that were consistent with the measurements of the five phenotypes within 5% of the experimental values. A cluster analysis enabled them to reduce this number to 82, by removing all redundant solutions. In order to further restrict the set of possible solutions, they compared the remaining models to the CMRs of two mutants carrying the double deletions *mad1mad3* and *mad2mad3*. Experimentally, they observed a strong buffering effect in the double mutants, meaning that the CMR was far lower than the product of the rates in the mutants carrying the individual deletions. Only two of the remaining networks accurately predicted this buffering effect. Of these, the authors in the end excluded one by arguing that it "appeared less plausible since it relied on highly improbable interactions and complexes" (Doncic et al. 2009, 7). The final result of this reverse engineering procedure, therefore, was one specific model that describes the structure and function of the (cell cycle arresting part of the) spindle assembly checkpoint network.

Concluding their article, the authors emphasize that their model reproduces established and previously hypothesized interactions of the real network even though they "made no assumptions about their existence" (Doncic et al. 2009, 9). It seems, therefore, that the strategy is successful in predicting structural features from behavioral data.

3.2.2 Discussion: Thin and Thick Modeling

In both of the examples that I just presented, modeling is used as a strategy for discovery. Models are not built based on known causal structures in order to investigate their properties. Instead, modeling is used to decide between possible structures. In both cases the starting point is a set of possible mechanisms that are consistent with known molecular knowledge and, at least at first sight, plausible explanations of observed behavior.

Doncic et al. (2005) provide three different models, each of which represents a whole 'class of mechanisms.' By abstracting from most of the molecular details, they are able to cover a large set of possible molecular structures with a small number of models. Each model is simple enough to derive fairly general claims about its behavior under varying parameter values. The goal of this work is not to *explain* the spindle assembly checkpoint,

even though, as a by-product, it might provide understanding by elucidating how certain causal structures bring about certain behaviors. Instead, its main interest is to compare different models with respect to their ability to fulfill specific design constraints. As we have seen, only one of the proposed models meets these constraints. Nevertheless, it is not proclaimed as the actual mechanism of the checkpoint. The more important result is negative: Certain types of causal structures are *not* able to account for the observed behavior, and every proposed molecular mechanism must be evaluated with respect to these constraints. The authors cannot guarantee that their selection of models exhausts all possible checkpoint mechanisms, and whether the actual checkpoint mechanism follows the successful 'emitted inhibition model' can be established only on the basis of additional molecular knowledge.

The authors can exclude some proposed mechanisms because they take into account constraints that do not appear in the mechanistic models of molecular biologists. Firstly, they quantify the observed behavior: it is not enough that the checkpoint is released after attachment, but it must be released *within a certain time*. Information about upper limits of the rates of chemical reactions are important to evaluate whether a proposed mechanism can fulfill this time constraint. Similarly, it is not sufficient to show that one type of molecule is able to inhibit another, but inhibition must be strong enough in terms of the fraction of inhibited molecules. The interaction between an inhibiting and an inhibited species is a chemical reaction that produces a dynamical equilibrium in which there always remain a number of uninhibited molecules. The chemical perspective, therefore, implies a reasoning in terms of *populations* of molecules, which we found to be largely missing in traditional molecular biology. Inhibition has different meanings depending on whether one talks about populations or about individual molecules. The activity of a single molecule is inhibited if it is bound to its inhibitor, whereas the activity of the population is inhibited if the number of uninhibited molecules is below a certain threshold. A further consequence of quantitatively accounting for the chemical reactions is that the strength of inhibition gets connected to the timing for the release from inhibition. This is exactly what was illustrated in Ciliberto and Shah's washbasin model (Figure 3.1). Aside from that, the models of Doncic et al. (2005) take into account the spatial properties of the system: inhibition must be strong everywhere in the nucleus, and not only near the

kinetochore. Limits of possible diffusion rates of proteins, therefore, set further important constraints on the possible signaling mechanism.

It has to be noted, though, that this strategy of thin modeling involves trade-offs. It seems that in order to serve as powerful heuristic tools, the proposed models must be of low complexity.[1] In Doncic et al.'s work we can find instances of both minimal and Galilean idealization (see Chapter 1). First of all, Doncic et al. lump a whole network of interactions into a minimal number of effective reactions. This is not an assumption they want to test, but it is a requirement of their strategy. Moreover, they make simplifying assumptions, such as the idealized spherical geometry of the system or the conservation of the numbers of all interacting particles. Especially this latter assumption is problematic since it has been shown that some of the components are actively degraded during and after a checkpoint arrest. In general, their models can capture only mechanisms that approximately fulfill these underlying assumptions. To the extent that these are unrealistic, the overall strategy cannot amount to a strict criterion to exclude candidate mechanisms.

Let us now turn to the reverse-engineering approach pursued in Doncic et al. (2009). Since it explicitly incorporates all relevant molecular species, it can be understood as an instance of 'thick' modeling. Its aim appears more ambitious than that of the previous project:

> Our previous work focused on the essential properties of the SAC, but did not attempt to capture the full details of the network. Here, we attempted to proceed beyond this general description and examine the possibility of deducing the detailed interactions between the checkpoint proteins using the quantitative phenotype of gene deletion mutants. (Doncic et al. 2009, 2)

For this reason, the proposed model is much more complex than the models in Doncic et al. (2005). In its general form it can describe all the possible interactions between different components, but it initially leaves open which of them actually do occur and with what strength. The point of this particular work is to show that reverse-engineering is able to detect a network topology that matches the findings of molecular biology. It is, therefore, best understood as a proof-of-principle for reverse-engineering as a general

[1] What I mean by the complexity of models is essentially the number of free parameters. For a more general discussion of why simpler models might be better tools, see Hitchcock and Sober (2004).

approach to biological discovery.

As I discussed in Chapter 1, the starting point of discovery is usually a vast number of possible causal structures. Traditional molecular biology aims at establishing the actual structure ultimately by directly checking the putative interactions between molecular components. As we have seen in the last chapter, this can be extremely tedious and requires sophisticated experimentation. Even though the discussed heuristic tools make the search much better than random, many findings in molecular biology appear to be largely the result of serendipity. The reverse-engineering approach, by contrast, promises an unbiased and systematic approach to discovery. It simply screens through all possible structures and selects those that are consistent with the observed systemic behavior. It relies only on information that is relatively easily accessible, and it seems to require much less tinkering with the experimental system. The success of Doncic et al.'s strategy raises the question of why one should go through the tedious experimental work at all. The reverse-engineering approach, however, is not as straightforward as it may seem, and I want to briefly highlight some crucial assumptions and possible drawbacks.

Obviously, screening through a large set of different parameter settings requires a lot of computing power. To check 30 million different sets, as done in this example, might not pose a serious problem. However, moving to more complex models and larger numbers of free parameters might quickly make computing power the limiting factor. Moreover, due to what I have referred to as 'the parameter problem,' one needs a suitable set of quantitative data to sufficiently constrain the search. Doncic et al. had to create seven different yeast single mutants and two double mutants in order to perform their screen. They observe:

> The number of free parameters over which we screened was rather large, and we compared them to only seven quantitative phenotypes that were derived to a limited resolution. In addition, some parameters not screened over were fixed by literature values, which are again, known only to some limit. It is interesting that despite these inevitable limitations, the reverse engineering theme was quite successful in pinpointing the key features of the checkpoint. This [makes] us optimistic regarding further developments in this direction. (Doncic et al. 2009, 2)

Note, however that a sufficient number of viable mutant strains (or other suitable perturbation datasets) cannot always be generated easily. Aside from that, it should be mentioned that Doncic et al.'s screening was efficient in part because of various simplifying assumptions in the model. First of all, they took into account only steady state solutions and did not capture any of the dynamic interactions involved in the initiation or relieve of the checkpoint. Second, they assumed that the activation levels of the molecular components were linearly dependent on each other and that there were no feedback loops in the network. A third crucial simplification was to assume the proportionality of APC level and chromosome missegregation rate. Moreover, as can be seen in the supplementary material to their article, Doncic et al. did not treat all possible networks in the same way. They distinguished between 'known' and 'putative' interactions, giving the latter less weight in the screening. Certain topologies were not allowed from the start on the grounds of empirical knowledge, and 'redundant' solutions, that is, solutions producing the same activation levels of the output variables or containing 'insignificant edges,' were excluded as well. Thus the procedure eventually yielded a unique solution not due to its 'algorithmic' nature, but because existing molecular knowledge was used to exclude many of the possible network topologies. I mention all this not to discredit the reverse-engineering approach, but rather to give an impression of how much simplification and biological input is needed to perform a sufficiently constrained parameter screen. The fact that simplifications are needed reveals the strategy as clearly heuristic in nature. These simplifications imply unwarranted assumptions, which makes it necessary to check specific results by other, more reliable means.

3.2.3 Conclusion

In this section I have discussed two examples of modeling the mechanism underlying the spindle assembly checkpoint. This mechanism is known to consist of a relatively small network of components, and a considerable amount of molecular knowledge has been accumulated regarding their interplay. As we have seen in Chapter 2, molecular biologists have gained a basic understanding of how the mechanism works, but they have not yet filled in all the molecular details.

Systems biologists directly build on the findings of molecular biology, and they are in-

terested in the solution of the same epistemic puzzle of how the mechanism works. Yet, they propose a different strategy for solving it. As I have shown, the primary role of systems approaches in this context is not simply to explain complex behavior on the basis of known molecular interactions, but to use mathematical modeling as an additional tool for discovery. Describing the phenomenon and the hypothesized causal structure quantitatively allows systems biologists to detect discrepancies between proposed mechanisms and reality. Moreover, the introduction of physical and biochemical constraints can lead to the exclusion of mechanistic models, even if they are considered plausible candidates by traditional molecular biologists.

More specifically, I have discussed two strategies, 'thin' and 'thick' modeling, even though the difference between them might be considered a matter of degree. Both strategies can be understood as strategies to exclude candidate mechanisms. Thin models are restricted to capturing the essential activities underlying a phenomenon and thereby stand in for large classes of possible mechanisms. Such models are evaluated with respect to very basic constraints, due e.g. to time, space, limits of possible reaction or diffusion rates, etc. In this way it is possible to uncover inconsistencies in the proposals of molecular biologists and to suggest candidate mechanisms that are consistent with these constraints. It has to be noted, however, that in order to be powerful as heuristics, thin models rely on, at times crude, idealizations. For this reason, molecular biologists might, sometimes rightly, reject such models as 'distortions' and consider the claims made by systems biologists as irrelevant to the plausibility of their proposals. However, a valuable role of thin models might be to raise the standards that proposed models must fulfill.

Thick modeling seems more likely to escape objections since it aims at describing mechanisms at the same level of detail as molecular biologists do. Due to the resulting increase in the number of free or underdetermined parameters, however, this strategy requires more quantitative data. With the help of modern computing power it seems feasible to screen high-dimensional parameter spaces and to find unique solutions, provided that the problem is sufficiently constrained. In the discussed example, however, we have seen that a solution could eventually be found only with the help of many simplifying assumptions and background knowledge.

Earlier in this section I quoted Ciliberto and Shah (2009) who promoted a modular ap-

proach to modeling the spindle assembly checkpoint. What distinguishes this approach from the traditional approach of molecular biology is not the fact that a system is decomposed into modules, but rather that the focus is on the organization of, or communication between, modules. As the discussion of my examples shows, this strategy involves a trade-off between the amount of molecular detail that can be incorporated and the level of organization that can be described. The thin models discussed in Doncic et al. (2005) comprise both the inhibition and the release parts of the mechanism, whereas the thick model in Doncic et al. (2009) only attempts to describe the system during an arrest, thereby focusing on only one of the modules of checkpoint activity.

In Chapter 2 I presented the general strategy of molecular biology as a hierarchal set of heuristics that together imply a particular view on the complexity and organization of living systems. The kind of modeling presented in this chapter can do without some of the more specific heuristics. However, as we have seen, it introduces other assumptions and effectively replaces some of the more specific strategies in the hierarchy with alternative heuristics. It is important to keep in mind this heuristic character of modeling and not simply to replace a 'molecular vision' with a 'modular vision of life' (cf. Hartwell et al. 1999).

3.3 Studying Large Networks

The examples of modeling that I discussed in the previous section are committed to the same basic decompositional strategy as the traditional accounts of molecular biology. They start with the goal of explaining a particular well-defined behavior in terms of a relatively small number of parts. Thus they rely on the same general perspective of functional modularity, even though they introduce different strategies to tackle complexity *within* the modules they are studying.

However, the system-wide study of the architecture of living organisms, as revealed in the various '-omics' projects, suggests that underlying many of the behaviors of biological systems are large networks of interacting components. It is not obvious whether the strategies to deal with small mechanisms can simply be scaled up:

> [M]ost biological characteristics arise from complex interactions between the

> cell's numerous constituents, such as proteins, DNA, RNA and small molecules. Therefore, a key challenge for biology in the twenty-first century is to understand the structure and the dynamics of the complex intercellular web of interactions that contribute to the structure and function of a living cell. (Barabási and Oltvai 2004, 101)

Albert-László Barabási, one of the authors of the article from which this quote was taken, is best known for his research in network theory. Along with others, he has proposed that the tools of this theory should be applied to biological systems as well.

Network theory was developed in the 1930s, largely within the social sciences. It became more widespread in its applications when connections with mathematics, especially graph theory, were established in the 1950s. The basic idea is to represent a system, in a very abstract way, as a series of nodes that are connected by links standing for pairwise interactions or relationships. One of the aims of the theory is to find quantitative measures of network properties in order to classify different types of networks. Network theory gained considerable popularity across the scientific community after it was shown around the turn of the millenium that networks as different as the world wide web, electrical power grids, and metabolic networks share some unexpected features, such as the property of being *scale-free* (Jeong et al. 2000). If a network is completely random, most nodes have roughly the same number of links, or *degree*.[2] In scale-free networks, by contrast, the degree distribution follows a power law, that is, the probability that a given node has k links is $P(k) \sim K^{-\gamma}$ for some positive exponent γ. This means that most nodes have only very few links, while there are a few nodes, called 'hubs,' that are highly connected. The fact that many different types of networks share this non-random property led many scientists, and notably biologists, to expect that the general concepts of network theory had the potential to reveal deep underlying principles and might increase our understanding of large complex systems (Keller 2005).

The study of universal properties of networks, however, did not turn out to be as fruitful as expected in biology. Nevertheless, many systems biologists hope that more specific network approaches, that elaborate on the concepts from network theory, will lead to important progress in the study of complex biological systems:

[2] 'Random' in this context refers to the process of network construction, i.e. links between nodes are created randomly, with every possible link having the same probability of being chosen.

> By itself, the fact that a network has scale-free properties is of limited use to biologists. Power laws occur very widely in nature and can have many different mechanistic origins. If we wish to obtain testable biological insights, we must probe further into the substructure of the network. (Bray 2003, 1865)

In the following, I will discuss two different approaches of building on the basic insights of network theory. The topic of Section 3.3.1 will be the search for *network motifs* which represents a strategy to decompose large systems into smaller units even when there is no intuitive functional decomposition. The network motif approach appears promising, even though it has to deal with a number of specific objections that are mainly linked to assumptions about the evolution of biological networks. However, some systems biologists argue more radically that one loses important biological understanding by decomposing large networks. In Section 3.3.2 I will discuss an alternative proposal according to which it is best to conceive of these networks as large dynamical systems since they show simple and coherent behavior at the macrolevel.

3.3.1 Decomposing Networks: Network Motifs

Network Motifs: The Basic Idea

The study of network motifs can be interpreted as an attempt to go beyond the investigation of global features of large networks. The approach has been developed mainly by the group of Uri Alon at the Weizman Institute in Tel Aviv, focusing on the properties of transcription networks of relatively simple, unicellular organisms such as *E. coli* or yeast. In his "Introduction to systems biology," Alon describes the ambition of his work as follows:

> Our goal will be to define understandable patterns of connections that serve as building blocks of the network. Ideally, we would like to understand the dynamics of the entire network based on the dynamics of the individual building blocks. (Alon 2007, 27)

On this view, networks are not simply large assemblages of interconnected genes or proteins, but they are constituted of 'building blocks,' that is, substructures that are situated somewhere between the level of the single node and the level of the whole network. These building blocks reveal themselves through recurring patterns of connectivity, or 'motifs.'

The key idea behind the search for motifs in a network might be called a 'reverse engineering' strategy. However, differently from the example I discussed in the previous section, its goal is not to determine the structure of molecular interactions underlying a particular behavior. Instead, one starts with the structure of a network and tries to make inferences on function. The search for motifs ideally begins with a complete description of the network's topology, that is, a map containing all the nodes and edges, and then applies a criterion of statistical significance to identify recurring patterns within subgraphs of the network. In order to find such a criterion, the network under study is compared to a computer-generated ensemble of *randomized* networks. These randomized networks share general characteristics, such as number of nodes and edges, with the real network, but the connections between their nodes are made at random. Network motifs are those patterns of connections that are found much more often in the real network than in the randomized networks. Their overrepresentation suggests that they are biologically meaningful and might play specific roles in the network, an idea that Alon supports with an evolutionary argument:

> [E]*dges in network motifs must be constantly selected* in order to survive randomization forces. This suggests that if a network motif appears in a network much more often than in a randomized network, it must have been selected based on some advantage it gives to the organism. If the motif did not offer a selective advantage, it would be washed out and occur about as often as in randomized networks. (Alon 2007, 29), emphasis in original)

To illustrate the basic idea of the approach, I will discuss the example of the *autoregulation motif* in the transcription network of the bacterium *E. coli* (Rosenfeld et al. 2002); Alon 2007, Chapter 3). This bacterium is one of the best studied organisms at the level of gene regulation, and a substantial amount of information regarding the interactions of regulatory proteins and their associated binding sites, as well as the organization of regulatory features have been integrated in the database *RegulonDB* (Huerta et al. 1998).

Transcription networks essentially consist of genes and transcription factors. By binding to specific regions of DNA, transcription factors, which are gene products themselves, can regulate the rate of transcription of a set of target genes. The bacterial transcription network can be understood as a *sensory* network, that is, its overall task is to respond to

signals such as changes in nutrient concentrations and external stresses. The autoregulation motif is very simple; it consists of only one node with an edge that originates and ends in that very node. At the biological level this corresponds to a transcription factor that directly regulates the transcription of its own gene. Alon and co-workers observed that the autoregulation motif occurs 40 times in the bacterial transcription network, in contrast with an expected number of one in a corresponding randomized network. The large overrepresentation of this motif translates into a statistically significant difference of 32 standard deviations! This finding suggests that the autoregulation motif is biologically meaningful.

One of the points that Alon stresses repeatedly is the fact that only a small number of all possible network motifs can be found in biological networks. This can best be appreciated when considering motifs consisting of several nodes. For instance, there are 13 possible three-node directed subgraphs, of which the simple feedback and the feed-forward loop (FFL) are well-known examples (Figure 3.4). By applying the same strategy used to discover the autoregulation motif, Alon and co-workers were able to show that the feed-forward loop is the only significant motif among the 13 possible three-node patterns in the bacterial transcription network. Even more striking is the case of four nodes: there are 199 possible four-node patterns, but only two of them turn out to be significant motifs in the transcription network. Apparently, among the many conceivable patterns, biological networks make use of only a few, or, as Alon puts it, "these networks are much simpler than they could have been" (Alon 2007, 45). A more exhaustive analysis shows that overall only four families of network patterns appear to be significantly represented in sensory transcription networks. The same kind of simplicity is found in other types of biological networks, such as developmental, signal transduction, or neuronal networks, even though they differ as to which are the significant families of motifs.

Functional Analysis of Network Motifs

Naturally, the question arises why it is that particular motifs are highly overrepresented in biological networks, while others are suspiciously absent. Addressing this question, however, forces biologists to go beyond a purely topological perspective on networks. The biological systems described as networks are obviously not static structures, but systems

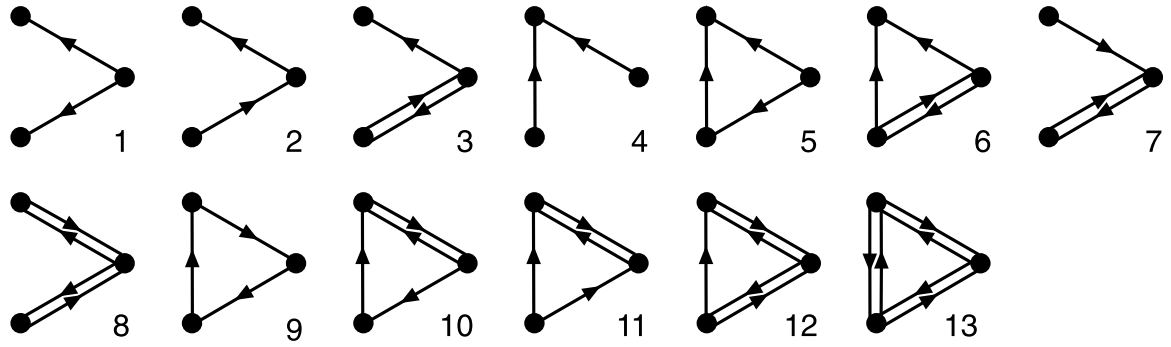


Figure 3.4: All types of three node connected subgraphs, including the feed-forward (5) and the feedback (9) motifs. *Source:* Milo et al. (2002)

whose components dynamically interact and thereby carry out complex functions, such as responding to environmental changes or controlling developmental processes. Therefore, in order to get at the putative biological significance of network motifs, one must focus on dynamic properties. By comparing the dynamic behavior of isolated motifs to dynamics produced by alternative possible structures, one might understand what functional benefit a particular motif provides for the whole network. This strategy was developed under the name of 'mathematically controlled comparison' in the 1970s and applied in the theoretical study of biochemical systems (Savageau 1976; see also Wall et al. 2004). Today, it can be applied on the basis of experimentally characterized systems, which has led to a complementary strand of the analysis of network motifs. To illustrate this type of analysis, I will again focus on the motif of autoregulation (Alon 2007, Chapter 3).

In our discussion we have so far neglected the fact that the edges in networks sometimes carry *signs*. These signs can become relevant when we turn to a functional analysis of network motifs. The directed edges in transcription networks, for example, can either represent *activation* or *repression*, based on the observation that a transcription factor can either increase or decrease the rate of transcription of a target gene. The large majority of the 40 autoregulatory proteins in *E. coli* have been found to repress their own transcription and are thus examples of *negative autoregulation*. In order to understand the putative functional significance of this motif, it is useful to compare it to the case of a *simply regulated* gene. Simple regulation means that a gene product Y is produced at a constant rate β_Y and degraded at a rate $\alpha_Y \cdot Y$ proportional to its own concentration. The

change in concentration over time can be expressed by the simple differential equation:

$$\frac{dY}{dt} = \beta_Y - \alpha_Y \cdot Y\,. \tag{3.1}$$

It can be shown that in the absence of external perturbations, a steady state is reached whose level is given by the ratio of the production and degradation rates:

$$Y_{\text{ss}} = \frac{\alpha_Y}{\beta_Y}\,. \tag{3.2}$$

A further quantity of interest is the *response time* of a variable in a dynamic process, defined as the time it takes to reach the halfway level between the initial and the final state. Regarding the regulation of a gene, one is usually interested in two cases: the process to reach the steady state from an initial concentration of zero, or the reverse case in which the concentration drops to zero after the production has been 'switched off.' It turns out that in the case of simple regulation, the response time for both processes is the same and given by:

$$T^Y_{1/2} = \frac{\log(2)}{\alpha_Y}\,. \tag{3.3}$$

Now consider the case of a negatively autoregulated protein X. Autoregulation means that the production rate is dependent on the level of X itself. Its dynamics can therefore be described by the following type of equation,

$$\frac{dX}{dt} = f(X) - \alpha_X \cdot X\,, \tag{3.4}$$

with some function f that describes this dependence. The simplest way to capture the idea of negative autoregulation is to use a *logical approximation* for the form of f. This means that if X is below a certain threshold K, its promoter is 'on' and X is produced at a constant rate, or $f(X < K) = \beta_X$. As soon as X reaches the threshold, the promoter is 'switched off' and the protein production instantaneously drops to zero, or $f(X \geq K) = 0$. It can be shown that a more realistic description of the production rate, for instance using Michaelis-Menthen kinetics for promoter activity, yields very similar results. A protein whose dynamics comes close to the logical approximation has a steady state that is equal

to the threshold value of its own promoter:

$$X_{ss} = K\,, \tag{3.5}$$

and the response time to reach half of the steady state level starting from zero is given by:

$$T^X_{1/2} = \frac{K}{2\beta_X}\,. \tag{3.6}$$

Comparing these expressions with the steady state and the response time of a simply regulated gene, one can draw several conclusions. One reason why the design of negative autoregulation might have been selected for is the robustness of its steady state level to fluctuations in other biological parameters. For a simply regulated gene, the steady state linearly depends on its production rate which can substantially fluctuate over time due to different stochastic effects, the availability of nucleotides and amino acids, the number of ribosomes, etc. The repression threshold K of a promoter, by contrast, is a 'hardwired' parameter since it depends only on the chemical structure of specific molecules and as a result shows much less fluctuation. Negative autoregulation has increased robustness because the steady state level depends only on K (Eq. 3.5).

Becskei and Serrano (2000) studied the dynamic properties of negative autoregulation experimentally by constructing a synthetic circuit in *E. coli* that could be directly compared to its simply regulated counterpart. The construct consisted of a fusion of the fluorescent protein GFP with TetR, a transcription factor that represses its own production. Simply regulated control circuits were produced either by mutating the DNA binding domain of the repressor, or by replacing the operator elements in the promoter. Their experiments revealed that autoregulation can dramatically reduce the variation in protein levels among different cells, thus confirming the robustness property of the motif.

A second possible advantage of autoregulatory design is its potential to speed up the kinetics of transcription. The response time in the simply regulated case is slow, especially for long lived genes since, according to Eq. (3.3), it is inversely proportional to the rate of degradation. Increasing degradation in order to achieve faster response times would require constant production and turnover of the protein and, therefore, impose substantial energy costs on the cell. For a negatively autoregulated gene, by contrast, the response

time is proportional to the promoter threshold K and inversely proportional to the production rate β_X (Eq. 3.6), and. Therefore, the response time can be speeded up, independently of the degradation rate, by using a strong promoter and a suitable threshold. The promoter guarantees a fast initial increase in concentration, but the repressor shuts off the production as soon as the desired steady-state level has been reached. Working with the same experimental system that Becskei and Serrano had used, Rosenfeld et al. (2002) showed that the response time of the negative autoregulatory circuit is indeed much shorter than that of the simply regulated one. Strikingly, the response time is reduced to about one fifth for the autoregulatory circuit, which matches the theoretically predicted (parameter-free) value of 0.21 for Michaelis-Menthen kinetics in the limit of strong autosuppression.

This brief discussion shows how the statistically identified network motifs can be analyzed with respect to their potential function in a biological context. Similar analyses have been carried out for other types of network motifs, and there is evidence that these have been selected due to their specific functional properties as well.

Recomposing the Network

The functional analysis of network motifs reveals aspects in the architecture of networks that appear beneficial from an engineering standpoint. It shows, for instance, that certain features convey robustness to perturbations, reduce detrimental variability, or are energy efficient solutions to a given task. However, with this result in mind, the question of how to explain the working of large networks is merely reformulated: How do we explain the behavior of the whole network in terms of the properties of the motifs? People advocating the network motif approach attempt to show that there is a tractable relationship between the behavior of individual motifs and the overall behavior of a network.

Alon (2007) argues that only four families of network motifs can be found in sensory transcriptional networks such as the one of *E. coli*. What is more, these families seem to cover virtually all of the genes in the network. Aside from the discussed example of the autoregulation motif, one finds feed-forward loops (FFLs), single-input modules (SIMs), and dense overlapping regulons (DORs). By aggregating the single nodes of the network into these four groups, one can achieve a relatively compact representation of the whole

network (Figure 3.5).

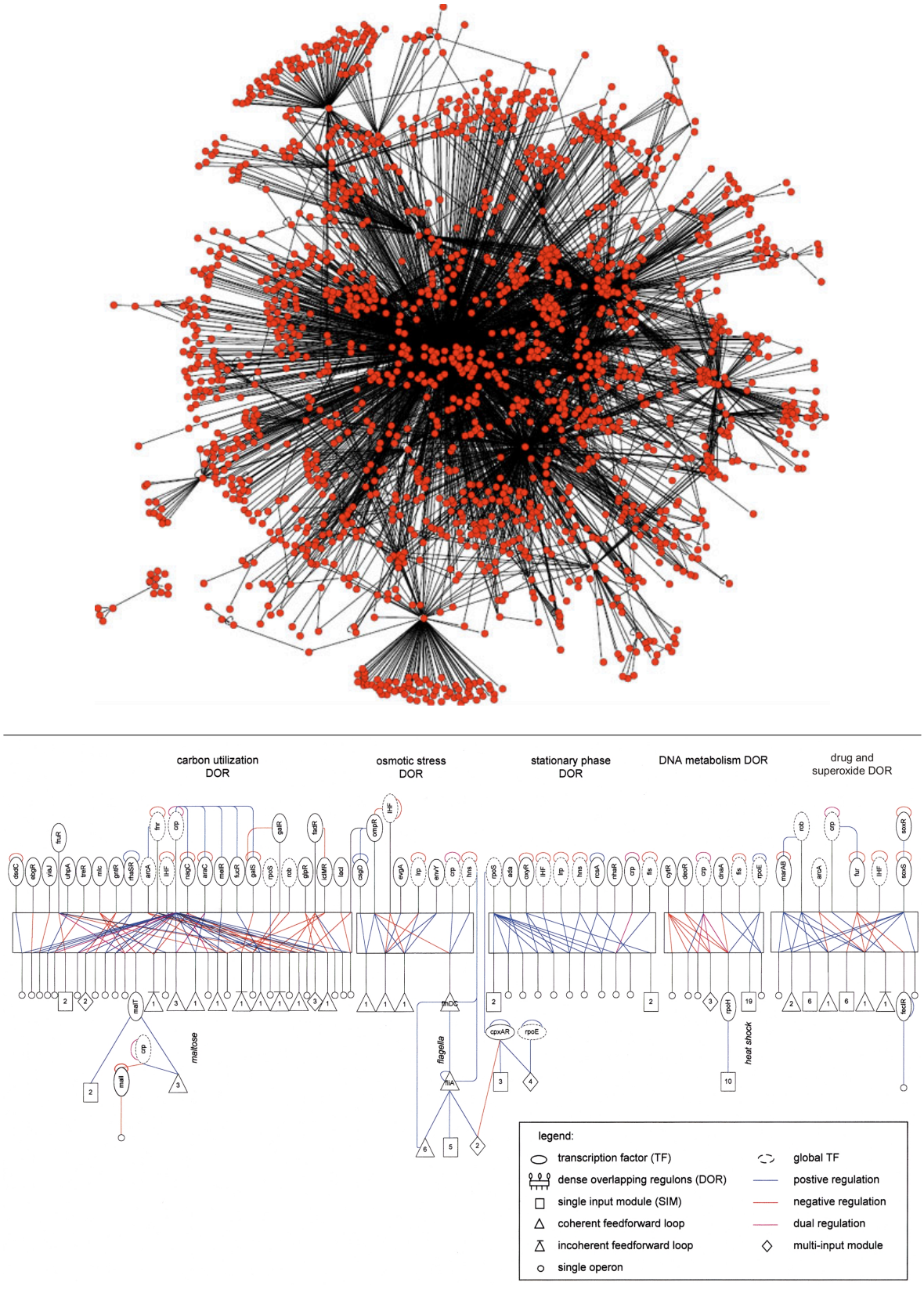


Figure 3.5: **Top:** Full representation of the transcriptional network of *E. coli*. *Source:* Freyre-Gonzalez and Trevino-Quintanilla (2010). **Bottom:** Compact representation of the same network using network motifs. *Source:* Shen-Orr et al. (2002).

DORs are considered to be the main 'computing' units of the network since each of

them regulates a whole set of 'output' genes by combining the incoming signals from a large group of transcription factors. A first important observation derived from the visual representation of the network is that it contains only one layer of DORs (represented by the big boxes in Figure 3.5, bottom). This is to say that there are no DORs regulating other DORs: "[M]ost of the computation done by the network is done at a cortex of promoters within the DORs" (Alon 2007, 90). Moreover, most of the other motifs appear to be integrated within this layer:

> The FFLs and SIMs are integrated within the DORs. Many of the FFLs are multi-output, with the same X and Y regulating several output genes. Negative autoregulation is often integrated within FFLs and also decorates the master regulators of SIMs. *Overall, the rather simple way in which the network motifs are integrated makes it possible to understand the dynamics of each motif separately, even when it is embedded within larger patterns.* (Alon 2007, 90), emphasis added)

Thus the structural analysis of a network in terms of motifs reveals a very particular and hierarchical organization of the whole network which also suggests modularity in functional terms: One can interpret the smaller motifs as carrying out particular dynamical subroutines, while the larger chunks in the network generate the complex computational responses that the cell employs to cope with environmental stimuli. This picture is further supported when taking into account the biological function of the output genes of the network. As the compact representation in Figure 3.5 shows, each of the DORs can be assigned to a particular biological function such as carbon utilization, osmotic stress etc. Overall, an idea of comprehensive understanding of the behavior of the whole network emerges, even though many details, notably the functional analysis of the less simple modules, still need to be filled in.

Heuristic Aspects of the Network Motif Approach

As we have just seen, the strategies of searching for network motifs and analyzing their dynamical properties serve the broader epistemic task of explaining the behavior of large networks. I have described them in some detail in order to give an impression of the way in which they reduce the complexity of this task.

Differently from the modeling strategies discussed in Section 3.2, the network motif approach does not apply the heuristics of decomposition and localization—apart from the fact that it takes the network as a whole as responsible for a particular function of the whole organism. Even though the aim is ultimately to explain the behavior of a complex system, this behavior does not directly guide the investigation as in the case of the traditional approach. Instead, the first step is to represent the structure of the complete system topologically and to look for peculiarities in this structure. The success of this strategy relies on several assumptions about the system, and the best way to reveal these is perhaps to discuss some of the criticisms that have been put forward against the network motif approach.

These criticisms come from various directions. Some scholars object to specific technical issues, while others raise more substantive concerns. As we will see, however, they are all related since the technical assumptions are often inseparable from the way in which the underlying biological system is conceived and represented. Evidence undermining the biological significance of network motifs is often produced with the help of computer simulations of network evolution, while other types of studies are based on the analysis of existing empirical data, or on targeted experiments.

The work by Artzy-Randrup et al. (2004), for instance, scrutinizes the network motif approach from a statistical perspective, interpreting Alon's strategy as an instance of hypothesis testing. The null-hypothesis in this context is that a particular structure in the network has no functional significance. The authors argue that the method of finding motifs based on comparison with randomized graphs "can lead to the wrong interpretations if the underlying null-hypothesis is not posed carefully" (Artzy-Randrup et al. 2004, 1107). In order to illustrate the potential dangers, they construct 'toy networks' with different generation rules and show that these can reveal significant abundance of network motifs when compared to the random graphs used, for instance, by Milo et al. (2002), even though they clearly have no function. One of their networks, for example, is based on a neighbor relation between different nodes, which can be interpreted as resulting from a spatial arrangement of the nodes, like that of neurons on a neural connectivity-map. In the generation of this network, the nodes are connected to nearby neighbors with higher probability. Another model follows the rule of 'preferential attachment,' which has

been argued to apply to various biological scale-free networks. According to this rule, new nodes connect preferentially to nodes that are already well-connected. In both of these cases the authors found that some motifs, notably the FFL, are significantly over-represented. Since these toy networks do not include any selective rules, they are alternative candidates for null-models in the search for network motifs. As the authors suggest, the choice of a different model for comparison potentially leads to very different results. Their overall point, however, is not to dismiss the network motif approach completely, but merely to point out the technical difficulties concerning the choice of the right null-model:

> As such, the actual process by which a network is generated, even if it is free of selection for or against particular motif functions, can strongly bias an analysis that seeks to determine the quantitative significance of motifs. (Artzy-Randrup et al. 2004, 1107)

They conclude with the cautious remark that "these techniques need to be developed further before design principles can be deduced with confidence" (Artzy-Randrup et al. 2004, 1107).

Other critics have objected more directly to the biological perspective underlying the network motif approach. The arguments they put forward, however, focus on the same basic issue of the right model for comparison and simply push the point further. The network motif approach presupposes that the processes generating the randomized and the real networks are equivalent in all relevant respects, except for the occurrence of selective pressures to which only the real network is subjected. Deviations in structural properties from the randomized networks are then interpreted as direct consequences of these selective pressures. Consequently, the overabundance of certain network motifs is taken as a strong indicator of adaptive value. This line of reasoning has been criticized on the grounds that the randomized models might be inadequate representations of the actual processes of network growth and evolution in the absence of selective pressures. In a short review article, called *Are network motifs the spandrels of cellular complexity?*, Solé and Valverde (2006) argue that network motifs are likely to be by-products of the rules of genome growth. This means that the motifs are indeed examples of *spandrels*, in the terminology that was introduced in a famous article by Stephen Jay Gould and Richard

Lewontin (1979). Originally, 'spandrel' is an architectural term that refers to the triangular space between two arches (Figure 3.6), and it was discussed by Gould and Lewontin

Figure 3.6: One of the spandrels of St. Mark's cathedral in Venice. *Source:* Gould and Lewontin (1979).

as an analogy for a particular type of evolved organismal feature. With reference to the particularly impressive exemples of spandrels that can be found in St. Mark's Cathedral in Venice, they write:

> The design is so elaborate, harmonious and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense of the surrounding architecture. But this would invert the proper path of analysis. The system begins with an architectural constraint: the necessary four spandrels and their tapering triangular form. They provide a space in which the mosaicists worked; they set the quadripartite symmetry of the dome above. (Gould and Lewontin 1979, 582)

By means of this analogy, Gould and Lewontin wanted to draw attention to the weaknesses of an exaggerated 'adaptationist program,' according to which each apparently useful feature of an organism requires an adaptive explanation. Against this view, which they perceived as mainstream in evolutionary biology, they pointed out that there were many organismal features that owed their existence to processes other than selection for adaptive value. Like the spandrels of St. Mark's cathedral, many traits appear simply as by-products of other, genuine adaptations.

Applying this term in the context of network motifs, Solé and Valverde directly object to the idea that motifs have been individually selected due to their functional benefits. While Artzy-Randrup et al. merely wanted to direct attention to the possible biases introduced by inadequate statistical methods, Solé and Valverde make the much stronger claim that network motifs are no more than epiphenomena of the process of network evolution. In support of this position, they cite recent work on artificial regulatory networks (ARNs) by Kuo et al. (2006). Such networks are generated by processes that simulate the evolution of real gene networks, involving, for instance, genome duplication events and sequence divergence. The resulting networks are found to share global features, such as the property of being scale-free, with real biological networks. But more importantly, they show distributions of network motifs that are very similar to those found in the regulatory networks of *E. coli* or *S. cerevisiae*. Since no selective processes were simulated in the generation of the ARNs, the authors conclude:

> [T]he topologies obtained are directly related to the method of construction. This might indicate that such topologies in natural networks may be a result of the way they are created rather than being explicitly molded by evolution. (Kuo et al. 2006, 192)

A complementary approach to study the functional role of network motifs *in silico*, simulating the application of selective pressures on artificial networks, was adopted by Knabe et al. (2008). They evolved groups of ARNs to exhibit particular behavioral responses and subsequently compared them among each other and to a group of randomly evolved networks. The initial expectation was that the algorithm would lead to the selection of useful motifs specific to the imposed functional requirements. Contrary to this expectation, however, they observed considerable variation of subgraph patterns between

the networks in each condition. Moreover, the differences in pattern distribution between the different conditions were not statistically significant. Reflecting on their results, the authors note:

> [O]ne might expect that motifs reflect evolved function. However our results show this view may be too naive—there was no convergence on the same single motif or a small set of switching motifs, and uniqueness of motifs was not observed. Instead a wide variety of network patterns and topologies was found. (Knabe et al. 2008, 73)

These results support the idea that network motifs are by-products of the network generating process and do not reflect design features that have been optimized by natural selection. By itself, however, this does not imply that they are functionally meaningless. In fact, the very comparison to 'spandrels' alludes to the possibility of playing particular roles within the network, in spite of not having been selected for these roles.[3]

Another article cited in Solé and Valverde's review, however, makes the even stronger claim that network motifs *are not* functional units. Mazurie et al. (2005) created and analyzed an 'integrated network representation' of *S. cerevisiae*, containing information about both transcriptional regulation and protein-protein interactions. Their analysis reveals that only four instances of motifs occur in isolation, whereas the large majority (= 500) are integrated into larger sub-networks. Furthermore, they studied "in detail the role of motifs in the case of the best-documented genetic sub-networks and biological functions where such motifs are found" (Mazurie et al. 2005, R35.5). In virtually all of the pathways and subsystems they looked at, they found that the motifs do not seem to play any central regulatory role. Considering these findings, they stress that usually several layers of regulation must be taken into account to get an adequate idea of biological function:

> At the moment, it is a fact that all the examples studied highlight the high level of integration of different regulatory mechanisms acting altogether [*sic*]. Reception and processing of cellular signals cannot be reduced to transcriptional regulation and protein-protein interaction switches. Other mechanisms

[3] One of the central points stressed by Gould and Lewontin was that there could be features with *adaptive value* that are not *adaptations*, the latter comprising only those that have been *selected for* their adaptive value.

> such as phosphorylation, triggered degradation, protein sequestration and transport, and higher-order multimerization are central to the logic of the sub-networks .... A qualitative impression surmised from the visible aggregation and nesting of the motifs with the rest of the network is that a 'pure' modular functional behavior is not very likely to occur. (Mazurie et al. 2005, R35.8–9)

This suggests that the network motif perspective underestimates the complexity of the actual system because it focuses only on some properties while ignoring others. In fact, this reveals the studied system as an 'interactionally complex' one (Wimsatt 1972; for further discussion see Chapter 1). Mazurie et al.'s results suggest that in order to understand its behavior, it is necessary to complement the network perspective with a molecular approach that is able to incorporate additional types of interactions.

The perhaps most interesting piece of evidence regarding the functional significance of network motifs was directly produced in the laboratory.[4] In an impressive series of experiments, Isalan et al. (2008) constructed 598 *E. coli* strains carrying artificial gene constructs to create the effect of network rewiring due to gene duplication events. Each of the constructs consisted of (the transcribed region of) a gene, coding either for a transcription factor or for a transcription initiation factor, fused to a new promoter. In this way new network paths were created connecting the inputs of the regulatory region to different outputs. Many of these insertions created radical changes in the network topology since they affected connections "at the top of the network hierarchy" (Isalan et al. 2008, 840), including transcription factors regulating hundreds of other genes. Strikingly, they observed that 95% of the new networks were well tolerated by the bacteria, and contrary to commonly held assumptions they found that, "at least when it comes to altering regulatory inputs, the hub genes do not appear to be the Achilles' heel of the network" (Isalan et al. 2008, 840). Even though the author's main interest was to assess the potential of such rewired networks for evolvability, they also addressed implications for the viability of the network motif approach:

> Overall, the results indicate a very complex rewired network response, sug-

[4] The work of Isalan et al. has also been discussed from a philosophical perspective by O'Malley and Soyer (2012), with a focus on the role of integration in systems biology.

> gesting that dissection into small network motifs may only lead to useful insights in some cases. (Isalan et al. 2008, 841)

And again towards the end of the article:

> [P]artition of a network into small modules (negative feedback, feed-forward, and so on) could in some cases be misleading, as the behaviour of these modules is affected to a large extent by the rest of the network in which they are embedded. (Isalan et al. 2008, 844)

The idea that a network can somehow respond 'globally' to dampen the effect of perturbations is in general not easily reconciled with a modular perspective. The sometimes puzzling nature of biological robustness is surely part of the reason why it is such a widely discussed topic in systems biology (e.g. Kitano 2004). In Chapter 4 the issue of explaining such robustness will be taken up again and discussed in more depth.

Overall, the discussion has revealed a variety of criticisms of the network motif approach. Some of them can be read as a request to be more careful in choosing a null-model, or at least to be more cautious about the interpretation of the results. These types of criticisms have inspired more refined strategies to prove the significance of network motifs. Other critics, as we have seen, more generally question the applicability of 'inverse approaches' to biological networks. Studies like the one by Knabe et al. (2008) suggest that a mapping from network topology to function is not obvious, and that often many different topologies are able to perform the same function. In response to this, there have recently been very detailed investigations of the potential complexity of 'function-topology maps' in biological networks. Ma et al. (2009), for example, systematically explored circuit architectures that are capable of a particular behavior called 'adaptation,' the ability of a sensory network to respond to a change in input stimulus and then to return to the initial level, even when the stimulus persists.[5] They computationally investigated the complete set of 13 608 three-node topologies and for each of these searched through 10 000 sets of kinetic parameters in order to find circuits displaying adaptation. With respect to the general idea of heuristics, it is interesting to see how they justify their exclusive focus on small topologies:

[5] Adaptation is a common feature of sensory systems in biology. A well-studied example is the phenomenon of chemotaxis in *E. coli* (e.g. Barkai and Leibler 1997). An analysis of this case study from a philosophical perspective can be found in Braillard (2010).

> Although most biological circuits are likely to have more than three nodes, many of these cases can probably be reduced to these simpler frameworks, given that multiple molecules often function in concert as a single virtual node. By constraining our search to three-node networks, we are in essence performing a coarse-grained network search. This sacrifice in resolution, however, allows us to perform a complete search of the topological space. (Ma et al. 2009, 762)

Their method thus involves an explicitly acknowledged trade-off between computability and resolution. As a result of their analysis, they find that there are only two classes of solutions that achieve robust adaptation: A negative feedback loop with a 'buffer' node and an incoherent feedforward loop with a 'proportioner' node. The fact that there is only a limited number of solutions for a particular functional task revives the hope of understanding network behavior in terms of motif decompositions.

The network motif approach is valuable as it suggests a way in which we can understand the behavior of large biological systems. It reduces the complexity of this task by assuming a particular organization that is, similar to the traditional approach, heavily based on functional modularity. However, some of the criticisms that I discussed seem to point towards the fundamental worry that the decomposition into modules is not a good heuristic strategy for the understanding of large biological networks. If modularity—as a topological property—turns out to be a mere by-product of network growth, and not the result of an explicit selection for specific motifs, then a separate argument is needed to license the inference from structural to functional decompositions. In the absence of such an argument, one might still hold that decomposition is the only way to obtain tractable representations of networks. Yet, one can also find approaches in systems biology that aim at an understanding of global network behavior without relying on any functional decomposition. One of these will be discussed in the following section.

3.3.2 Tracing Global Network Behavior: Cell Fate Attractors

In his 2004 article *Back to the biology in systems biology: What can we learn from biomolecular networks?*, the systems biologist Sui Huang draws attention to 'globalist' approaches within systems biology. He observes a clear divide between two camps of biologists in

general (Fig. 3.7) and argues that, regarding their vision of biological complexity, even systems biologists mostly remain within a 'localist' perspective that continues in the path of molecular biology:

	The 'localist' ('particularist') view (Those who see the trees first)	**The 'globalist' ('generalist') view (Those who see the forest first)**
Level of original focus	Gene- and pathway-centric	Network-centric
New field created	'Systems biology'	'Biocomplexity'
Use of hypothesis	Hypothesis at level of individual pathways. No systems-level hypothesis: research becomes 'discovery-driven'. Example hypotheses: 'Gene A inhibits Gene B, is required for function X, etc.	Hypothesis at systems level concerning generic design of network and network position of genes Example hypotheses: 'Hub proteins are important' 'Power-law architecture favours ordered dynamics'
Philosophy	System is complicated Properties of systems lie in the property of the components Comprehensiveness. The whole equals the sum of the parts	System is complex Higher-order system properties emerge from collective behaviour of components Holism The whole is different from sum of parts
Practical aims of study	To characterise exhaustively the biochemistry of (all) individual pathways and their 'functions' To describe idiosyncrasy	To understand generic aspects of genome-scale networks as an entity with its higher-order properties To understand universality
Gene identity in models	Of primary interest. Specific models with nominal genes and their idiosyncratic properties	Of secondary interest. Models with anonymous genes as generic entities may sometime suffice
Network topology	Precise biochemical characterisation and categorising of physical and regulatory interactions in specific pathways	Analysis of large scale features, based on global statistics of local network features (degree distribution, modularity, clustering)
Network dynamics	Detailed modelling of individual small circuits (modules) in separation as a low-dimensional dynamic system	Global dynamics of network maps into whole cell behaviour (cell fates)
Function	Focus on local cellular functions (eg protein synthesis, vesicle transport, filopodia extension, DNA repair, etc) associated with a specific pathway	Emphasise emergent whole-cell behaviour, such as switch between discrete cell phenotypes (cell fates)
Typical non-biologist partners	Computer scientists, engineers	Physicists

Figure 3.7: Huang's distinction between localist and globalist views in biology. *Source:* Huang (2004).

> [T]he localists' view is rooted in classical molecular biology, hence is shaped by decades of devotion to the study of individual cellular pathways that represent to them linear causal relationships. But the prevalence of pleiotropy and convergence in cell signalling, and of crosstalk between pathways, has led to the increasing awareness that understanding gene function requires that one reaches beyond the narrow focus on individual pathways. (Huang 2004, 284)

According to Huang, a 'globalist' approach is necessary to capture higher-order properties of genome-scale behavior. Following such an approach, however, implies leaving behind the 'engineering perspective' inherited from traditional molecular biology with its focus on small circuits and its assumptions of local optimization. He urges instead that systems biologists move towards a 'physics perspective' that investigates the generic properties of large complex systems in the spirit of Stuart Kauffman (e.g. Kauffman 1969, 1974, 1993). The results of the application of network theory to biological systems play an important

role in his own approach. However, he appears unconvinced of the network motif approach and above all emphasizes those features of large biological networks that defy the localists' efforts of decomposing them into small modules. An example of such a feature is the existence of so-called *giant components* in many biological networks. In the yeast protein-protein interaction network, for instance, a significant fraction (> 75%) of all proteins belong to one large cluster of connected nodes (e.g. Yook et al. 2004), a finding which, even though it does not conflict with network modularity *per se*, points to the relevance of coordinated behavior at the systemic level. Moreover, differently from the network motif approach, which took network *structure* as the exclusive starting point, Huang underlines the importance of system *behavior*. A global approach is necessary, on his view, to account for the fact that complex wholes, such as cells, tissues, or organs, often show a coherent behavior that appears to be both simple and robust. The prime example for Huang is cellular differentiation:

> [C]ells in multicellular organisms exhibit a simple, coherent whole-cell behaviour which may precisely reflect a higher-order dynamics of the global network: the switching between cell fates. This strictly regulated, rule-based systems behaviour is robust and *remarkably simple compared with the complexity of the underlying molecular network*. (Huang 2004, 291), emphasis added)

Huang's strategy to understand this simple behavior at the cellular level, is to apply the conceptual apparatus of dynamical systems theory to the representation of large networks. The basic idea is to describe the state of the network at a given moment as a vector $S(t)$ in a high-dimensional state space, whose elements are the states of all the individual components of the network (e.g. molecular concentrations). The state of the system can thus be thought of as a point that moves in the state space, along a trajectory that is dictated by the dynamical rules. It has to be admitted that the idea of directly investigating the dynamic properties of whole-cell models consisting of thousands of variables is, at least at present, far-fetched (for the state of the art, see Section 3.4). On the one hand, because precise measurements of most of the kinetic parameters to build such a model are missing, on the other hand—and more importantly perhaps—because the investiga-

tion of such a model would by far exceed the currently available computational powers.[6] In spite of these obvious shortcomings, the theoretical framework of dynamical systems theory can nevertheless be of use as it provides a number of conceptual tools for the rationalization of qualitative systems behavior. Perhaps the most prominent of these is the notion of an *attractor*. An attractor is a set in state space towards which the system evolves if it is initially within a nearby region (called the attractor's *basin of attraction*). The most straightforward examples of attractors are stable equilibrium states (point attractors) and periodic orbits such as limit cycles, but more complicated cases have been found, such as the *strange attractors*, with fractal structure, that are investigated in chaos theory. Attractors are stable by definition, which is to say that a system will remain in or near an attractor, even if it is slightly perturbed. This property is especially interesting when it comes to understanding the robustness of biological phenomena.

Huang's aim is to show that the concepts of dynamical systems theory, notably the notion of attractor, can be productively applied to the investigation of cellular development:

> [T]he dynamics of a network with attractor states naturally captures the essential properties of cell fate dynamics, including mutual exclusivity, robustness and all-or-none transitions between cell fates in response not to a single 'specific' instructive signal but to a large variety of signals. (Huang 2004, 292)

At the same time, it is a perspective on networks that goes beyond the purely 'topological' approaches that mainly focused on the patterns of connections between nodes:

> Most functional interpretation of networks has been based on their topology alone . . . . Much of the topology-based reasoning about function rests on the unarticulated premise that the molecular network acts like a communication network in which some information 'flows' in the links from node to node. (Huang 2004, 289)

[6] In an earlier article, called *The practical problems of post-genomic biology*, Huang makes the following estimate: "Take the human genome with 100,000 genes and let every gene be simply either 'on' (expressed) or 'off' (silent). This minimal, idealized, and discrete setting alone would lead to the astronomical number of $10^{30,000}$ possible gene expression profiles! The computing and testing of all these patterns with the existing serial computers would take more time than the age of the universe" (Huang 2000, 471). The fact that the number of genes in the human genome is nowadays known to be substantially smaller ($\approx 30,000$) does not affect the general conclusion of the argument. Another, very thoroughly derived, estimation of the computing power necessary to simulate complete models of biological networks, with the same qualitative conclusion, can be found in Gatherer (2010).

Discussing the mathematical models of the spindle assembly checkpoint in Section 3.2, I noted how dynamic features put in question the use of the informational perspective on biological systems by revealing how different activities in a process can be dependent on each other. This becomes even more pronounced when considering even larger dynamical systems. The movement towards a high-dimensional attractor, for instance, does not correspond to a well-defined sequence of molecular events. In theory, there are an infinity of possible ways in which the attractor state can be reached.

Historical Precursors of the Attractor View in Systems Biology

The idea of describing cell fates as attractor states of a dynamical system reveals a direct connection with ideas that had been entertained as early as in the 1930s by theoretically minded biologists, notably by Conrad H. Waddington (1905–1975).[7] Waddington was considered the leading British embryologist and geneticist from the 1930s throughout the 1950s, and is today best known for coining the visual metaphor of the 'epigenetic landscape.' This metaphor is often used to illuminate processes of biological development by comparing them to a marble rolling down an inclined surface (Figure 3.8). The particular shape of the surface, with hills and valleys, creates preferred paths and branching points for the marble, corresponding to developmental trajectories and decision points that eventually lead the developing system towards one of several possible end points or 'fates.' The relationship to concepts from dynamical systems theory becomes obvious when considering that the landscape in the picture corresponds to a potential energy surface for the marble in a gravitational field. The local minima on such a surface are straightforward examples of attractor states.

Waddington himself was very interested in a mathematical formalization of his ideas on development and inspired other scholars to engage with him in this task. Among them was René Thom, a French mathematician, who is considered one of the founders of 'catastrophe theory' which gained considerable popularity in the 1970s. Thom attended the famous Bellagio conferences on theoretical biology that were organized by Waddington, and he developed many of the notions of catastrophe theory, such as the attractor concept, in close correspondence with Waddington (Aubin 2004).

[7] Others who entertained similar ideas are the physicist Max Delbrück and the biologist Jacques Monod (for more on the historical background, see e.g. Keller 2002, Chapter 5).

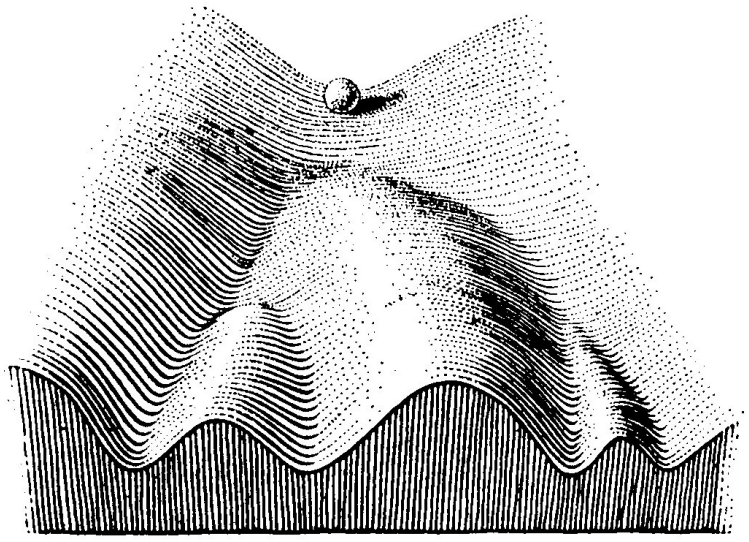


Figure 3.8: Representation of Waddington's epigenetic landscape. *Source:* Waddington (1957)

Another scholar who was directly inspired by Waddington's ideas was Stuart Kauffman, who has already been mentioned several times. A participant at the Bellagio meetings like Thom, he started investigating the properties of large Boolean genetic networks in the late 1960s (Kauffman 1969, 1974, 1993). Differently from Waddington and Thom, however, he directly built on knowledge that by then had been accumulated by molecular biology. For example, he made explicit reference to Jacob and Monod's work on gene regulation to motivate an analogy between genetic switching circuits and computers and to defend his use of a Boolean approximation in building his models (Kauffman 1969, 438).

When Kauffman developed his first models, there was, however, virtually no empirical information about the overall architecture of genetic networks. Hence, his idea was to build virtual networks from possible 'small scale elements,' even if the details were not fully known, and to see whether such networks would predict some of the large scale properties observed in biological organisms. He deemed such a perspective on large scale properties necessary since by trying to understand the system piece by piece one might end up missing important aspects:

> [W]e should consider ways to construct an adequate picture of the architecture of cell control systems whose full details may never be directly known. In addition, incomplete knowledge of those control systems poses the critical problem that there are likely to be dynamic properties of central biological importance which depend in some way on large portions or on the whole

> organization of the control system, not on small isolatable fragments of it. (Kauffman 1974, 168)

Thus one might say that by neglecting the specific molecular details of biological networks, Kauffman was simply making a virtue of necessity. His main heuristic move was thus to tentatively interpret the dynamics of living systems as guided by what Weaver (1948) called 'disorganized complexity' (see Chapter 1). Just like theoretical physicists had been able to master the apparent complexity of large disordered systems, one might be able to find conceptual tools to reduce the complexity of large biological systems. Consequently, Kauffman referred to his strategy as an "ensemble approach" (Kauffman 1974), adopting this terminology from statistical mechanics.

The networks he investigated are randomly generated Boolean networks with nodes representing genes that can be in two different states, 'on' or 'off.' In these networks the dynamics unfold by updating the state of each gene at successive discrete time steps. The state of each gene is determined by the states of a specific set of 'input genes' at the preceding time step. These can be interpreted as transcription factors regulating the activity of a common target gene. Formally, each gene involves mapping of set of binary arguments on a single binary value, that is, it realizes a Boolean function. A particular network is constructed by first specifying the number N of genes and the number K of inputs to each of them. To each gene then are assigned K inputs and one of the 2^{2^K} possible Boolean functions. Starting from an arbitrary initial condition, the state of the network develops by evaluating the Boolean function for each gene at each time step $t = 0, 1, 2, \ldots$, and by assigning to it the resulting value at $t + 1$. Kauffman's model only considers the interactions between genes in the network and does not take into account any external inputs. Since the system is deterministic and has a finite number of possible states, it will unavoidably return to a state that it has already previously passed, and from then on repeat the same sequence, or *cycle*. To each cycle (which may consist of only one state) corresponds a set of states leading into that cycle, which Kauffman refers to as a *confluent*. These cycles are examples of attractors in a discrete dynamical system, and the confluents are their basins of attraction.

Investigating networks of low connectivity ($K = 2$ or $K = 3$), both analytically and numerically, Kauffman found some of their properties reminiscent of the behavior of biolog-

ical cells. In most networks he found a surprisingly small number of cycles, compared to what would have been combinatorially possible. Moreover, these cycles tended to be very short. He interpreted these cycles as corresponding to the different 'cell types' that can be exhibited by genetically identical cells. Extrapolating from the relationship between the size of the network and the number of cycles derived from his simulations, he was able to predict reasonable estimates for the number of different cell types in various species of multicellular organisms (Kauffman 1969).

Kauffman's approach relies on the assumption that by investigating *typical* instances of a random collection virtual networks, one can get an understanding of the *particular* instances found in nature. This in turn presupposes that, during the process of their evolution, these networks have retained, or acquired, an essentially random structure at the large scale. If, on the contrary, we assumed that the structure of biological organisms was precisely specified by the influence of natural selection, it would be of little use to study networks created at random. Encouraged by his results, Kauffman turns this line of reasoning around: the fact that random networks explain many aspects of living organisms, such as stability, multicellularity, etc., suggests that nature exploits the orderly properties of random structures:

> Large, randomly assembled nets of binary elements behave with simplicity, stability, and order. It seems unlikely that Nature has made no use of such probable and reliable systems, both to initiate evolution and protect its progeny. (Kauffman 1969, 466)

Seemingly complex behaviors of biological systems might thus find an explanation that does not require uncovering all of the underlying molecular details.

Kauffman's theoretical work has exerted a direct influence on contemporary systems biology, and researchers like Sui Huang see their work in direct continuity with his modeling efforts.[8] However, the increasing interest in global perspectives on cellular behavior has also been due to findings in experimental stem cell biology.

[8] In fact, there are a number of articles that are the result of a direct collaboration between the two researchers (e.g. Huang et al. 2009, Foster et al. 2009).

Connections with Contemporary Stem Cell Biology

Waddington's ideas on biological development have also been revisited in the recent experimental literature on stem cell biology and cellular differentiation. Observations of cellular plasticity, de-differentiation, trans-differentiation, and notably, the 'reprogramming' of terminally differentiated cells into a pluripotent state, have led to a major rethinking of some of the basic ideas of the field. A quote by stem cell biologist Peter Andrews may serve as an illustration of this phenomenon:

> [T]he recent reports of stem cells from different adult tissue, displaying quite unexpected plasticity and apparent lack of specific commitment, [suggest] that perhaps the concepts of unidirectional, irreversible differentiation along distinct cell lineages should be revised. (Andrews 2002, 412)

The fact that plasticity and lack of commitment is perceived as a 'quite unexpected' finding reveals a number of implicit assumptions in the traditional picture of cellular development. Assuming 'unidirectionality' and 'irreversibility' suggests the idea of cellular differentiation as a deterministically executed program. The roots of this view can be traced back to the early period of molecular biology:

> During the 1950s and 1960s, many [molecular geneticists] treated differentiation in terms of the regulation of protein synthesis. The underlying hypothesis was that differentiation is an irreversible commitment of a cell lineage to the manufacture of a coordinated set of "luxury" proteins—i.e., specialized proteins not needed to maintain the life of the cell. Thus, the primary differences among nerve, kidney, skin, and blood cells were thought to depend on the specialized sets of proteins that they make, which, in turn, affect their morphologies, interactions with other cells, and responses to biological signals and stimuli. (Burian 1993, 391)

In this perspective, we clearly see the heuristics of decomposition and localization at work. First of all, genes are identified as the locus of control of differentiation, essentially following Jacob and Monod's suggestion that "differentiation operates at the genetic level, using elements basically similar to those found in bacteria" (Jacob and Monod 1963, as cited in Keller 2002, 166). Differentiation is conceived in terms of the 'switching on' or

'off' of specific genes. The distinct phenotypes and observed behaviors of differentiated cell types are localized in the activities of disjunct sets of specialized genes. This allows biologists to understand cellular differentiation as the execution of a 'genetic program,' and to reduce the epistemic complexity of understanding differentiation by assuming that this program controls distinct modules, or subsystems of genes, and is triggered in response to external stimuli via specific signaling pathways.

The observed cellular plasticity challenges the notion of a differentiation program since it implies that the cell can deviate from, and even revert, its typical path of differentiation. In order to accomodate these findings, it is necessary, according to Andrews, to adopt a broader perspective on the mechanisms of cell differentiation:

> When considering the factors that regulate cell behaviour, whether commitment and determination, or differentiation, attention commonly focuses on individual signalling pathways by which cells respond to external cues, e.g. growth factors, the extracellular matrix, or interactions with other cells. To keep the analysis simple, such signalling pathways within a cell are often considered in isolation, and are also considered as simple switches—either 'on' or 'off'. However, any molecules within a signalling pathway will obey the normal chemical laws affecting reaction rates and equilibria. The activity of particular regulatory molecules will be influenced by the overall state of all the other regulatory and metabolic reactions taking place within the cell. (Andrews 2002, 412)

This suggests moving away from a conception of 'genetic regulation' towards one of 'cellular regulation' (cf. Keller 2002, Chapter 5). Even though Waddington's image emphasizes the stability of developmental pathways, it also lends itself to intuitive ideas of plasticity and reprogramming. The hills between the branching valleys are not necessarily insurmountable obstacles, and one can easily imagine that the marble, when given the right kind of 'push,' can end up in a different valley, or even back at the beginning of the track. The metaphor of the epigenetic landscape, therefore, provides an important tool for stem cell biologists to conceptualize their experimental findings (cf. Figure 3.9).

The question that arises in the context of systems biology, however, is whether such an updated picture of the epigenetic landscape can go beyond the role of a metaphor and

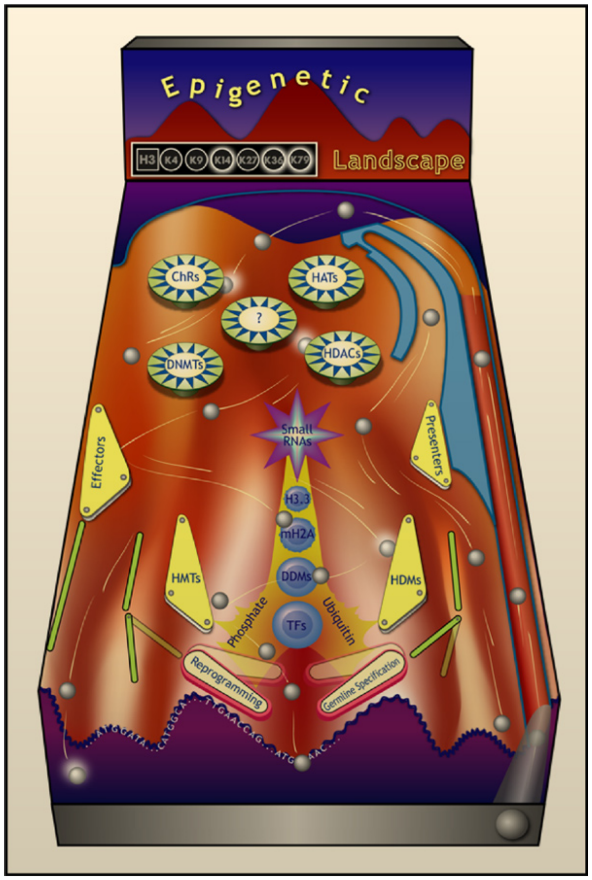


Figure 3.9: Contemporary interpretation of the epigenetic landscape, adding stochasticity and recent concepts and ideas from stem cell biology to Waddington's picture. *Source:* Goldberg et al. (2007).

suggest more concrete ways towards a global understanding of cellular development. Can network models, such as Kauffman's, provide an adequate 'mathematization' of Waddington's landscape? And can such models be brought into contact with the experimental activities of contemporary stem cell biologists?[9]

Empirical Investigations of Cell Attractors

A theoretical model like Kauffman's is very attractive because it proposes to explain a variety of observed properties at the macroscopic level while making only very few assumptions about the processes at the molecular level. Yet, it is not clear whether the adequacy of the proposed explanations can be established in practice since all simulations are performed on statistical ensembles of generic models. The ideal test, presumably, would be

[9] An interesting discussion of the integrative role of Waddington's Metaphor in current stem cell and systems biology can be found in a recent article by Fagan (2011). However, the article's conclusion regarding systems biology, according to which the landscape "is a derivational consequence of the ODE framework for representing molecular interactions, which visualizes the predictions of mathematical models in an accessible way" (Fagan 2011, 211), seems to understate much of the complexity of mathematical modeling.

to construct a complete and realistic dynamic model of a cell and to show that the attractor states of this model coincide with the measured molecular properties of the different cell fates. Of course, such a model is currently not available, for reasons that have already been discussed several times.

What is currently feasible, however, is to adopt a 'phenomenological' approach that takes advantage of the ability to study whole systems at the molecular level by means of high-throughput experimentation. Notably DNA microarrays, whose development dates back to the mid-1990s, have revolutionized the analysis of gene expression.[10] They allow biologists to monitor the expression of thousands of genes at the transcriptional (mRNA) level simultaneously. In this way the dynamics of gene expression can be tracked via time-series data of the transcriptional state. If the entire network of gene regulation is represented as one large dynamical system, one might consider such microarray experiments as a proxy for the state vector of this system. By studying the behavior of experimentally observed trajectories in the 'gene expression space,' one might be able to investigate the attractor landscape of cellular differentiation:

> [E]ven in the absence of knowledge of the specific network architecture, it is possible to use genome-wide gene expression profiling to probe the state space structure of a natural complex network and extract characteristic signatures of a stable high-dimensional attractor. (Huang et al. 2005, 3)

In order to illustrate this strategy, I will discuss an experiment that investigates the differentiation of *neutrophils* (Huang et al. 2005). Neutrophils are the most abundant type of white blood cells in mammals, and they are derived from a particular type of progenitors, called *promyelocytic* cells. It has been observed that these progenitors can be induced *in vitro* to differentiate into neutrophils by a variety of different stimuli. The starting point of Huang et al.'s study was the idea that monitoring different trajectories of the transcriptional state during the process of differentiation might reveal coherent genome-wide dynamics. Interpreted within the dynamical systems perspective, the state of the progenitor cell is initially a stable attractor. The different stimuli provide external perturbations that drive the system away from this initial state and towards the basin of attraction of

[10] Even though it seems that high-throughput sequencing techniques, such as RNA-Seq, are displacing microarrays as experiments of choice for transcriptome analysis.

the differentiated neutrophil state. One might understand the basic idea better by thinking of a marble that is trapped in a small depression on (a slightly modified version of) Waddington's landscape. An external 'push' can help it to continue its path along one of the valleys. Different kinds of pushes from the same starting position might result in very different trajectories, even though the eventual 'fate' of the marble is the same. What Huang et al. showed is that cellular differentiation towards the same cell fate can occur along different trajectories, just like the movement of the marble.

They induced the differentiation of neutrophil cells using two biochemically distinct stimuli (called atRA and DMSO), and monitored the transcriptional state of the cells over time with microarrays. An obvious difficulty consisted in the interpretation of the large and high-dimensional gene expression datasets that were thereby produced. Most commonly, microarray experiments are used to generate lists of 'signature genes' for a particular phenotype, that is, one looks for individual genes that significantly change in expression between different experimental conditions. However, Huang et al. started with the assumption that information about the behavior of individual genes would not reveal any clues about the global dynamics of differentiation. Moreover, such information is not very reliable due to the high levels of noise and intrinsic variability at the level of the single measurements.[11] There are, however, statistical techniques of dimensionality reduction that can be used to transform large datasets into a more manageable form, while retaining information about global behavior. Many of these methods rely on 'distances' in state space, which are calculated as straightforward generalizations of distance measures in geometrical spaces. One of the methods that was used by Huang et al. was principal component analysis (PCA), in which correlations within a large set of variables are exploited in order to find a representation in terms of a smaller set of uncorrelated variables (the principal components). If there is substantial correlation in the dataset, then two or three of the principal components can be sufficient to explain most of the variation between different datasets. A second method involved self-organizing maps, which are essentially artificial neural networks that are trained to produce low-dimensional representations while preserving some of the topological properties of the initial data. They can be used to create visual representations of large datasets that allow for a qualitative comparison

[11] Wimsatt (2007a) provides an interesting discussion of high-throughput experiments and ways to cope with their unreliability.

of different transcriptional states by eye.

Huang et al. applied these methods to the transcriptional data obtained from the two differently induced neutrophil populations. As can be inferred from the representations in Figure 3.10, the trajectories initially separate and move towards different regions of the state space. After a certain time, however, they begin to converge until the two populations show virtually identical expression patterns. From this the authors conclude:

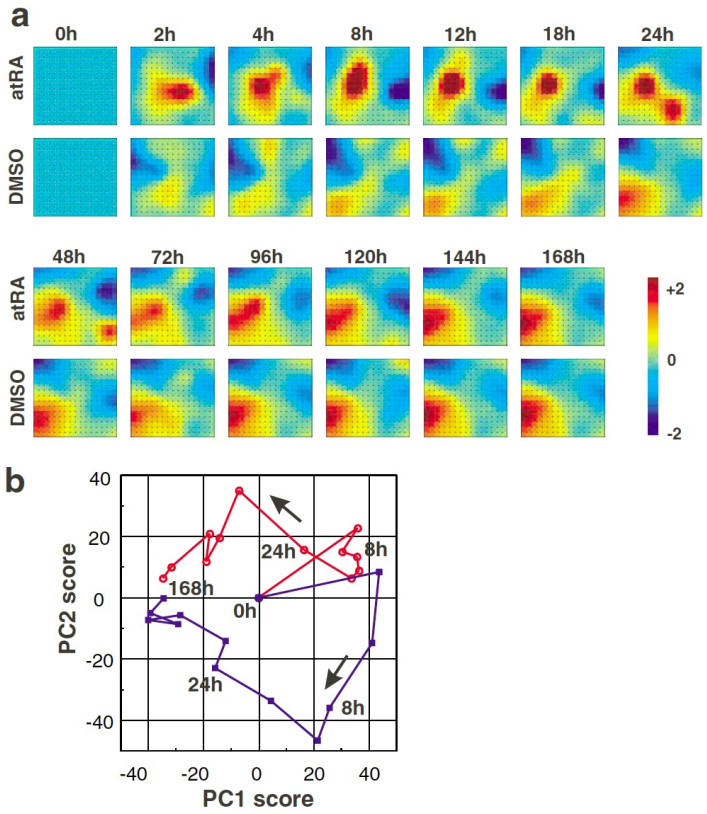


Figure 3.10: Convergence of two gene expression trajectories for a subset of $N = 2773$ genes during neutrophil differentiation. (a) The genes are clustered by a self-organizing map. Each of the 'mosaics' represents a snapshot of $S(t)$. Tile colors indicate the expression levels. (b) Principal component analysis. Each point represents an individual expression profile $S(t)$ within one of the two differentiation processes (red circles: atRA; blue squares: DMSO) projected onto the first two principal components (PC1 and PC2). *Source:* Huang et al. (2005).

> The convergence of trajectories from different directions across a large number of gene dimensions is a necessary condition for a high-dimensional attractor state and cannot be easily explained by the existing notion of a specific, unique "differentiation pathway" as the common target of the two drugs. (Huang et al. 2005, 1–2)

Thus they suggest that cellular differentiation should not be understood as driven by a 'program' that specifies the exact sequence of steps in the process. The alternative they

put forward instead conceives of it as a spontaneous relaxation process towards a state of 'minimal energy' at which all components of the system are in dynamic equilibrium.[12]

How 'Holistic' are Attractors?

Earlier, I introduced the attractor perspective as being in opposition to a 'localist' view that looks for understanding in terms of individual components or small modules of a system. In this regard, it is an interesting aspect of Huang et al.'s analysis that the identity of individual genes appears to be completely irrelevant. Genes and their products are treated as 'anonymous' particles of a system whose interactions generate regular behavior at a higher level of observation, almost like the molecules in a gas. Huang considers this as one of the characteristic features of a global approach (cf. Figure 3.7). The statistical data analysis in the neutrophil experiment revealed that "convergence occurred with respect to a large portion of the genome, i.e., to a high number of state space dimensions" (Huang et al. 2005, 2-3), and one might be tempted to infer that the global dynamics of the system is due to spontaneous self-organization of the parts in the spirit of Stuart Kauffman's theoretical analyses. It must be noted, however, that the experimental evidence provided by Huang et al. (2005) is purely correlational. Even if the low-dimensional description of the results, obtained e.g. by the principal component analysis, reveals that "thousands of genes in the complex network exhibit a globally coherent dynamic pattern of attraction to a common stable state" (Huang et al. 2005, 3), this does not imply the absence of hierarchical organization in the network. It might be that there are a small number of 'master regulator genes' that enslave large portions of the network, thereby causing them to follow their dynamics. At any rate, it is not obvious how one might distinguish between the two possibilities, hierarchy or self-organization, by solely observing the behavior of the trajectories in state space. Experiments of this kind, therefore, do not show that the strategy of localizing specific contributions to the behavior of a system in single genes or small modules is doomed to failure. As with the network motif approach, the utility of the attractor perspective depends on particular assumptions about biological organization.

Kauffman called his strategy an 'ensemble approach' since it uses the idea of statistical

[12]The interpretation of the epigenetic landscape as a potential energy surface is not straightforward. A potential function can only be found if the set of equations defining the system has very special properties. For an attempt to interpret it as a quasi-potential landscape that describes 'altitude' in terms of the probability of noise-induced transitions between attractor states, see Wang et al. (2010).

mechanics to acquire knowledge about a system even if many of the details are not known or impossible to incorporate in a tractable model:

> The trick of statistical mechanics is not to study a single system, but a large collection or *ensemble* of systems. Where understanding a single system is often impossible, one can often calculate the behavior of a large collection of similarly prepared systems. (Sethna 2006, 1)

It might help to illustrate this with an example. A theoretical model in statistical physics with widespread applications are *random walks*. These are processes with successive steps going in random directions. While it is impossible to predict the behavior of an individual instance of a random walk, one can nevertheless derive simple relationships about the statistical properties of an ensemble of random walks. For example, the endpoint of a random walk has a probability distribution that can be described in terms of the diffusion equation. Since such behaviors at the ensemble level are largely independent from the microscopic details of individual walks, physicists refer to them as *universal* (Sethna 2006, Chapter 2). Kauffman's theoretical analyses attempted to show that the number of cell types in a multicellular organism and their stability can be understood as universal properties of a particular type of network.

Researchers advocating the attractor view can potentially pursue different lines of research. First, they might try to turn Kauffman's speculative models into an actual explanation of cellular behavior. In order to do so, they must show that real biological networks, which are not necessarily approximated well by Boolean models and whose patterns of connectivity are expected to be different from those assumed by Kauffman, are nevertheless typical members of a statistical ensemble. Since the time of Kauffman's early works, biologists have accumulated considerable knowledge about transcriptional regulation in gene networks and might now be in the position to revise some of Kauffman's assumptions. In this context, Roger Sansom (2008) has recently challenged the applicability of Kauffman's models on empirical grounds, arguing that it is more adequate to describe them as *connectionist* models, known from artificial intelligence and cognitive science. Ultimately, there seem to be competing views about the primary explanatory role of natural selection in shaping the behavior of biological networks. In a recent article, Kauffman summarizes the alternative positions as follows:

> There is a fundamental ontological assumption underlying [the ensemble] approach, and it is not known if that assumption is true or false. Is it the case that the genetic network in an organism, or a species, or family of species, after 3.8 billion years of natural selection and evolution, is a highly crafted, "one off" design, brilliantly tuned by selection to achieve its functions? Or might it be the case that real genetic regulatory networks are more or less "typical" members of some class, or ensemble, of networks which selection has modified to some degree? In the latter case, we may be able to gain very considerable insight into the structure, logic, and dynamics of gene regulatory networks by examining the typical, or generic properties, of ensemble members. (Kauffman 2004, 582)

Even though I will not pursue this line of reasoning further, this strongly suggests that fundamental questions about biological evolution can be of considerable importance for debates on complexity in systems biology.

An alternative strategy for systems biology is to suspend the issue of whether attractors in networks are truly generic features, but nevertheless try to understand the behavior of large networks from a dynamical systems perspective. The recent work by researchers like Sui Huang mainly illustrates this second strategy. However, it seems that in order to obtain a more realistic picture of the dynamical system that underlies the processes of cellular development, biologists must zoom into smaller sections of the whole network and at least partly give up their 'globalist' ambitions and the anonymity of the genes. An example of this is provided by recent investigations of binary cell fate decisions.

Binary Cell Fate Decisions

The pool of specialized cells in vertebrate tissues is maintained due to the presence of stem cells and progenitor cells. Within the landscape metaphor, the progenitor states can be interpreted as 'branching points' at which a cell can commit to either of two distinct lineages. An example in the context of blood cell development is the common myeloid precursor cell (CMP) that can continue its differentiation path either in the erythroid/megakaryocyte lineage or in the myelomonocytic lineage. Two transcription factors, GATA1 and PU.1, have been shown to control the lineage specification for these developmental

paths (Orkin 2000). For instance, overexpression of either factor induces the differentiation towards one of the lineages. Moreover, each of the two factors can suppress the expression of the other. This simple schema of mutual inhibition supports the idea of a 'cell fate switch' that underlies the decision between mutually exclusive, distinct and robust cell fates. Even though the image of the switch appears to be perfectly compatible with the attractor view, it gives rise to several questions. One question regards the way in which the decision between the two fates is made. For instance, is it caused by a specific external stimulus or is the process essentially stochastic? Another question regards the progenitor state itself. Should it be considered an attractor state as well, and if so, how can its stability be explained given the tendency of the switch to push the cell towards a specific lineage?

In order to address these questions Huang et al. (2007) built dynamic models of the GATA1-PU.1 circuit. The first model they investigated described only the mutual inhibition of the transcription factors. Here they observed, as expected, a bistable behavior: the system has two stable attractor states in which the level of one factor is high while the level of the other is low, and one unstable equilibrium state with similar levels for both. These attractors can be interpreted as the erythroid lineage and the myelomonocytic lineage, respectively. Huang et al. then showed that adding positive autoregulation to this model can create a basin of attraction for the formerly unstable equilibrium state, turning it into a third attractor state. This state can be interpreted as the CMP progenitor state, which fits the observation that the progenitor cells express both GATA1 and PU.1 at intermediate levels (Cross and Enver 1997) and provides a possible explanation for the stability of the progenitor state.

Next, Huang et al. turned to the question of how the system manages to leave the progenitor attractor in the process of cell fate commitment. They hypothesized that the differentiation signal corresponds to a parameter change in the model and thus acts by affecting the topology of the attractor landscape. They envisaged that this could happen in two different ways. The signal might correspond to a directed, or asymmetric, parameter change that enlarges one of the basins of attraction of the differentiated states and thereby causes the system to 'drop' into that state. In this scenario the external signal is considered *instructive* because it clearly determines the fate, and trajectories leading to

different fates would be expected to initially go in opposite directions. Alternatively, the signal might lead to a symmetric modification of the attractor landscape, in which the progenitor state is destabilized by being transformed from a 'well' into a 'hill top' or 'watershed.' In this scenario the initial transcriptional changes should be similar for both fates: after receiving the signal, the progenitor cell remains for some time in an unstable state 'on the top of the hill.' In order for the cell to subsequently 'drop' into one of the remaining attractor valleys, a symmetry breaking event is needed. In the real system such an event might occur due to stochastic fluctuations of regulatory factors inside the cell.

Huang et al. went on to study the qualitative behavior of the model when simulating the two hypothesized kinds of signals. They found that the resulting trajectories in the GATA1-PU.1 phase space could be classified into qualitatively distinct types. Notably, when they triggered differentiation with the symmetric destabilization of the progenitor attractor, they found a characteristic loop in one of the two trajectories (Figure 3.11, **A**). Using microarrays, they experimentally monitored the trajectories of differentiating cells in the GATA1-PU.1 plane and compared them to the theoretically obtained results. Interestingly, they observed that the real cells show the same kind of loop that was predicted by the model with symmetric destabilization (Figure 3.11, **B**).

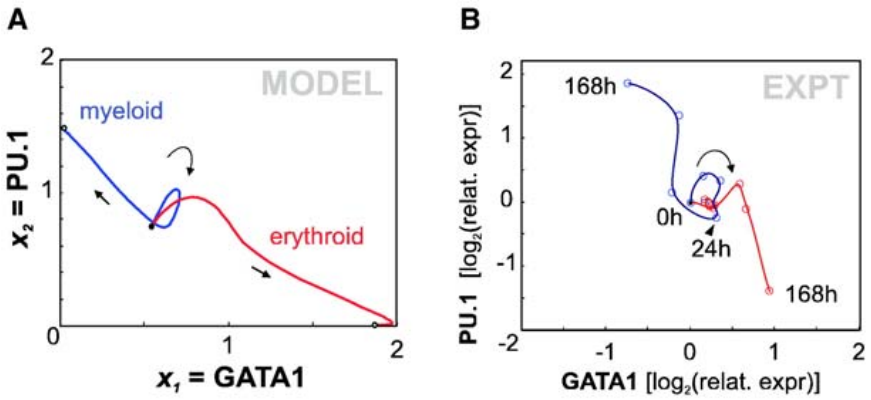


Figure 3.11: Predicted and observed trajectories during binary cell fate decisions. **A**: Typical trajectories for system differentiating into the myeloid fate (blue) or the erythroid fate (red). **B**: Observed trajectories using mRNA levels of GATA1 and PU.1 during differentiation.

The similarity of predicted and observed behavior thus supports the idea that lineage commitment is caused by a (near-) symmetric destabilization of the progenitor attractor. Other possible scenarios seem less likely because they lead to qualitatively different predicted behaviors in the model. In particular, the idea of a directly instructive signal

is rendered implausible since the trajectories corresponding to different lineage fates initially both remain in the central area of the phase plane and only later diverge towards their respective attractor states. The model thus suggests an explanation for the observed stochasticity in cell fate regulation, while at the same time allowing for the possibility that specific signals sometimes 'tip the balance' in favor of one particular fate.

Even though the model provides an attractive mechanistic account of cell fate decisions in terms of the competition between two transcription factors, Huang et al. emphasize that it is not built on the basis of known biochemical interactions. Instead, they think of their model as describing the "functional relationships of nodes in an influence network" (Huang et al. 2007, 698). These relationships are inferred from *in vivo* experiments and, therefore, incorporate the systemic context of the modeled components. The contrast with the 'structure-centered' network motif perspective comes out well in the following quote:

> A network topology motif, such as the 2-gene circuit ... is not necessarily a functionally independent module. The transcription factors GATA1 and PU.1 regulate and are regulated by many other genes and hence, are embedded in an almost genome-wide gene regulatory network ("giant component")..., which establish and maintain the cell-type-specific transcriptomes. (Huang et al. 2007, 704)

On this view, it would be wrong to consider the whole-cell behavior as an epiphenomenon of the activity of a few master regulators. On the contrary, the effective behavior of small circuits is determined by the systemic context.

In order to take the behavior of the whole network into account, the authors went on to study the differentiation trajectories of the transcriptional state of the whole network. They observed roughly the same behavior as that shown by the key regulators GATA1 and PU.1: The trajectories do not immediately diverge into opposite directions of the state space, but show an initial transient phase in which they behave very similarly. Even though the authors stress that it would be highly desirable to have models that describe the process in a higher-dimensional state space, they bring themselves to a conciliatory remark in the end, acknowledging the value of localist strategies:

> Nevertheless, it appears that the dynamic properties of local circuits of 2-3

> genes considered in isolation can have biological relevance. (Huang et al. 2007, 711)

3.3.3 Concluding Remarks on Networks

The theoretical tools of network theory provide potentially powerful heuristics for an understanding of large and complex systems. Representing such a system as a collection of nodes and edges can reveal interesting features that might go unnoticed when working with a more detailed model. At the same time, the network representation can strongly bias the analysis of a system by neglecting important features of the system. The following quote nicely illustrates this point with regards to biological systems:

> [A] graph is a static projection of possible interactions. The analysis of regulatory processes varying in space and time requires additional information not usually included in the topology of biological networks. Indeed, the very representation in the form of a unique network entails the integration in space and time of the interactions taking place during the cellular lifetime. Some of the patterns of interaction might then be spuriously due to a projection effect, whereas they actually take place at different times and/or locations within the cell. (Mazurie et al. 2005, 6)

This indicates that the network perspective should be complemented with other strategies that can reveal potential 'projection effects.' Alternatively, given sufficient data, one might try to circumvent bias by building networks that include many layers of information. A recent study in the context of the ENCODE project has investigated the properties of a 'meta-network' of human transcriptional regulation that includes non-coding RNA regulation, protein-protein interaction, and protein phosphorylation (Gerstein et al. 2012).

Both approaches discussed in this section go beyond the general topological perspective in order to gain understanding of functional aspects of biological networks. Uri Alon's search for network motifs is based on the idea that networks are decomposable into functional units, but he gives up Bechtel and Richardson's strategies of decomposition and localization (discussed in Chapter 2). Instead of functionally decomposing system behav-

ior and subsequently localizing the hypothetical sub-operations in system structure, his starting point is a structural analysis of network topology. The guiding idea is that structural differences in the comparison with randomized networks can reveal clues about biological function. Generic features of large networks serve as a null-model for the detection of biologically meaningful patterns. Sui Huang's attractor perspective is diametrically opposed to the network motif approach since it precisely emphasizes the biological importance of generic features. On his view, the right way to approach the study of biological networks is to focus on simplicity that emerges at the level of the whole system. Such behavior might be explainable without getting into the gory details of the network's structure.

As we have seen, however, both camps deem it necessary to complement their general approaches with more fine-grained analyses. Alon analyzes the dynamical properties of motifs in order to substantiate claims about their functional role, while Huang tries to demonstrate the utility of the attractor view by modeling small networks of transcription factors. This suggests that the use of small mathematical models is the common denominator of many approaches in systems biology. It leaves open the question, however, of a proper way of accounting for systemic context. But shouldn't this be one of the aims of an endeavor that calls itself 'systems biology?'

3.4 Whole Cell Modeling

Up to now we have exclusively dealt with 'partial' models of living organisms. The spindle assembly checkpoint models discussed in Section 3.2 captured only a small aspect of cell cycle regulation, which in turn is only one among many processes occurring inside a cell. The sensory network of *E. coli* and the gene regulatory networks underlying cellular differentiation encompass a much larger number of components, but still account for only some aspects of the overall behavior of an organism. The systems that are modeled in all of these cases are thus treated as functionally independent modules that can be studied in isolation, even though they are clearly embedded in a larger 'super-system' and must be integrated in some way or other to produce the behavior of the whole organism.

William Bechtel and Adele Abrahamsen have recently emphasized the importance of

recomposing as one of the tasks of mechanistic research:

> Reductionist inquiry, which involves decomposing a mechanism into its parts and operations, is only one of the tasks of mechanistic research. A second task (which may be undertaken largely simultaneously) is recomposing it—conceptually reassembling the parts and operations into an organized arrangement that constitutes the mechanism. (Bechtel and Abrahamsen 2009, 177)

In Chapter 2 we have seen that molecular biologists do not simply decompose mechanisms into parts but also elaborate organizational schemes to explain how the parts produce the behavior of interest. However, it seems that the recomposition of different mechanisms into an integrated account of the organism is largely missing from the project of traditional molecular biology. The idea of producing a 'complete model' of a biological organism has considerable allure, and even though we have seen that modeling in systems biology is employed to address a wide variety of specific explanatory tasks, arguably the achievement of completeness would for many systems biologists represent the most impressive way in which their field could fulfill its promises. However, one may ask what the purpose of such a model would actually be. Is there actually an interest in explaining the behavior of a whole organism *in terms of the whole organism*? Would a successful model represent something like a *proof of principle* or *consistency check*, licensing biologists to proceed by showing that all the pieces nicely fit together? Or would it rather serve as a tool that can be interrogated for the generation of new hypotheses and the motivation of interesting experiments? The Japanese systems biologist Masaru Tomita calls whole-cell simulation a "grand challenge for the 21st century," and notes:

> Suppose that a certain organism's genome has been completely sequenced. Then suppose that structures and functions of all its gene products have been thoroughly identified. Suppose further that a giant map of the entire metabolic pathways has been drawn flawlessly. Then what? Would we have conquered the cell? The answer is clearly 'no' because the overall 'behavior' of the cell would still not be understood. (Tomita 2001, 205)

Framing the issue in this way suggests that whole-cell simulations are actually the *only* way to understand the integrated behavior of a cell. Provocative statements like this and

notably Hiroaki Kitano's idea of the 'Human Systeome Project,' whose ambition is "to complete a detailed and comprehensive simulation model of human cells at an estimated error margin of 20% by the year 2020 and to finish the identification of the system profile for all genetic variations, drug responses, and environmental stimuli by the year 2030" (Kitano 2002a), have led some philosophers to question the epistemic value of such efforts. Krohs and Callebaut, for instance, criticize "the project of a 'realistic' representation of all metabolic processes in a 1:1 manner as lacking explanatory power and, more generally, as being epistemologically misguided" (Krohs and Callebaut 2007, 209). More specifically, they argue:

> The systeome project aims to collect data without providing a strategy to arrive at explanatory models. Though coming under the label of systems biology, it turns out to be a purely 'omic' project, as is also made clear in its name. The only improvement with respect to other 'omic' projects is that it integrates a dynamic perspective, but instead of taking explanatory advantage from this perspective as systems biology proper does, the systeomic project degrades network dynamics to another source of large data sets. (Krohs and Callebaut 2007, 207–208)

In order to get a better impression of the scope and value of whole-cell projects, it might be helpful to look at an example of how such projects are actually pursued in practice. This reveals that the ambition of whole-cell models is not so much a 'realistic' representation of a cell, but rather to integrate existing partial models of a cell in order to account for the organization of processes at the organismal level.

In July 2012, a joint group of researchers from Stanford University and from the J. Craig Venter Institute published in the journal *Cell* the first whole-cell model for a complete organism based on detailed and exhaustive empirical information (Karr et al. 2012). A truly astonishing achievement, this model describes the life-cycle of the pathogen *Mycoplasma genitalium* by including all known molecular components and interactions. Incidentally, *M. genitalium* is one of the bacterial species that were used in Venter's spectacular experiments of synthesizing and transplanting entire genomes (Gibson et al. 2010), and it is also the experimental organism for his 'minimal genome project' which seeks to determine the minimal set of genes that can sustain life. Clearly, among the reasons for

choosing this particular organism as a target for the whole-cell model is the fact that *M. genitalium*'s genome, which consists of only 525 genes, is the smallest of any known functioning cell found in nature. Moreover, systematic research efforts have accumulated vast amounts of data about the transcriptome, proteome, and metabolome of bacteria of the genus mycoplasma. Most of the necessary information for building a complete model was therefore already available. One of the biggest challenges, however, consisted in integrating the very different and heterogeneous kinds of datasets. Another integrative and non less important task was the coordinated connection of the various different computational methods that are normally used for the 'small' models of specific processes and conditions.

In what follows I will describe the general modeling strategy and discuss the goals and possible value of such a project. Since the model itself is very large and complex, I will not be able to explain it in full detail, but restrict myself to discussing some broader issues that are of relevance for the questions addressed in this thesis.

3.4.1 Modularity formalized

The key concept underlying Karr et al.'s modeling strategy is modularity. Differently from what some might perhaps expect from a complete model, they did not simply lump all the molecular components together to create one big system of equations. Instead, they constructed the model—very much in the spirit of Simon's watchmaker—by first putting together sub-assemblies, each comprising a substantially smaller number of parts in comparison with the whole system. The decomposition into these sub-assemblies was of course not arbitrary, but based on prior knowledge about the functional processes inside the cell:

> Our approach to developing an integrative whole-cell model was to divide the total functionality of the cell into modules, model each independently of the others, and integrate these submodels together. (Karr et al. 2012, 389–390)

Aside from reducing the complexity of the overall modeling task, the modular approach had the further crucial advantage of facilitating the integration of different methods. As already mentioned, building the whole-cell model required the integration of very dif-

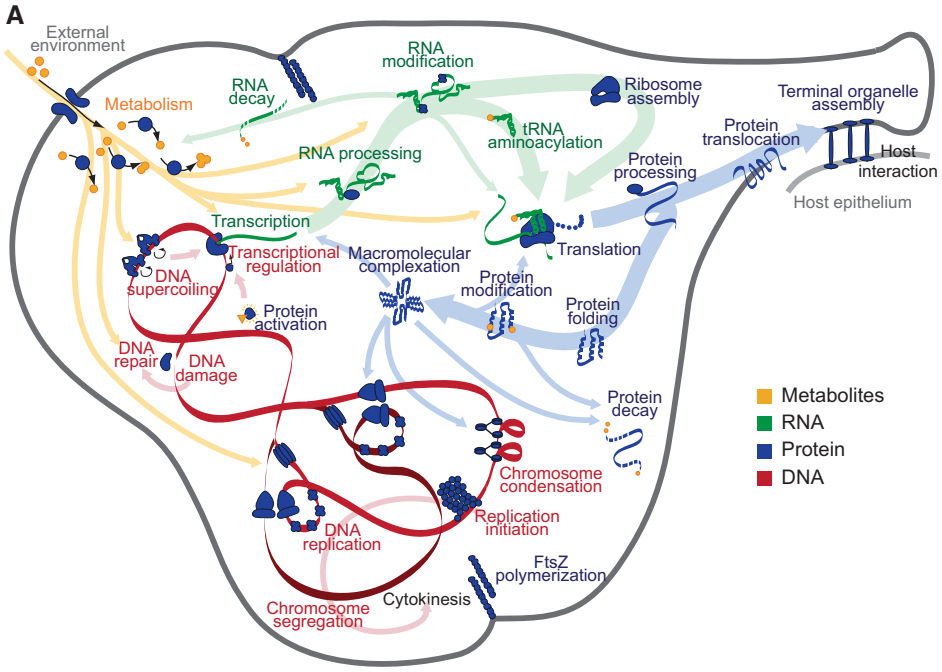


Figure 3.12: Depiction of the 28 submodels and their integration in a *M. genitalium* cell. *Source:* Karr et al. (2012)

ferent styles of mathematical modeling. The reason underlying this necessity lies in the insight that apparently *one size doesn't fit all*, as far as such methods are concerned:

> [N]o single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations ..., were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative approaches were developed that require fewer parameters, including Boolean network modeling ... and constraint-based modeling .... However, the underlying assumptions of these methods do not apply to all cellular processes and conditions, and building a whole-cell model entirely based on either method is therefore impractical. (Karr et al. 2012, 389)

The decomposition allowed the researchers to construct the model by choosing for each module the most adequate style of mathematical representation. However, the real technical problem was the next step: to integrate the modules and to allow them to interact. In order to achieve integration, they drew on the assumption that the processes by which different functional modules interact can be described on a much longer time scale than the processes occurring within each module. The way in which they describe their strat-

egy reads almost as if it was directly inspired by Herbert Simon's reasoning about *near-decomposability*:

> We began with the assumption that the submodels are approximately independent on short timescales (less than 1 s). Simulations are then performed by running through a loop in which the submodels are run independently at each time step but depend on the values of variables determined by the other submodels at the previous time step. (Karr et al. 2012, 390)

The processes communicate with each other by accessing and updating shared *state variables*. These state variables hold the information about the different kinds of entities inside the cell and their configurations. In the supplementary material to the article, the authors compare their method to the numerical algorithms that are used to solve systems of ordinary differential equations. The 28 cellular processes can be considered as 'meta-equations' that are solved independently for each time step, while the 16 state variables figure in different processes, like variables in a set of equations, and therefore represent interfaces between these processes. Of course, the state variables are not simply real numbers as in the case of 'ordinary' ODEs. The *Chromosome* state, for instance, "represents the polymerization, winding, modification, and protein occupancy of each nucleotide of each strand of each copy of the M. genitalium chromosome, and the (de)catenation status of the two sister chromosomes following replication" (Karr et al. 2012, S10). Mathematically speaking, this object is a set of 12 tensors (multi-dimensional arrays of numbers), each of which stores specific information about every nucleotide of the *M. genitalium* genome. Most of the other states, such as the *RNA*, *Metabolite*, or *Polypeptide* states, are of similar complexity.

3.4.2 The Purpose of Whole-Cell Modeling

Karr et al. consulted over 900 primary sources, reviews, and databases in order to gather as much information as possible. More than 1900 observed parameters were incorporated to specify the organization of the *M. genitalium* chromosome, the structure and function of each gene product, metabolite, and their interactions and reactions. The sheer amount of detail should, however, not lead to the impression that the model is in some

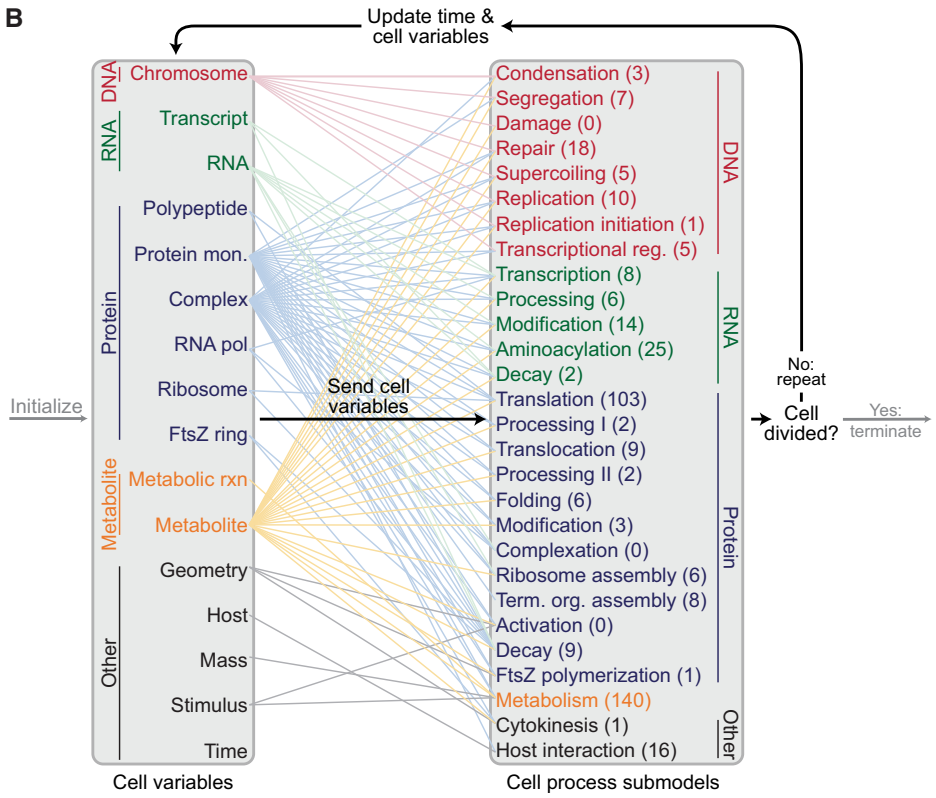


Figure 3.13: Basic flow chart of the whole-cell model. The number in parentheses after each process indicates the number of genes associated with each submodel. *Source:* Karr et al. (2012)

sense complete with respect to molecular detail, and that the behavior is produced as a self-organized product of all the individual molecular interactions. As Harvard systems biologists Jeremy Gunawardena puts it:

> The expectation that, with enough details, a model will miraculously spring to life … is the stuff of fiction. (Gunawardena 2012a, 839)

Instead, the model can only recapitulate biological processes to the extent that they are currently understood. 'Completeness' in the case of this model, therefore, does not refer primarily to the amount of molecular detail that is incorporated, but rather to the fact that all known higher-level processes are taken into account. The way in which the individual modules are represented is not necessarily more advanced than the models we have discussed in the previous sections. The pragmatic spirit of the project is well expressed in the following commentary:

> For every module, there will likely be some expert who will present a fair criticism of the module's mathematical representation or parameter estimation, even though at present they appear to represent the best available attempt

> at balancing realism, computational complexity, and number of free parameters. (Freddolino and Tavazoie 2012, 249)

Clearly, the regulation of the processes that are incorporated in the individual modules does not simply emerge 'from the bottom up.' Moreover, there are many processes that are simply 'black-boxed' or represented in a very coarse-grained way because they are currently not understood well enough. The *Protein Folding* process, for example, represents the three-dimensional configuration of each protein as a two-state Boolean variable: 'folded' or 'unfolded.' The folding rate is a Boolean valued function that increments the copy number of folded protein depending on the amount of unfolded protein, of metabolites, and of chaperones that assist the folding. On the other hand, processes that are better understood, such as chromosome replication, are modeled in considerable detail. Every single process is implemented according to the best available modeling strategy, but all of them heavily rely on simplifications, and there are many remaining gaps in the model. What is more, the virtual cell does not simply start 'living' once all the empirical information is fed in. Despite the available information, the parameter values for many processes are still not known well enough. These parameters, therefore, have to be fit or adjusted in order to fulfill certain basic observational constraints and to be consistent with the other processes. For example, the *Metabolism* process, that describes the import of nutrients and their conversion into building blocks for macromolecules, was fit to match the observed mass doubling time of *M. genitalium* by means of flux balance analysis (FBA).

While this might come as a disappointment to those who are dreaming of virtual organisms that are complete in every respect, the result of Karr et al.'s modeling efforts is far from a trivial aggregation of smaller models. Instead, it is a strategy that has the potential to overcome the biases of decomposition and localization. When we discussed small mechanistic models in Section 3.2, we noticed that biologists mostly investigate mechanisms as individual modules and treat the rest of the organism as simply providing an input for, and receiving an output from, the particular mechanism under study. In this way, the epistemic task is considerably simplified because any complexity that results from the communication *between* modules is ignored. In Karr et al.'s model, by contrast, the inter-module communication is explicitly taken into account. Thus the whole-cell model is not

simply a digital summary of everything that is currently known about a certain small parasite, but it provides a consistency check of the way in which biologists currently grasp the organization of living organisms:

> [W]e anticipate that the construction of whole-cell models and the iterative testing of them against experimental information will enable the scientific community to assess how well we understand integrated cellular systems. (Karr et al. 2012, 399)

An integrated representation of a whole organism imposes additional constraints on the included models of individual processes. The synthesis of enzymes, for instance, consists of several steps each of which involves a number of chemical reactions that require the presence of particular metabolites. These metabolites in turn have to be produced by other processes that require the presence of particular enzymes. The organism as a whole can sustain itself only if all of the different processes occur in a coordinated fashion such that the output of each process matches the demand of those processes that depend on it. This is precisely what a whole-cell model has to account for.

With regard to integrated behavior, the model makes some interesting predictions. The authors noticed, for instance, that the overall length of the cell cycle in the simulation showed considerably less variability than the single stages of the cycle alone. Thus cell cycle length appears to be regulated in some way, even though no regulation was explicitly incorporated in the model. By analyzing the output of their simulations, Karr et al. found that the availability of single DNA nucleotides seems to be responsible for this phenomenon. They observed that the lengths of two stages of the cell cycle, replication initiation and replication, are inversely related to each other. If replication initiation is slow, a large pool of nucleotides builds up in the meantime which in turn speeds up the subsequent replication process:

> The whole-cell model ... presents a hypothesis of an emergent control of cell-cycle duration that is independent of genetic regulation. (Karr et al. 2012, 393)

In a commentary on Karr et al.'s work systems biologist Mark Isalan, emphasizing especially this result, writes:

> So perhaps the most exciting thing about a whole-cell model is that it may allow us to look beyond the direct molecular 'cogs and wheels' that drive biology and into the emergent properties of biological systems. (Isalan 2012, 41)

Note that 'emergent' here is not taken in a strong sense, as something that cannot be explained or predicted on the basis of underlying molecular processes. Instead, what is meant is the fact that the cell cycle control can be understood only when different modules of the system are integrated. This suggests that what systems biologists call 'emergent' are often those behaviors that are left out of the traditional picture of molecular biology due to the biases of decomposition and localization.

Karr et al. further tested their model by comparing its 'phenotypic' behavior against direct experimental observations. The most impressive result in this regard seems at first glance to be the model's ability to predict the essentiality of genes with 79% accuracy. However, this result has to be put into perspective. Among the experimentally tested genes, about 85% turned out to be essential for the bacterium, compared to 71% in the model. If we randomly assigned the genes in the model to the two groups 'essential'/'non-essential,' while keeping group sizes constant, we would obtain an accuracy of 65% by chance![13] This is not to say that Karr et al.'s result isn't highly statistically significant (that is, it cannot be explained by chance alone), but it is maybe not as striking as it might seem at first. Rather than to celebrate this as a big predictive success of the model, it might therefore be more useful to focus on the reasons for deviations between the model and the real system. The authors suggest that such deviations can be exploited for "model-driven biological discovery" (Karr et al. 2012, 396). In particular, they looked more closely at three genes whose disruption resulted in discrepancies between model prediction and observation. In one of the cases this prompted them to consider an additional enzymatic reaction that had not been included in the model before, while the other cases suggested slight parameter changes, consistent with the rest of the model's performance. Overall, they found:

> In each of these three cases . . . , identifying a discrepancy between model pre-

[13] The expected overlap was calculated using the hypergeometric distribution, which is a discrete probability distribution that describes the probability of k successes in n draws from a finite population of size N containing m successes without replacement.

> dictions and experimental measurements led to further analysis, which resolved the discrepancy and also provided insight into *M. genitalium* biology. (Karr et al. 2012, 397)

The real strength of whole-cell modeling might therefore lie in its ability to accelerate biological discovery by including additional constraints that are invisible when looking at individual chunks of a system. Deviations between prediction and model can provide clues on where our knowledge about a system is incorrect or incomplete.

3.5 Conclusion: Alternative Heuristics?

In this chapter I looked at several case studies corresponding to what I take to be different perpectives on the problem of biological complexity. In doing so, I have mainly focused on the contribution of systems biology to discovery. As I have discussed at length in Chapter 1, it is useful to distinguish between the 'epistemic complexity' of a particular scientific task and the 'intrinsic complexity' of the system under study. These two can differ mainly because the researcher's information about the system is usually incomplete. I have considered cases where the scientific task consists in identifying the relevant components and interactions of a mechanism and in explaining how these bring about the behavior of interest. Heuristics are used as tools to simplify this task. I will now summarize how the different approaches discussed in this chapter tackle the problem of complexity and how they differ from the traditional approach of molecular biology.

3.5.1 Small Models: Catalysts of Search

The models discussed in Section 3.2 both target the spindle assembly checkpoint mechanism. Even though molecular biology has acquired considerable knowledge about this mechanism, many open questions remain. The two kinds of models, that I have classified as 'thick' and 'thin,' respectively, can be understood as tools applied at different stages in the discovery process. Both, as I have shown, should be understood as mainly addressing the problem of finding the structure of the mechanism, and only secondarily as tools for understanding its behavior.

The case of thin modeling (Doncic et al. 2005) largely abstracts from molecular detail and focuses on the role of physical constraints. Radical idealizations are introduced, partly to keep calculations tractable, but mainly to sufficiently constrain the model behavior. By evaluating model performance with respect to quantitative empirical constraints, this strategy allows the authors to exclude large classes of possible mechanisms. This strategy involves a clear trade-off: The models incorporate additional information with respect to the mechanistic accounts of molecular biology (reaction kinetics, diffusion, geometry) and thus are in certain respects more sensitive tools to detect deviations between predicted behavior and observation. On the other hand, the idealizations that are introduced might make the candidate models questionable representatives of the target system. For this reason, thin models are usually inadequate to directly *find* the right mechanism. However, their value might lie in raising the standard for candidate explanations by adding additional requirements. For example, every account of the spindle assembly checkpoint mechanism that relies on an autocatalytic loop (as in Doncic et al.'s 'Self-Propagating Inhibition Model') must explain how the system manages to shut off the inhibition fast enough.

Thick modeling is possible when sufficient molecular detail about a mechanism is available, and promises to circumvent some of the problems of the thin approach. Usually, the relevant molecular components have been identified and the basic scheme of the mechanism established, but the precise way in which the components fill in this scheme has not yet been figured out. Discussing Doncic et al. (2009)'s strategy, we saw that the problem of search for the mechanism is reduced to a problem of search through parameter space. Unless sufficiently constrained, this search itself requires heuristic strategies. The idiosyncratic way in which the problem is solved in the example suggests that general strategies to approach the parameter problem in systems biology have yet to be developed. In this regard, Jeremy Gunawardena notes that systems biology will need to start "harmonizing [the] cacophony" of "concepts and techniques that are coming into the subject from the physical sciences and computer science" (Gunawardena 2010, 42).

Both strategies can be seen as complementing the traditional approach and heavily rely on its results. There is thus no question of *replacing* molecular biology in this context. The models retain the general framework of decomposition and localization, but can

afford to drop some of the more specific heuristics of molecular biology (see Chapter 2), owing to a quantitative mode of representation. Doncic et al. (2005), for instance, go beyond purely sequential models by considering the role of an autocatalytic loop. Furthermore, biochemical constraints and population (i.e. concentration) effects are explicitly taken into account. We have seen, however, that in order to be efficient, modelers have to make use of alternative heuristics, mainly in the form of simplifications that allow for analytical or numerical tractability and make the modeling problems well-constrained.

3.5.2 Large Networks: Struggling with Modularity

The challenge posed by large networks is that they do not easily allow for the application of decomposition and localization in the traditional way described in Chapter 2. Large systems often cannot be functionally decomposed in an intuitive way. The examples I discussed in Section 3.3 represent two different perspectives on networks that lead to different strategies of understanding network behavior.

The network motif approach discussed in 3.3.1 relies on the idea of functional modularity, just like the heuristics of decomposition and localization. But instead of starting with system behavior, it applies a structural criterion to identify modules in the network. As discussed at length, this strategy can only work if a number of assumptions on the evolution of networks are justified. In particular, it requires that the freedom of natural selection in 'engineering' functional units is not unlimited:

> Evolution appears to have *converged* on the same network motifs again and again in different systems, suggesting that they are selected because of their function. (Alon 2007, 233, emphasis in original)

Thus, the idea is that a given biological activity constrains the set of possible topologies that can produce it. The way biological evolution produces such networks and the material properties of the components it uses, on this view, give rise to an inherent simplicity of biological networks which may ultimately allow us to understand them:

> There is no *a priori* reason that immensely complex biological systems would be understandable. But despite the fact that biological networks evolved to

> function and not to be comprehensible, simplifying principles can be found that make biological design understandable to us. (Alon 2007, 233)

In spite of its slightly different interpretation of the effects of biological evolution, the network motif approach retains a general informational perspective on biological systems. Even if networks contain feedback and feed-forward loops and thus cannot be forced into the sequential schemes of traditional mechanistic accounts, they are understood as 'information processing' or 'computing' devices. As we have seen, these metaphors suggest intuitive ways in which different functional units can be combined to produce the overall functionality of the network.

The attractor perspective (3.3.2) is in many respects diametrically opposed to this vision of biological organization. Sui Huang stresses precisely those ways in which networks might be more complex than the proponents of a 'localist' vision assume. Since his 'philosophical reflections' address the heuristic aspects of this vision, I will quote him in some detail:

> Much of the topology-based reasoning about function rests on the unarticulated premise that the molecular network acts like a communication network in which some information 'flows' in the links from node to node. Although this may be appropriate for metabolic reaction networks, it certainly does not apply to networks of regulation, like the protein or transcription networks, where a link represents an influence rather than a flow. (Huang 2004, 289)

The relevant processes in a network should not be understood as chains of signaling reactions, but rather as coherent movements of the state of the whole system along a trajectory in the state space. As a result, instances of the same higher-level process can be radically different at the molecular level. This puts into question the traditional approach of understanding a system by accumulating detail about its molecular parts. In a more recent article, Huang writes:

> It is obvious to many biologists that increasing the density of the molecular *fuzzball* by ceaseless discovery of new regulatory relationships … is inapt for providing an intuitive grasp of the observable, emergent stem cell behaviours that are actually quite simple and readily described in few words. The concep-

> tual simplicity of such nested binary choices at the cell behaviour level stands in stark contrast to the vastly complicated molecular network with countless circular control loops which, one naively hopes, may offer linear causal explanations when carefully combed. (Huang 2011, 2247–2248, emphasis added)

This sounds as if the need for an alternative perspective arises not so much from a particular idea of biological organization, but from the limitations of our cognitive abilities:

> An explanation of a phenomenon that exceeds in complexity the phenomenon itself that it seeks to explain will not afford a natural, satisfactory understanding. There is no understanding without simplification .... Thus, we propose that any efforts to achieve satisfactory explanation for how a cell-fate decision ultimately results from the collective action of the molecular interactions must be dedicated to the identification of more abstract, generalizable patterns or principles that are simple enough to be grasped by the human mind notwithstanding the complexity of the impenetrably entangled network of molecular interactions. (Huang 2011, 2248–2249)

According to Huang, the idea of an 'attractor landscape' provides the adequate theoretical framework to understand simplicity at the macrolevel in terms of "well-known 'first principles' of mathematics and physics of dynamical systems" (Huang 2011, 2249). It provides a natural explanation of the robustness of cell types and differentiation processes. However, we have seen that it is not obvious whether a perspective that largely abstracts from molecular details can provide more than a very general idea of these processes. Ultimately, Huang calls for an integration of different perspectives:

> It is of course still necessary to work out the molecular details of the specific pathways that were faulted as not being explanatory in the opening of this paper. Knowledge of the precise molecular pathway diagrams with specific details is still indispensable for designing methods to interfere with cell-fate regulation in order to steer their development into a particular, useful state. If characterization of specific pathway diagrams provides a road map, the study of the state space will one day reveal the topography, exposing the valleys in hidden dimensions and the possibly surmountable hills between them. Such

> information on the structure of the epigenetic landscape will be needed for harnessing the natural forces and constraints that drive cell state changes in order to reprogramme cell fates. (Huang 2011, 2256–2257)

Thus despite appearing as a harsh critic of the traditional approach, Huang does not think that it is without value. What he suggests, though, is that it should be complemented and possibly counterbalanced by a global perspective that keeps an eye on the features that might be lost by decomposing a system.

3.5.3 Whole-Cell Models: The Future?

Both the small mechanism and the large network models retain the problem of neglecting the organismal context in one way or another. I have emphasized the potential of 'complete' models to take into account the interactions between the modules. In this way, whole-cell models promise to escape the biases of decomposition and localization, and at the same time do not rely on the alternative assumptions of the network approaches. So are whole-cell (or whole-organism) models a way to pursue unbiased mechanistic discovery in biology? There are several problems that must be mentioned. The first is practical: It might simply exceed any available computational power in the foreseeable future to scale up from *M. genitalium* to more complex organisms. An obvious next project for whole-cell modelers would be the standard model bacterium *E. coli*. Yet, this step would already correspond to a ten fold increase in genome size. Putting this concern aside, there are also some more principled issues that have to be taken into consideration. Karr et al.'s example shows us where some of the weaknesses of such approaches may hide. One problem is due to the size of the model. Even if investigating a model is usually a much more tractable problem than studying the real system directly, it is not obvious whether one can easily localize the cause of a deviation between prediction and observation in the model. With larger and larger models, one will eventually need heuristics for this problem of search as well. Moreover, there is the general risk that one will always find changes in some of the parameter values to obtain a fit with the empirical data, even if the actual cause of the deviation lies in the structure of the model. In this regard, one might expect that the in-built modularity of the model can greatly facilitate the 'debugging' of the model. Yet, how can we test this assumption of modularity itself? The authors do not pro-

vide any further justification for their particular interpretation of biological modularity and confine themselves in this regard to the following statement:

> Because biological systems are modular, cells can be modeled by the following: (1) dividing cells into functional processes; (2) independently modeling each process on a short timescale; and (3) integrating process submodels at longer timescales. (Karr et al. 2012, 399)

However, even if it is true that biological systems are modular, it is an altogether different question of whether a particular decomposition into modules is correct. To construct their model, Karr et al. had to build on a particular organismal decomposition that was based on the results of previous biological research. For this reason, the model might have inherited some of the biases of the strategies of decomposition and localization. It is difficult to imagine how one could make the model sensitive enough to detect these biases, unless all the molecular properties were known with very high precision.

Whole-cell modeling appears to be a very promising approach to integrate smaller models, to test the consistency of our current knowledge, and to detect some of the biases hidden in approaches that focus on individual mechanisms. However, it has to be taken into account that this approach is not without bias itself.

3.5.4 Alternative Strategies in Systems Biology?

One of the key roles of mathematical modeling in systems biology at present lies in the contribution to the development of mechanistic models of biological behavior. Clearly, we are not yet at a stage where all the parts of biological systems have been identified and the remaining task consists in reproducing the behavior from their interactions. Instead, systems biology continues the project of discovering the causal structures underlying various biological phenomena. The different approaches that I have discussed follow different strategies to tackle the complexity of discovery. These strategies rely on different, and at times competing assumptions about the organization and intrinsic complexity of living systems. With the help of growing amounts of available empirical data and powerful analytical and computational methods it becomes increasingly feasible to directly study the tenability and scope of these assumptions. We have seen examples of this in the con-

text of large network approaches. Importantly, such endeavors often need to integrate an evolutionary perspective on biological systems.

All the discussed examples heavily rely on the previous and ongoing experimental work of molecular biologists, even though they reveal some of the shortcomings of the traditional approach. What all of them have in common is that they relax some of the assumptions implicit in the heuristics of traditional molecular biology. On the other hand, we have seen that all approaches have to introduce strong simplifications in order to arrive at formal and tractable representations of the systems under study. Even though it is probably not possible to make a rigorous claim about this, my case studies suggest that the gain in overcoming the shortcomings of the traditional approach has to be paid by introducing other potential biases. However, my discussion has revealed on several occasions that a combination of different approaches can be productive at reducing bias, provided that the scientists acknowledge the heuristic character of their approaches and allow to be criticized by those who follow alternative strategies. The formulation of mathematical models enables researchers working with different strategies to speak a 'common language,' or at least to find points of contact between their accounts. This possibility is demonstrated in integrative projects such as the whole-cell model. Moreover, a formal mode of representation, whether quantitative or not, forces researchers to make their assumptions explicit which can lead to constructive criticism instead of pointless battles.

4

THE RELEVANCE OF IRRELEVANCE: EXPLANATION IN SYSTEMS BIOLOGY

Summary

After having discussed strategies of developing and revising mechanistic explanations, I now return to the topic of mechanistic explanation itself. In particular, I investigate explanations in systems biology that rely on the tools of dynamic modeling. I argue that accounts of mechanistic explanation that are based on 'change-relating relationships' between the components of a mechanism do not easily make sense of certain features of dynamical patterns that mathematical models can account for. Moreover, I suggest that when investigating the use of such models, one should distinguish between the ideas of 'causal relevance' and 'explanatory relevance.' I show that the explanatory function of mathematical models often consists in elucidating relationships of non-dependence. Notably, the robustness of biological systems is often best accounted for in this way, and not by invoking separate mechanistic features. Drawing on examples from the literature in systems biology, I show that an important aspect of explaining the behavior of a biological mechanism consists in elucidating how in the systemic context components are not, or only weakly, dependent on each other.[1]

[1] An earlier version of this chapter has been accepted for publication (Gross, forthcoming).

4.1 Introduction

The starting point of my analysis in the preceding chapters was that scientists make use of heuristic strategies to facilitate the discovery of mechanisms. I mentioned that heuristics go along with particular assumptions about the organization and complexity of the system under study. Some of these assumptions are openly acknowledged by scientists to be 'working hypotheses,' or even known to be wrong, while others are so entrenched that they are rarely put into question. Moreover, they are often heavily influenced by the social context and can change over time:

> From the universe of the *Timaeus*, through the Archimedian analogues of Galileo and the clockwork universe of Newton, to the recent focus on servomechanisms and computers, the available analogues were important factors in determining which mechanistic models scientists advanced. (Bechtel and Richardson 1993, 17–18)

Systems biology, as we have seen, challenges some of molecular biology's fundamental assumptions about the organization of living systems. In particular, it emphasizes the importance of dynamic features of biological mechanisms. Bechtel and Abrahamsen (2010) recently argued for the importance of such features and proposed an account of 'dynamic mechanistic explanation:'

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism, *manifested in patterns of change over time in properties of its parts and operations*, is responsible for one or more phenomena. (Bechtel and Abrahamsen 2010, 323, emphasis in original)

The patterns of change over time are accounted for by building a quantitative model in which the properties of known components and operations figure as variables and parameters. Such a model can explain phenomena, such as oscillations, that are not easily understood by "mental simulation of the mechanism's functioning" (Bechtel and Abrahamsen 2010, 332). I fully agree with the observation that mathematical models can allow scientists to explain the behavior of very complex mechanisms. Here, I want to draw

attention to a further feature of dynamic modeling. In line with my general focus on the ways in which scientist attempt to reduce complexity, I want to suggest that dynamical models often account for unexpected simplicity.

It is perhaps as a result of the hype around chaos theory that 'nonlinearity' is usually associated with the idea that small changes can have large effects. However, in nonlinear dynamical systems the converse is also possible: large changes with negligible or small effects. In systems biology such cases are widely studied under the label of 'robustness' (e.g. Barkai and Leibler 1997, Carlson and Doyle 2002, Kitano 2004, Daniels et al. 2008, Gunawardena 2010). Molecular biologists, by contrast, usually focus on relationships where a change in one factor leads to a change in another. This goes back to the very intuitive idea of understanding *causation* in analogy with *manipulation*, that is, with the idea of bringing about a change by intervening on some object. This idea has been developed into a philosophical conception of causation and causal explanation by James Woodward (2003). More recently, Carl Craver (2007) elaborated on it in his manipulationist account of mechanistic explanation. The claim I want to defend in this chapter is that such an account relies on a clock-like picture of biological mechanisms by assigning importance solely to *change-relating* relationships. I argue instead for the explanatory value of relationships that are not change-relating,[2] especially when it comes to the explanation of the behaviors of seemingly very complex mechanisms. Examples of explaining biological robustness in systems biology will provide the right kind of illustrations for this point.

The chapter is organized as follows. In the next section I discuss some general intuitions about biological robustness. Next, in section 4.3, I introduce the manipulationist account of mechanistic explanation and discuss how biological robustness might be accounted for from its perspective. In Section 4.4 I turn to examples of explanation in systems biology. The first is a simple model of gene expression. Here my aim is to show that the significance of non change-relating relationships arises especially in the context of dynamic modeling. Dynamic (or steady-state) equilibrium is arguably the simplest case of dynamic stability, and it reveals some important features of the explanations that are given in systems biology. Afterwards, I present a real case study from systems biology that illuminates how information about non-dependence plays an essential role in our under-

[2] I will synonymously speak of "non change-relating relationships" and "relationships of non-dependence".

standing of biological mechanisms. Finally, in section 4.5, these issues are connected to a general perspective on robustness and the global architecture of living systems.

4.2 Biological Robustness

The molecules inside a living cell do not behave like the molecules in a gas. In a gas the individual particles freely move around and interact randomly (if at all), while showing no apparent organization. A gas seems to be the perfect example of a simple aggregate (Wimsatt 1997, 2007b) whose macro-level properties are invariant under many changes at the micro-level. Obviously, the cell is not such a simple aggregate. However, a cell does not appear to behave like a mechanical clock either. The mechanism of a clock almost certainly breaks down if we remove one part or try to exchange two different components, whereas living systems are often surprisingly stable under a wide range of perturbations. An impressive example of such robustness is revealed, for instance, in the experiments of rewiring the *E. coli* network by Isalan et al. (2008) that I discussed in Chapter 3. From such observations one may infer that, with respect to organization, cells assume a position on a spectrum somewhere between gases and clocks, between what Warren Weaver called "disorganized" and "organized complexity" (Weaver 1948, see Chapter 1). Arguably, however, some people would object to this classification. Aren't biological systems much *more* complex and organized than the artifacts of mechanical engineering? And aren't there very specific kinds of perturbations, such as mutations, to which living systems can react in very sensitive ways? Taking this into account, one might conclude that living systems are very complex mechanisms, but differently from clocks they have additional features that account for their particular ways of resisting perturbations.

The coexistence of extreme complexity and robustness is undoubtedly one of the most fascinating features of life. As the physiologist Walter Cannon remarked:

> When we consider the extreme instability of our bodily structure, its readiness for disturbance by the slightest external forces and the rapid onset of its decomposition as soon as favoring circumstances are withdrawn, its persistence through many decades seems almost miraculous. (Cannon 1932, 20)

The scientific work by people like Cannon has shown that the stability of many physiolog-

ical processes can be explained by specific mechanisms. For example, the homeostasis of blood sugar levels can be explained with reference to a simple feedback mechanism that involves the hormones insulin and glucogen. Systems biology has recently started to investigate similar phenomena at a more fine-grained level, such as the robustness of genetic or metabolic networks. Can these phenomena be understood in the same way as the homeostatic processes at the organismal level studied by physiologists? In what follows I want to show how some of the insights gained by systems biologists challenge widely held intuitions about mechanistic explanations in the life sciences. By doing so, I do not want to suggest that systems biology will eventually give rise to an alternative, non-mechanistic paradigm of explanation in biology, but merely draw attention to certain issues that mechanistic accounts will need to address in order to capture the explanatory ambitions of systems biologists.

4.3 Manipulation and Mechanistic Explanation

It seems natural that real understanding of a system implies the ability to predict how it will respond to various kinds of interventions. To understand a phenomenon means to know how changes in it can be brought about, and this idea seems, at least implicitly, to underlie many of the recent conceptions of mechanistic explanation in the philosophy of science. The relationship between intervention and explanation has been made most explicit by James Woodward (2002, 2003) in his manipulationist account of causation and explanation which was subsequently adopted and further developed by some of the main proponents of mechanistic explanation (e.g. Glennan 2002, Craver 2007). On Woodward's account a causal relationship holds between two variables or events if it is possible (at least in principle) to systematically bring about changes in one by intervening on the other. The importance of these relationships for explanation, according to Woodward, lies in the fact that they allow us to answer a range of counterfactual *what-if-things-had-been-different* questions about the explanandum. We may, for instance, explain why a particular person has contracted lung cancer by referring to the fact that the person was a heavy smoker. The causal knowledge that the occurrence of cancer can be influenced by intervening on smoking behavior increases our understanding since it allows us to infer

the counterfactual claim that the person (probably) wouldn't have gotten the cancer if she hadn't smoked.

However, mechanistic explanations are not simply explanations of effects in terms of their causes, but usually are understood as explanations of the properties of a whole in terms of the properties of its parts. This distinction can be further illuminated by invoking the different types of questions that explanations are supposed to answer. While explanations of effects in terms of their causes are directed towards *why-questions*, such as 'why did this person get lung cancer?', the description of the mechanism underlying a behavior may be understood most intuitively as answering a *how-question*, such as 'how does the heart pump blood?'. In this context, Wesley Salmon referred to explanations in terms of underlying structure as *constitutive* and distinguished them from *etiological* explanations that cite the causal history of an event or phenomenon (Salmon 1984, 275). More recently, Craver (2007) has argued that constitutive dependencies between mechanisms and their components are metaphysically distinct from causal dependencies holding between objects at the same level. He argues, however, that usually both causal and constitutive relationships are employed in mechanistic descriptions, and that both can be understood within Woodward's general manipulationist perspective.

4.3.1 Manipulationist and Explanatory Relevance

In Chapter 1 I already discussed the most influential accounts of scientific explanation. In particular, I mentioned nomological and causal conceptions of explanation. According to the deductive-nomological model (e.g. Nagel 1961), explaining consists in logically deriving the explanandum from premises that include law-like generalizations. Among the second category, we find the traditional accounts of causal-mechanical explanation (e.g. Railton 1981, Salmon 1984), according to which an explanation has to show how the explanandum was produced by citing its causal history, where 'cause' is understood, roughly, as physical influence.

The manipulationist conception proposes an alternative to both nomological and causal-mechanical accounts of explanation. It promises to solve some important conceptual problems, notably the question of how to exclude irrelevant factors from an explanation. According to the nomological account, for instance, the following logically valid argument

is an acceptable explanation:

> This sample of table salt dissolves in water, for it has had a dissolving spell cast on it, and all samples of table salt that have had dissolving spells cast on them dissolve in water. (Kyburg 1965, 147)

But clearly, mentioning the spell is not explanatory since *all* samples of table salt—whether or not they have had a spell cast on them—dissolve in water.[3] The traditional causal-mechanical accounts struggle with a slightly different, though no less worrying problem: If everything that has had a causal influence has to be cited in the explanation of an event, where do we stop? Do we, for instance, have to include the gravitational influence of remote stars when explaining a car accident?

The manipulationist conception proposes a solution to these problems of irrelevance by referring to a counterfactual criterion, according to which only those factors are relevant that *could have made a difference* to the explanandum:

> [W]e see whether and how some factor or event is causally or explanatorily relevant to another when we see whether (and if so, how) changes in the former are associated with changes in the latter. (Woodward 2003, 14)

For instance, the spell is irrelevant because by modifying or omitting it we cannot change the dissolving of salt in water. Similarly, manipulating the position of a remote star, apart from being difficult to achieve, is not expected to make a difference to the occurrence of a particular car accident (unless one believes in astrology).

Carl Craver has recently argued that an analogous criterion can provide an account of explanatory relevance in mechanistic explanations of phenomena exhibited by a system. The core idea can be illustrated with the following quote:

> One need not be able to derive the phenomenon from a description of the mechanism. Rather, one needs to know how the phenomenon is situated within the causal structure of the world. That is, one needs to know how the phenomenon changes under a variety of interventions into the parts and how the parts change when one intervenes to change the phenomenon. When one

[3] More generally, the problem arises from the fact that if $P \Rightarrow C$ is valid, then also $P \wedge Q \Rightarrow C$ is valid, where Q can be *any* proposition.

> possesses explanations of this sort, one is in a position to make predictions about how the system will behave under a variety of conditions. Furthermore when one possesses explanations of this sort, one knows how to intervene into the mechanism in order to produce regular changes in the phenomenon. (Craver 2007, 160)

Craver thus thinks of *explanatory relevance* (which components and relationships should figure in a mechanistic explanation) in terms of *manipulationist relevance* (which factors can be manipulated to change the phenomenon).

If one conceives of biological systems as clock-like, it is plausible to equate manipulationist relevance with explanatory relevance. In a clock it seems that exactly those interventions that bring about changes in the overall behavior are the ones that reveal the relationships one needs to know in order to grasp the underlying mechanism. For instance, if the balance spring in a clock is replaced by an otherwise similar spring with greater stiffness, the balance wheel will oscillate with increased frequency, and, as a consequence, the hands of the clock will move faster. Generalizing from this example, one may say that machines like clocks are fragile in a certain sense because changes in the components are connected with systemic behavior in a 'rigid' way. The observation that clocks do not easily fall apart, in spite of this fragility, is explained by the fact that the properties of the parts are not easily changed in the first place. Consider the effect of temperature on a clock. A clock made of metal can work reliably in most climates because the temperatures that could significantly deform its components lie far beyond the typically encountered range. Similarly, most mechanical devices owe their robustness to the fact that properties of their parts are insensitive to a wide range of external perturbations or changes in external conditions.

It seems that biological systems are not fragile in the same sense. Robustness in biological contexts is often taken to mean roughly "that some property of the system remains the same under perturbation" (Gunawardena 2010, 35). In this context, however, 'perturbation' is usually understood as a change in the components or the structure of the system itself. In other words, robustness implies that certain interventions on the components do *not* bring about changes in a phenomenon, which implies that there are relationships between properties of the system and its components that are *not* change-relating. What

role do such relationships play in our attempts to understand biological systems? And how could they be interpreted within a manipulationist account of mechanistic explanation?

There seem to be two strategies of dealing with such relationships within a manipulationist framework. On the one hand, one may argue that they simply fail to meet the criterion for explanatory relevance. For example, a clock's behavior will not be altered by changing the color of the balance spring. Consequently, the color of the spring is considered irrelevant when it comes to explaining how the clock works. However, there might be occasions where the fact that something *doesn't change* itself is of explanatory interest. The manipulationist will then set out to look for an explanation of this behavior in terms of underlying relationships that actually *are* change-relating, such as in the case of blood sugar homeostasis. Insofar as robustness is a somehow "surprising" or "almost miraculous" (Cannon 1932) property of living systems, she will attempt to explain it by looking for specific mechanisms that are responsible for the resistance to change. To sum up, for the manipulationist change-relating relationships are the fundamental building blocks of mechanistic explanations. Relationships that are not change-relating are either irrelevant for our understanding, or themselves have to be explained in terms of change-relating relationships.

By investigating examples of dynamical modeling in systems biology, I will show that relationships that are not change-relating (relationships of non-dependence) point to something deeper and draw our attention to complementary aspects of scientific understanding and explanation that have been neglected in recent discussions on mechanistic explanation. Before turning to these examples, I will discuss in more detail the connection between change-relating relationships and explanation according to the manipulationist picture.

4.3.2 Explanation and Invariance

To illustrate his counterfactual account of causal explanation, James Woodward (2003, 187) makes use of a simple example from physics that probably can be found in any textbook on electrostatics (Figure 4.1). A very long straight wire carries a uniformly distributed electric charge with density λ. The explanandum in this example is the force

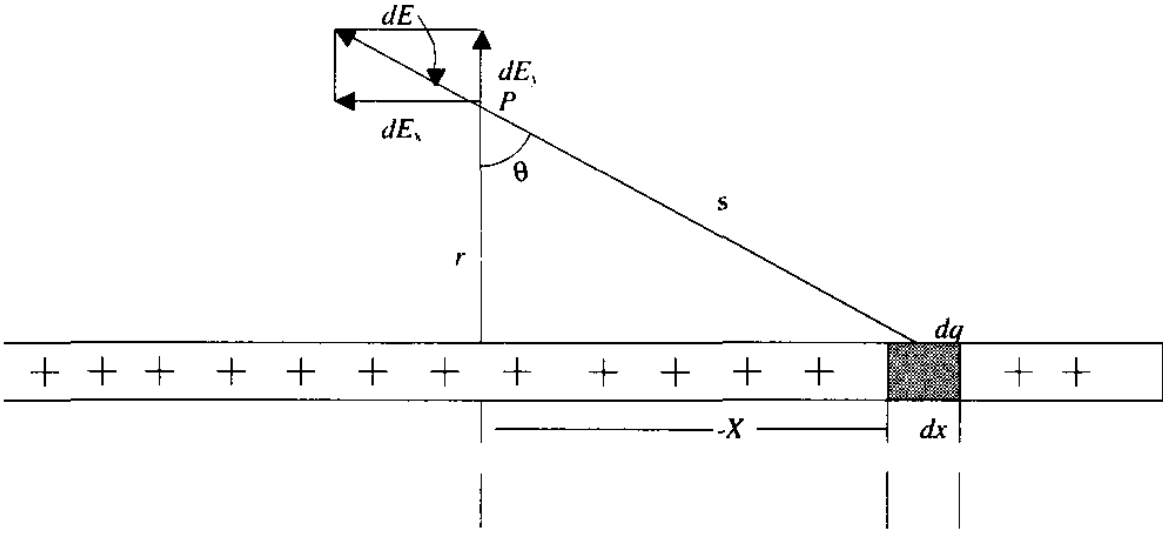


Figure 4.1: Woodward's example of the charged wire. *Source:* Woodward (2003)

of the electric field on a test charge at position P at a perpendicular distance r from the wire. Woodward describes a derivation that is based on Coulomb's law and determines the strength of the field at P by summing up the contributions dq from all the infinitesimal sections dx of the wire. As a result he obtains the following expression:

$$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r}. \tag{4.1}$$

Woodward argues that this relationship, together with its derivation, explains the field intensity at P because it allows us to predict how the value of E changes if we intervene on the system in various ways. For instance, if we increase the distance between the wire and the test charge, the formula tells us that the field intensity decreases proportionally to the reciprocal of the distance. Changing the relative charge λ, on the other hand, results in a proportional change in intensity.

Generalizing from this example, Woodward proposes that explanation amounts to exhibiting the systematic patterns of counterfactual dependence that can be expressed as functional relationships between variables. In doing so, however, he restricts himself to relationships that are change-relating, that is, to those relationships in which an intervention on one variable brings about a change in the other. If we consider the derivation of (4.1), however, we notice that it also elucidates relationships that are not change-relating. For instance, we learn that moving the test charge to a new position P' at the same distance from the wire will not change the value of the field intensity because, as can easily be

shown, such a transformation would not affect the result of the calculation. Likewise, we can infer that the x-component E_x of the field vector will never change to a value different from zero, no matter how we intervene, provided that we do not destroy the symmetry of the geometrical setup.

In Woodward's picture information about such non change-relating relations plays no direct role in explanation. However, the notion of *invariance*, that figures prominently in his account, seems closely related. He argues that causal claims are always associated with claims about invariant relationships:

> Invariance under at least one testing intervention (on variables figuring in the generalization) is necessary and sufficient for a generalization to represent a causal relationship or to figure in explanations. (Woodward 2003, 250)

The idea is roughly the following. The generalizations on which causal claims are based can be described as functional relationships between two variables of the type $Y = G(X)$. It is not necessary that a generalization holds under all circumstances, instead it is required only that there are *some* possible changes of X under which it continues to hold. Invariance obviously comes in degrees, but as long as there is a minimum of invariance, a relationship is causal and, therefore, potentially explanatory. Highly invariant generalizations, such as the fundamental laws of physics, do not necessarily give rise to better explanations, even though they might have other desirable features.

In a more recent article Woodward uses the terms of *invariance* and *stability* interchangeably, and gives a slightly different characterization in terms of background circumstances. He argues that in order to qualify as causal, it is sufficient that a relationship of counterfactual dependence holds in some set of circumstances B_i. He then states:

> The *stability* of this relationship of counterfactual dependence has to do with whether it would continue to hold in a range of other background circumstances B_k different from the circumstances B_i. (Woodward 2010, 291-292, emphasis in original)

According to this characterization, a claim of invariance or stability can be formally expressed as:

$$Y = G(X, B_i) = G(X, B_k) \text{ for all } k \text{ in some set } K, \tag{4.2}$$

which implies the existence of a relationship $F(B) = G(X, B)$ of non-dependence, that is, $F(B_k) = F(B_l)$ even if $B_k \neq B_l$. Woodward's account, therefore, relies on both change-relating and non change-relating relationships! However, the two seem to play very different roles in an explanation. Change-relating relationships, on Woodward's view are the crucial elements; they provide the *content* of the explanation, so to speak, and elucidate the features of the explanandum phenomenon by giving information about what would have changed if things had been different. Relationships of invariance, by contrast, largely keep in the background. They are necessary for specifying the range of application, or generality, of an explanatory claim, but strictly speaking do not provide any explanatory information.

There is thus a clear conceptual separation between the two types of functional relationships reflecting Woodward's distinction between causal explanatory claims, on the one hand, and claims about invariance, on the other hand. However, if we think of the derivation of the field strength in the wire example, we observe that it implicitly also provides information about relationships of non-dependence. If we keep track of both the x- and y-components of the field strength, we notice that all the infinitesimal contributions to E_x exactly cancel out, independently of the position P at which the field is evaluated. The relationship

$$E_x(P) = 0 \text{ for all } P, \tag{4.3}$$

however, does not seem to be irrelevant in the same way as, for instance, the color of the wire. We are therefore inclined to conclude that both change-relating and non change-relating relationships are potentially important for our understanding of mathematical structures like the one given in the example. It is not clear why one type of relationship should be somehow more interesting or informative than the other. The reason why non-change-relating relationships are often neglected might be due to the following feature: information about non-dependence can be represented in a very compressed way—and we have seen that it is often left implicit. As Herbert Simon put it: "Mother Hubbard did not have to check off the list of possible contents to say that her cupboard was bare" (1962, 478). But it is important to see that this property pertains to the way in which we describe a phenomenon, and it should not be conflated with explanatory irrelevance.

Woodward's account implies that change-relating relationships exhaust all that is nee-

ded for explaining the behavior of a system. But if relationships of non-dependence can contribute to our understanding of mathematical models, why shouldn't they be taken as contributing to our understanding of phenomena that are explained by means of such models? As will be further illustrated later in this chapter, the functional relationships that play a role in the models of systems biology often are change-relating in some particular range of values while being non-change relating in a different range. I will show that usually both types of information are crucial for an understanding of complex behavior, without one necessarily being reducible to the other.

By that I do not want to deny the important role that change-relating relationships play in determining the causal or constitutive links within a mechanism. There is no doubt that these relationships provide explanations by allowing us to answer to why-questions of a particular type. But this alone does not entail the equivalence of information about manipulationist relevance and explanatory information when it comes to more complex mechanistic explanations.

A related issue, that Woodward's account leaves unclear, is how invariance or stability itself is explained. As Robert Batterman notices:

> Woodward stresses the importance for explanation of a kind of invariance and robustness that may be present in a given regularity to some degree or other. Thus, he discusses how "nonlaw-like" regularities may, because of their robustness, play crucial explanatory roles. Woodward is not concerned to answer why-questions about the universality or degree of universality of the regularities that he discusses. That is, he does not, as far as I can tell, ask the question why the regularity has the robustness that it has or has it to the degree that it has. (Batterman 2002, 59)

Batterman argues that in the explanation of a phenomenon one has to distinguish between two different kinds of why-questions:

> A type (i) why-question asks for an explanation of why a given instance of a pattern obtained. A type (ii) why-question asks why, in general, patterns of a given type can be expected to obtain. Thus, a request to explain an instance of universality is a request to provide an answer to a type (ii) why-question. (Batterman 2002, 23)

Batterman's ambition to explain universality and Woodward's efforts to elucidate explanation in terms of contingent causal generalizations point to different but possibly complementary aspects of scientific curiosity. These may be seen as loosely related to the different types of questions that are typically asked in the physical and the biological sciences, respectively. It is a philosophically interesting question how the new field of systems biology locates itself on this spectrum since, with regards to its methodological and explanatory resources, it has often been perceived as pushing biology more towards a physics attitude (see e.g. Poon 2011). A closer look at some examples may help to shed light on this issue.

4.4 Explaining Robustness in Systems Biology

4.4.1 Explaining Equilibrium

Let us start with a very simple case and consider the following minimal model of gene expression. The system consists of a protein with concentration X that is synthesized at a constant rate $S = \sigma$, while its degradation rate, $D = \delta \cdot X$, is proportional to the concentration. Figure 4.2 graphically represents the qualitative features of this model. The dynamics of X is captured by the following differential equation:

$$\frac{dX}{dt} = S - D(X) = \sigma - \delta \cdot X\,. \tag{4.4}$$

Solving this equation allows us to obtain the temporal behavior of X depending on a given initial concentration X_0 at time $t = 0$. As can be checked, its explicit solution is given by:

$$X(t, X_0) = \left(X_0 - \frac{\sigma}{\delta}\right) \exp(-\delta t) + \frac{\sigma}{\delta}\,. \tag{4.5}$$

After sufficient time, the value of the exponential will become very small and the first part of the right hand side of (4.5) can be neglected. Formally,

$$X(t, X_0) \to \frac{\sigma}{\delta} \quad \text{for} \quad t \to \infty\,. \tag{4.6}$$

We notice that the expression to which X converges does not contain X_0. This means that the protein concentration in the long run does not depend on its initial value, but assumes an equilibrium (or steady state) value $X_S = \sigma/\delta$ that depends only on the protein's rates of synthesis and degradation. A further consequence is that, whenever the system is perturbed by changing the concentration to some value $X \neq X_S$, it will always return to X_S eventually. At least at first sight this derivation seems to provide a perfectly satisfactory explanation of equilibrium.

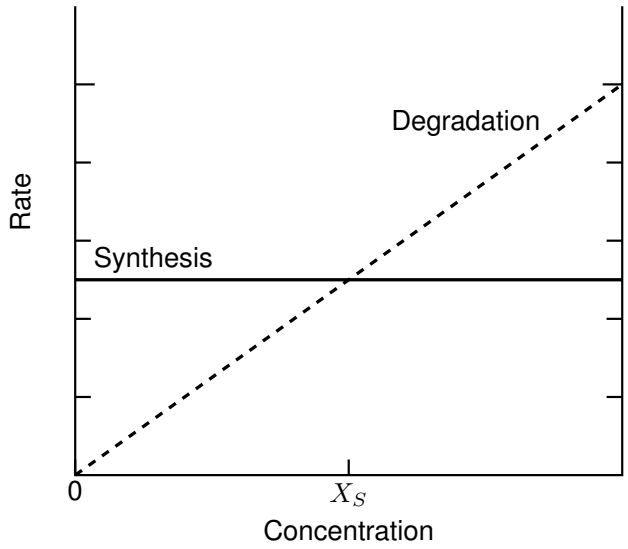


Figure 4.2: Rate balance plot for the simple gene expression model. Synthesis is constant, while degradation depends linearly on protein concentration. Both can be represented as straight lines. The intersection of the two lines corresponds to the equilibrium state. The stability of the equilibrium can be inferred from the sign of the resulting rate when subtracting degradation from synthesis.

The model just described is very similar to an example that Elliott Sober (1983) used to raise some questions about causal-mechanical approaches to explanation. He refers to an explanation given by R. A. Fisher for the 1:1 sex ratio observed in many sexually reproducing species. Instead of providing a particular causal history for the occurrence of the ratio, Fisher points out why the long run ratio in many sexually reproducing populations does *not* depend on particular causal details. As Sober reports:

> Fisher's account shows why the actual initial conditions and the actual selective forces don't matter; whatever the actual initial sex ratio had been, the selection pressures that would have resulted would have moved the population to its equilibrium state. Where causal explanation shows how the event

> to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired. (Sober 1983, 202)

Sober concludes that equilibrium explanations are not causal explanations in the traditional sense:

> The causal explanation focuses exclusively on the actual trajectory of the population; the equilibrium explanation situates that actual trajectory (whatever it may have been) in a more encompassing structure. It is in this way that equilibrium explanations can be more explanatory than causal explanations even though they provide less information about what the actual cause was. This difference arises from the fact that explanations provide understanding, and understanding can be enhanced without providing more details about what the cause was. Equilibrium explanations are made possible by theories that describe the dynamics of systems in certain ways. (Sober 1983, 207)

Sober thus hints at a discrepancy between information about particular causal events and information that is relevant for explanation. He seems to suggest that to explain equilibrium means to show why particular causal facts do not make a difference to the outcome. The question thus arises how this idea relates to Woodward's account according to which such facts are simply explanatorily irrelevant. Is it straightforward to capture equilibrium explanations within the manipulationist framework?

Before trying to determine what kinds of explanations they are—or aren't, we should clarify what it is that equilibrium explanations are supposed to explain. Regarding the sex ratio, the general question is 'Why is there an equilibrium at a sex ratio of 1:1 in so many species?.' However, this can be interpreted as including actually three different calls for explanation, depending on where we put the stress in the sentence. First, it can be read as the question of why it is one and the same ratio that is observed across a wide range of sexually reproducing species. In other words, why does the rate not assume different values for different species? Second, it may express an interest in explaining why the ratio has the particular numerical value of $r = \#\,\text{males}/\#\,\text{females} \approx 1$, and not some other number in the interval $(0,\infty)$. Third, one may ask why the observed ratio represents an equilibrium point, that is, why it is stable and adjusts itself after perturbations.

Each way of interpreting the question calls for an account that makes use of different explanatory resources. The first interpretation, even though interesting in its own right, is not relevant for the current discussion since it seems to mainly depend on empirical facts that are specific to evolutionary biology. For this reason, my focus will be on the differences between the second and the third interpretation that more directly pertain to the phenomenon of equilibrium in general, and roughly correspond to Batterman's type (i) and type (ii) why-questions. I will discuss these differences in more detail using the particularly clear example of the gene expression model.

Let us look at the explanation-seeking question and the two relevant interpretations when transferred to this example. The general question is, 'Why is there an equilibrium at a concentration $X = X_S$?', and it can be interpreted as expressing an interest either in the particular numerical value or in the fact that there is an equilibrium. Responding to the first, the derivation of (4.6) can be taken to show why the protein concentration at steady state is given by the particular ratio σ/δ. This seems to represent a paradigmatic case of a Woodwardian explanation since the steady state concentration is explained in terms of the dependency relations characterizing the system. It clearly allows us to answer a range of counterfactual what-if-things-had-been-different questions. For instance, we can predict how the steady state value would change if we were to intervene on the synthesis or degradation constants. Differently from the type of causal explanation that are the target of Sober's argument, however, this explanation refers to structural features of the model rather than to causal history. In the terminology introduced earlier, this explanation might, therefore, best be understood as constitutive. This is the way in which Kuorikoski (2007) interprets equilibrium explanation within a manipulationist framework:

> If explanations indeed track dependencies instead of persistence, the interesting explanatory relationship cannot be the one between the initial conditions and the equilibrium state, as might first be surmised, and indeed as seems to have been Sober's view. Instead, what the equilibrium state does depend on are the structural features of the system. *Equilibrium explanations are not causal explanations of events but structural or constitutive explanations of system-level properties.* (Kuorikoski 2007, 154, emphasis in original)

However, stating the dependency relations between parameters and steady state value alone arguably does not give an answer to Batterman's type (ii) question of why the pattern, in this case equilibrium, obtains in the first place. Instead, as we have seen, equilibrium seems to be explained precisely by deriving a relation of *non-dependence* between the initial conditions and the long-run concentration. Is there another way in which we can understand this aspect of equilibrium within a manipulationist framework of causation while avoiding Sober's puzzle about the irrelevance of particular causal facts?

To maintain a contrastive focus, one might try to interpret the existence of a single stable equilibrium as a property that systems either do or do not possess, and determine exactly what this property depends on. It turns out that in the present example this property depends only on the structure of the model.[4] This dependency may be expressed in terms of a binary variable $P \in \{0,1\}$ in the following way:

$$P(\{S,D\},\{X\},\{\sigma,\delta\}) = 1, \tag{4.7}$$

where $S = \sigma$ and $D = \delta \cdot X$ represent the particular types of functions used to express the dynamic relationships, while $\{X\}$ and $\{\sigma,\delta\}$ stand for the sets of variables and parameters that appear in the model. By modifying this structure in particular ways, one may obtain a different model for which $P = 0$, that is, a model without an equilibrium state, or perhaps with more than one. An example of such a modification is the complete disruption of degradation, i.e. setting $D = 0$, or the addition of a more complex dependency $S(X)$ of synthesis on the concentration. This reasoning suggests that in principle it might be possible to find a representation of the (potentially very complicated) dependency relation between P and the structural properties of the model. Subsequently, one could make use of this relation to explain why a particular instance of the model does or does not possess the equilibrium property P. Furthermore, one may argue that P's structural dependency explains equilibrium by showing how it appears when the structural parameters are changed in particular ways. But have we thereby really explained equilibrium? It seems that by using the complex dependency relation, we have at best been able to give a more sophisticated answer to a type (i) why-question. That is to say, we have explained

[4] Since the equilibrium is global it does, for instance, not depend on the initial concentration being within a particular range. However, similar arguments can be made for cases of non-global equilibrium.

that a particular system shows equilibrium because it belongs to a particular structural class. If we intervene on the structure of the system in such a way that it no longer belongs to this class, it will exhibit qualitatively different behavior. In the theory of dynamical systems the investigation of equilibrium states when varying the parameter values is known as bifurcation analysis. However, this type of analysis is carried out to investigate the circumstances under which a system shifts between qualitatively different behaviors, not to explain the behaviors themselves. By using a manipulationist strategy, we do not reach beyond the explanation of instances of equilibrium.

To summarize, a satisfactory explanation of equilibrium in causal terms fails for the reasons discussed in Sober's paper. In order to explain equilibrium constitutively, the manipulationist may invoke relationships that relate quantitative or qualitative changes in behavior to changes in structural features of the system, but she thereby fails to give an account of how the behavior is produced in the first place. As I argue, and as Sober suggests, equilibrium is best explained by referring to a relationship of non-dependence.

As mentioned several times, Woodward's manipulationist account implies that only change-relating relationships are doing real explanatory work. Mathematically speaking, this amounts to restricting oneself to *injective* functional relationships that are defined by the property that

$$\text{for} \quad x \neq y \Rightarrow f(x) \neq f(y)\,. \tag{4.8}$$

The mathematical derivation of the equilibrium state, however, makes use of a relationship between possible initial conditions and the long term behavior of the system that is non-injective in the limiting case of $T \to \infty$, that is, for large times T we have that for two initial concentrations X_0 and Y_0

$$X(T, X_0) \approx Y(T, Y_0)\,, \quad \text{even if} \quad X_0 \neq Y_0\,. \tag{4.9}$$

The case of equilibrium shows how knowledge about such relationships can be relevant information for the explanation of a phenomenon.

I briefly discussed in Section 4.3.2 that the manipulationist picture might be defended by maintaining that a relationship of non-dependence simply expresses the fact that some element is explanatorily irrelevant. For instance, even though physics tells us that the cur-

rent positions of remote stars exert a non-vanishing gravitational force on objects on the earth, we do not mention them in our accounts of biological phenomena because we do not think that they make a difference. But I propose that there are interesting 'non-dependencies' just as there are uninteresting dependencies. Many phenomena depend on factors that we would not want to include in their explanations. The croaking of a frog, for instance, depends on whether the frog has just been run over by a car, but we do not cite facts about cars when we explain how a frog croaks. Craver (2007) points out that this problem of 'extravagant causes' follows automatically once one allows for negative causation, and he admits that he does not have a general solution to deal with it. However, extravagant causes seem to be threatening only if one insists on equating the notions of manipulationist and explanatory relevance. If it is true that information about causal irrelevance can be explanatory, then obviously the line between what is relevant for explanation and what is not must be drawn elsewhere. Craver in the end has to resort to a pragmatic notion of changes that *typically* occur in a system. But if one is forced to acknowledge that pragmatic criteria are necessary *anyway* in order to distinguish relevant from irrelevant factors, the objection that information about non-dependence only points to explanatorily irrelevant factors seems much less convincing. Why not accept the pragmatic criteria as primary and consider both relationships of dependence and of non-dependence as potentially providing explanatorily relevant information?

4.4.2 Dissecting a Dynamic Switch

After these initial considerations about dynamical equilibrium, one may ask whether they are of any importance for the description of actual scientific explanations. For this reason I will now turn to a real example taken from the scientific literature. The biological phenomenon I will discuss is an instance of so-called *bistable switching* which plays a role in many important biological processes, for instance in the control of gene expression, in cellular differentiation, cell-cycle progression, and in neural signaling. It is thus representative for a class of phenomena that are biologically relevant and widely discussed among theoretically minded molecular biologists (see e.g. Bhalla and Iyengar 1999, Ferrell and Xiong 2001, Savageau 2001, Novak et al. 2007). My aim is to show how in the explanatory practice of systems biology manipulationist reasoning about causal mechanisms is inte-

grated with dynamical modeling. Notably, it will become clear that relationships of non-dependence are crucial to understand systemic behavior, and not only used to establish the invariance of the causal or constitutive relationships. Conveniently, the philosophically interesting features of this example can be elucidated without going too much into the mathematical details.

At a particular stage during the process of egg formation in the frog *Xenopus laevis*, oocytes are arrested in an immature state. When exposed to the hormone progesterone, they undergo maturation and complete the first meiotic division. The maturation of oocytes has been observed to occur in a switch-like manner, which is to say that cells are either in the immature or in the mature state, but apparently cannot be in intermediate states for extended periods of time (Ferrell and Machleder 1998). A crucial step in triggering maturation is the phosphorylation of the protein kinase p42 MAPK. When treating individual oocytes with intermediate doses of progesterone, Ferrell and Machleder observed either very high (> 90%) or very low levels (< 10%) of phosphorylated p42 MAPK. In the following they were interested in understanding how "a continuously variable stimulus—the progesterone concentration—is converted into an all-or-none biological response" (Ferrell and Machleder 1998, 895). The all-or-none character of the phenomenon subsequently led them to the hypothesis that the underlying process is characterized by bistability, that is, it can be understood as a system shifting between two alternative stable equilibrium states. In what follows I will present the way in which these and other scientists have explained the switching behavior in oocyte maturation.

Bistability can arise in certain types of dynamical systems that involve nonlinear relationships between their variables. It is easy to see how the existence of multiple equilibria is possible when we consider a rate balance plot in which we do not restrict ourselves to straight lines (cf. Figure 4.3). The concentrations of the unphosphorylated and the phosphorylated protein are denoted by A and A^*, respectively, and the total concentration, which is assumed to be constant, by $A_{\text{tot}} = A + A^*$. The particular nonlinear behavior of the forward reaction curve shown in the figure may, for example, be due to the presence of positive feedback. If for small proportions A^*/A_{tot} the slope of this curve is less steep than the slope of the back reaction curve, there can be three intersections of the two curves and hence three equilibrium points. The one in the middle is unstable, however,

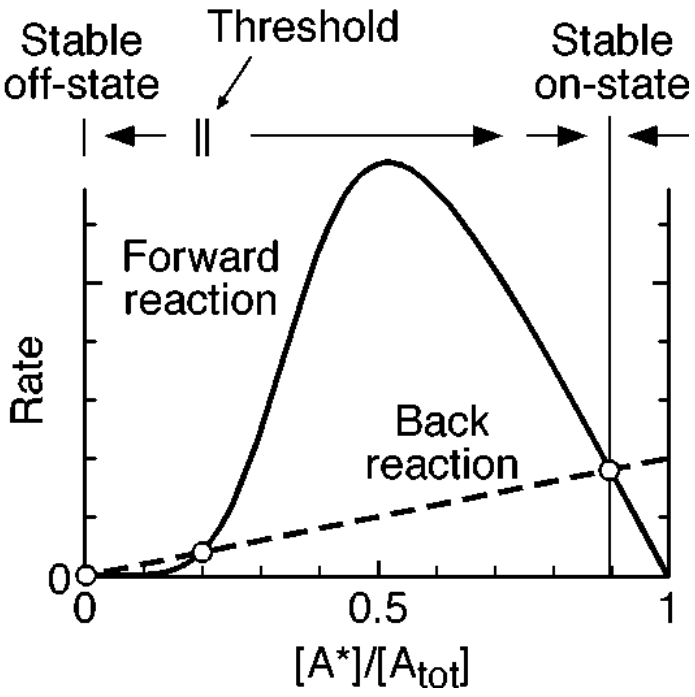


Figure 4.3: Rate balance plot for the oocyte maturation model. Due to non-linearities the forward reaction is not a straight line. The three intersections correspond to three equilibrium points. The one in the middle is unstable. Note that, instead of balance of degradation and synthesis, equilibrium in this case requires equal rates of the forward and the backward phosphorylation reaction. *Source:* Ferrell and Xiong (2001).

since in its vicinity the resulting rate will always drive the system away from it, towards one of the two outer equilibria. Ferrell and Xiong (2001) suggest that in oocyte maturation several mechanisms are probably jointly responsible for the bistability. Notably, p42 MAPK is involved in a positive feedback loop by contributing to the accumulation of Mos, its upstream activating kinase.

In my description I have so far established only that there can be two stable equilibria at low and high concentrations of phosphorylated kinase, respectively. These can be interpreted as *off* and *on* states of a switch; but how can the maturation process be switched on? It turns out that such a shift from *off* to *on* can occur at a critical level of progesterone concentration. This is because the basal rate of the reaction is proportional to the level of the activating progesterone stimulus. The basal rate is the rate at which the reaction would proceed in the absence of the feedback mechanism, and its dependence on the stimulus also affects the shape of the total forward reaction curve. Figure 4.4 illustrates how different levels of progesterone correspond to curves with different shapes. With this representation we can grasp what happens when the level changes: As the stimulus increases, the *off* state and the unstable equilibrium point come closer together until, at

a certain critical level, the two points coalesce. The curves corresponding to even higher levels of stimulus have each only one intersection with the back reaction curve. Therefore, if the system was initially in the *off* state, it will at some level of stimulus jump to the *on* state which then is the only remaining equilibrium. This shows that the all-or-none behavior is due to the existence of a particular threshold level at which one of the equilibria is destabilized.

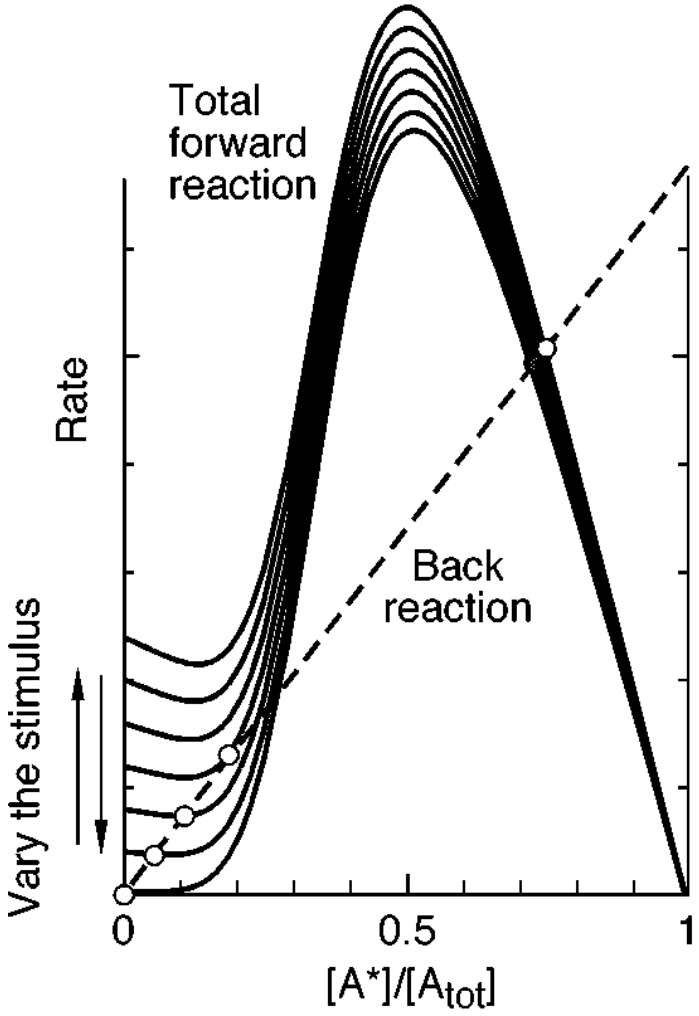


Figure 4.4: Rate balance plot for varying progesterone concentration. Above a critical level of stimulus, the *off* state disappears. *Source:* Ferrell and Xiong (2001).

The behavior of the switch can be further illustrated by representing the position of the stable equilibria as a function of the stimulus (Figure 4.5). This plot elucidates another important property of the switch: After the system has been driven from the *off* state to the *on* state by continuously increasing the stimulus, it will remain in the *on* state even if the stimulus is subsequently decreased again, a behavior known as *hysteresis*. In this way the system is prevented from shifting back and forth between the two states. As a consequence of hysteresis, once the oocyte receives a hormonal stimulus of sufficient size, it is

irreversibly committed to maturation even if the stimulus is later withdrawn. This kind of irreversibility is of course crucial for the reliability of developmental pathways.

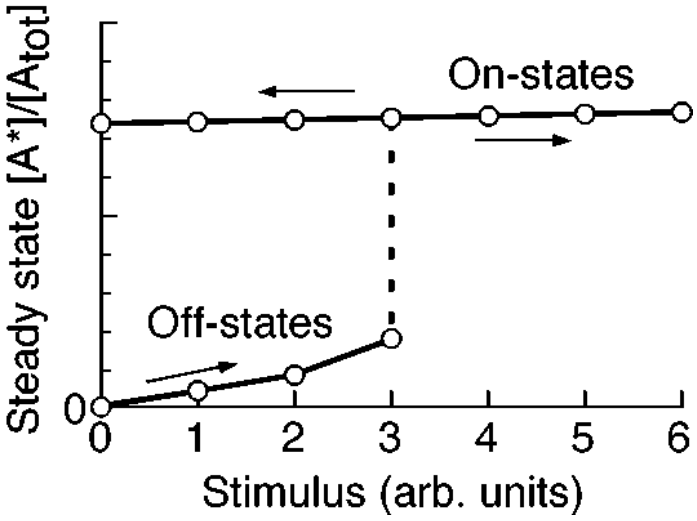


Figure 4.5: Stimulus response curve for the oocyte maturation model. Once the stimulus has reached the threshold level, the system is locked in the *on* state in which the concentration of the phosphorylated kinase A^* is always high. *Source:* Ferrell and Xiong (2001).

Let us now try to understand in more detail how the given account explains the initiation of maturation. If we first consider only the "switching on" part of the story, we can represent the mechanism in terms of a simple causal relationship between two binary variables: A stimulus variable that can take on the values 'below threshold' or 'above threshold', and a kinase activity variable that accordingly assumes either of the values 'on' or 'off'. Obviously, this is exactly what we expect from a simple switch. Note, however, that this behavior is exhibited by a system with a high number of degrees of freedom. The simplicity of the behavior, as will be shown, arises from the fact that possible dependencies among the variables are removed or attenuated.

Let us go back to Figure 4.4 from which we can infer how the total forward reaction rate curve changes as the stimulus is varied. The first thing to notice is that the important changes concern only the lower left portion of the plot.[5] Which are the relevant features for the behavior of the switch? First of all, it is necessary, as we have seen, that there exists a threshold level for the stimulus above which the curves do not intersect in this region of the graph. The value of this threshold is biologically important since it determines the sensitivity of the switch. A very low threshold, for instance, would cause the system to

[5]This is because the stimulus significantly affects the forward reaction only if most of p42 MAPK is still in its unphosphorylated form A.

shift already at small levels of hormone, which might lead to premature differentiation. Second, it is crucial that the threshold level of $A^\star$, that is, the highest value it can reach while still in the *off* state, is not so high as to activate the maturation process. Otherwise, the oocyte would start the maturation process even before the system switches. As long as these conditions are met, however, the details of the relationship between stimulus and *off* state concentration do not matter. The organization of the mechanism, notably the particular type of feedback involved, ensures that there is a range within which $A^\star$ depends only weakly on the stimulus. Note here, that the role of feedback is not merely to confer robustness to particular features of the system. Instead, it is an integral part of the switching mechanism since in its absence the system would not show bistability in the first place.

Similar considerations can be made for the *on* state. Once the system is switched on, the level of $A^\star$ is practically independent from the level of hormone, as can be seen in Figure 4.5. This is once more due to the fact that the stimulus affects the forward reaction only at very low $A^\star$. The hysteresis or memory effect is, therefore, best understood in terms of the *loss* of a dependency relation.

To sum up, what this example shows is that the explanation of complex dynamical behaviors requires information both about relations of dependence and of non-dependence. In order to understand features of persistence, such as robustness or memory, we have to illuminate how some variables in certain ranges do not or do only weakly depend on others. Moreover, this kind of knowledge allows us to explain how systems built of many parts may show behaviors that can be described in comparatively simple terms. The simplicity of the behavior at the level of the whole mechanism is due to the fact that many changes at the level of the components are not constitutively relevant, in the sense of not being change-relating. We cannot fully comprehend how this behavior is brought about if we restrict ourselves to information about manipulationist relevance. This suggests that the manipulationist conception of mechanistic explanation is insufficient to account for many aspects of phenomena that involve dynamical patterns.

As already noted, the description of the switching behavior itself can be taken as representing a change-relating generalization. Therefore, it can be used as a basis for further explanations. For example, one may explain why one particular oocyte did not initiate

maturation by referring to this generalization plus the fact that the given hormonal stimulus was not sufficient. Relationships of non-dependence partly account for the invariance of this generalization. They illuminate, for instance, why different oocytes initiate maturation even when given slightly different doses of stimulus. It might be argued, therefore, that information about change-relating generalizations is sufficient to explain the phenomenon of interest, and that information about non-dependence comes into play only if we want to generalize for further purposes of explanation. But as I hope to have shown, both kinds of relationships are in fact already used in the explanation of the basic features of the switch. The mechanistic explanation that shows *how* the system brings about the behavior contains answers to both of Batterman's types of *why*-questions. Systems biologists want to understand the factors on which changes in observed dynamical patterns depend, but they also want to explain why these patterns are the way they are.

The discussion of this particular mechanism has touched upon the concept of robustness on several occasions. In the following section I will return to the idea of robustness as a fundamental property of living system and show that relationships of non-dependence play an important explanatory role here as well.

4.5 Robustness and the Architecture of Living Systems

Investigating robustness is often invoked as one of the key motivations for research in systems biology. Hiroaki Kitano, for instance, holds that "[it] is one of the fundamental and ubiquitously observed systems-level phenomena that cannot be understood by looking at the individual components" (Kitano 2004, 826). How does this idea of robustness as a fundamental property of living systems connect to the discussion about relations of non-dependence in the preceding sections? We have seen in the example of the switch that the particular dynamical organization of a mechanism can lead to weak relationships between variables or components, which in turn confers reliability and robustness to the system as a whole. However, in the just cited article Kitano notes:

> Robustness is often misunderstood to mean staying unchanged regardless of stimuli or mutations, so that the structure and components of the system, and therefore the mode of operation, is unaffected. In fact, robustness is the

> maintenance of specific functionalities of the system against perturbations, and it often requires the system to change its mode of operation in a flexible way. (Kitano 2004, 827)

This seems to imply that it would be overly simplistic to explain robustness by referring to the causal or the constitutive irrelevance of particular factors under certain conditions. Instead, the quote suggests that the reliable performance of a system requires sophisticated underlying structures. A more refined view of the mechanistic structure of living systems would, therefore, consist in holding that robustness can be explained by invoking particular 'robustness mechanisms.' Indeed, Kitano mentions four different features that could play the role of such mechanisms: system control, redundancy, modularity, and decoupling (Kitano 2004, 827). Even though it may not have been his intention, the fact that Kitano is speaking of "mechanisms that insure the robustness of a system" (Kitano 2004, 827) suggests a very particular biological picture: A living system may at its core be clock-like, but reliable functioning in environments that are characterized by uncertainty and noise is guaranteed by an intricate machinery of additional features that has evolved around this core. If this picture were accurate, the general strategy of understanding mechanisms in terms of change-relating generalizations alone might be justified after all. Robustness would not be a fundamental property of the mechanisms themselves, but rather a separate phenomenon that could be explained by referring to independent mechanistic features. Yet, we have seen in the previous section that there are at least some cases where it is not possible to separate the explanation of a behavior from an explanation of its robustness. Moreover, a closer look at Kitano's alleged robustness mechanisms suggests that this conceptual separation might in general not be obvious. What he means, for example, by 'systems control' is the use of certain control strategies in the building of biological circuits, something that also fits the example discussed in the previous section. One therefore gets the idea that his robustness 'mechanisms' are probably better understood as 'design features' of biological mechanisms. Just as in the case of oocyte maturation, robustness is *in-built* and not in any obvious way *added* to the mechanism.

Redundancy, on the other hand, is often taken as a straightforward mechanism that can explain the observed robustness of large biological networks to perturbations in individual elements. However, the idea that redundancy can explain most of the robustness

of biological networks has recently been challenged. Since this debate sheds some additional light on the distinction I wish to draw in this chapter, I will address it in some detail.

4.5.1 Redundancy and Degeneracy

In systems with redundancy the overall performance can be maintained even if a component is broken or damaged because of the presence of structurally identical or sufficiently similar features that compensate for the loss in function. Redundancy explains, for instance, why people can survive after one of their kidneys has been removed. Similarly, redundancy at the molecular level has often been invoked as an explanation for the robustness of genetic networks. In a large scale perturbation study in yeast cells in which all genes on chromosome V were systematically rendered dysfunctional one by one, Smith et al. (1996) report that almost 40% of their mutants do not show any significant fitness defects. Given that large duplicated chromosomal regions have been found in the yeast genome (Seoighe and Wolfe 1999), this surprising result appears to be most naturally explained by redundancy. Like in the kidney case, one may expect that the effect of knocking out a gene will be mitigated if there is a functional copy of that gene at a different location in the genome. However, as Andreas Wagner (2005) has pointed out, many of the genes whose elimination does not lead to a decrease in system performance are genes without duplicates. He argues that a principal cause of mutational robustness is due to what he calls 'distributed robustness:'

> In distributed robustness, many parts of a system contribute to system function, but all of these parts have different roles. When one part fails or is changed through mutations, the system can compensate for this failure, but not because a "back-up" redundant part takes over the failed part's role. Distributed robustness is a fairly poorly understood cause of mutational robustness, because it requires a detailed, quantitative understanding of the inner workings of a genetic network. (Wagner 2005, 176)

Wagner's main interest are the consequences of mutational robustness for the evolvability of biological systems, and not primarily a mechanistic explanation of this property. He

starts from the idea that a system with a null-effect mutation represents an alternative 'solution' to the problem of matching a particular behavior. He holds that it is a common feature of biological architecture that the same phenotype can be realized by a vast number of different genotypes. The existence of such alternative states, that can be reached by 'neutral' changes, promotes evolvability because it gives a population quick access to large amounts of phenotypic variation. A particular genotype is robust if it lies within a 'neutral network,' a large connected set of alternative genotypes with the same phenotype. A similar view has been expressed by Ralph Greenspan in an article called 'The Flexible Genome:'

> In [a network], the same output can be produced in various ways. This property, particularly when discussed in the context of knockout mutations with no apparent effect, has often been called redundancy. But the compensation that occurs in a network after removal of elements is not redundancy. Redundancy implies substitution of identical elements to preserve the same overall structure, as well as the same outcome. (Greenspan 2001, 385)

Greenspan proposes to talk about *degeneracy* instead, whereby he means "the capacity to produce the same result by different strategies" (Greenspan 2001, 385). This idea connects robustness at the level of larger systems to the discussion about relationships of non-dependence in the previous sections. Degeneracy implies that a property of the system at the level of the observed phenotype is independent from certain changes at the level of the components.

Degeneracy can, however, still be interpreted in different ways. On the one hand, we may take Greenspan's talk of 'different strategies' seriously and hold that a system maintains performance by shifting between different modes of operation depending on the particular type of perturbation it encounters. Consider as an example the capacity of a yeast cell to produce energy using different metabolic pathways depending on whether or not oxygen is present. Systems that are robust in this sense arguably have evolved (or are designed) to respond by adjusting their behavior in specific ways. On the other hand, degeneracy is sometimes understood as an even more fundamental property of complex systems. In line with Wagner's idea of a neutral network, degeneracy may be taken as invariance of performance with respect to a large class of changes in the underlying struc-

ture. In other words, the system might maintain performance, not because it has evolved to cope with particular situations, but because the space of solutions realizing this performance is generically large and connected. An organism would then be protected from a wide range of possible perturbations, even if it has never encountered them before. It is of course nevertheless legitimate to ask whether this architecture is the result, or the by-product, of an evolutionary process, or whether it is "order for free" in Stuart Kauffman's sense (Kauffman 1996), that is, a typical property of a certain class of networks. In the remainder of this section, I will illustrate how mathematical modeling has recently been applied to elucidate the features underlying biological robustness of this kind. There is evidence that the more fundamental idea of degeneracy might play a substantial role in the architecture of living systems.

4.5.2 The Sloppiness of Biological Networks

In the attempt to simulate the interactions among the genes responsible for segmentation in *Drosophila*, von Dassow et al. (2000) developed a dynamical model and systematically investigated its behavior under changes in parameters. The segment polarity network described by this model generates a periodic expression pattern across cells early in development. Initially, von Dassow et al. had hoped that the requirement to reproduce the behavior of the target system would impose sufficient constraints on the model to obtain reasonable estimates for the nearly 50 parameters of the model. Consequently, they expected that only a relatively small subset among all the states in the high-dimensional parameter space would lead to biologically meaningful versions of their model. Strikingly, however, they found that solutions in this space were not rare at all:

> Among 240,000 randomly-chosen parameter sets we found 1,192 solutions (~1 in 200). This is very frequent; as this search involved 48 parameters, on average a random choice of parameter value has roughly a 90% chance of being compatible with the desired behaviour.
> (von Dassow et al. 2000, 189)

Apart from their abundance, solutions are apparently not isolated in parameter space. For many of them the model was found to be tolerant to variation of individual parameters

over several orders of magnitude. Thus the scientists concluded that the model's ability to reproduce the target behavior is "intrinsic to its topology rather than to a specific quantitative tuning" (von Dassow et al. 2000, 189).

The case of the segment polarity network, therefore, supports the idea that robust behavior is not always achieved by adding structural components to an otherwise fragile mechanism. Robustness, therefore, cannot necessarily be analyzed as a separate feature, but instead appears to be entangled with a system's overall functionality. We have seen this clearly in the example of the bistable switch in Section 4.4.2, where the particular feedback, that certainly has an influence on robustness, is necessary for the basic behavior of the mechanism. In general, one gets the idea that the particular organization of many biological systems somehow weakens the dependence of the behavior of the system on the detailed behavior of the components. Note that this does not imply redundancy, according to which some of the system's components are simply dispensable for the behavior. The following quote nicely illustrates how the scientists' initial assumptions about robustness were overturned by their detailed investigation of the mathematical model:

> We originally expected the core topology to be frail and easily perturbed, and expected to achieve robustness only by adding additional complexity; we expected the reconstitution approach to tell us which architectural features confer robustness. Confounding that expectation, the simplest model that works at all emerged complete with unexpected robustness to variation in parameters and initial conditions. (von Dassow et al. 2000, 191)

Robustness of this kind does not seem to be restricted to the generation of developmental patterns in *Drosophila*. Gutenkunst et al. (2007) investigated 17 different systems biology models and systematically examined the sensitivity of their behavior to parameter changes. The set of models covered a wide range of different biological systems and, aside from von Dassow et al.'s network, included models of circadian rhythm, metabolism, and signaling. In all of them they found what they call 'sloppy parameter spectra:' the behavior of the model is sensitive to variation along a few 'stiff' directions in parameter space, but insensitive along a large number of 'sloppy' directions. It is important to emphasize that these directions do not correspond to individual model parameters but rather to combinations of parameters:

> Naively, one might expect the stiff eigenvectors to embody the most important parameters and the sloppy directions to embody parameter correlations that might suggest removable degrees of freedom, simplifying the model. Empirically, we have found that the eigenvectors often tend to involve significant components of many different parameters. (Gutenkunst et al. 2007, 1873)

This means that the systems do not react in a clock-like fashion to most perturbations on individual components. Therefore, in order to bring about significant changes in systemic behavior, it is necessary to intervene on multiple components simultaneously. Obviously, such a feature provides resilience towards many disturbances at the molecular level, but this does not automatically imply that it is an evolved feature of living systems. Daniels et al. (2008), for example, conjecture that sloppiness might be a universal property of a particular class of dynamical models which naturally accounts for many types of robust behavior with no need to invoke separate robustness mechanisms. With regard to von Dassow et al.'s case of the segment polarity network they state:

> The model is robust in these [sloppy] directions not because of evolution and fitness, but because of the mathematical behavior of chemical reaction networks, which are naturally weakly dependent on all but a few combinations of reaction parameters. (Daniels et al. 2008, 393)

In general, however, it is clear that the investigation of biological robustness must pay attention both to evolved robustness mechanisms and to generic features of biological organization such as the one discussed in this section.

All of the models investigated by Gutenkunst et al. rely on mechanisms whose organization is essentially well-known. The principal issue, therefore, is not about whether the proposed components and interactions do in fact bring about the observed behavior of the system. The question is rather how well one needs to know the precise structural features of these components in order to understand the overall working of the mechanism and to be able to make predictions about its behavior. Their results suggest that the relation between components and system behavior is not as straightforward as analogies to machine-like mechanisms would make us believe. Robustness is not just an interesting feature of living systems that requires mechanistic explanation. Instead, thinking about

robustness may have a profound influence on the way in which we should conceive of mechanistic explanations in the life sciences.

4.6 Conclusion

In this chapter I have tried to assess particular accounts of mechanistic explanation, according to which explanatory relevance relies on change-relating generalizations, by looking at dynamical modeling in systems biology. The motivating question was whether this framework can adequately account for what we know about the robustness of living systems, a property that is extensively studied by systems biologists. I have argued that certain aspects of the explanation of dynamical patterns, first and foremost simple dynamical equilibrium, are not captured by approaches that solely focus on change-relating relationships. Instead, the explanation of such features relies on information about relationships of non-dependence, that is, on information about factors or relationships that are irrelevant from a manipulationist standpoint. Next, by presenting the example of a bistable switch in oocyte maturation of *Xenopus*, I have shown that this kind of reasoning is actually applied in mechanistic explanations as they are found in the scientific literature on systems biology. I have tried to illuminate how in the case of this mechanism the discussion of robustness cannot be separated from its functional behavior. I have then turned to a discussion of robustness in larger systems and distinguished between different versions of the concept. There is evidence that at least some of the robustness we find in biological systems cannot be accounted for by invoking separate 'robustness mechanisms.' Instead, it might often be explained by the fact that the interactions among the components are reducible to only a few significantly sensitive dependencies. Therefore, if we want to mechanistically explain how these systems work, we have to understand how the direct dependencies between the parts are weakened due to their organization and how this results in coherent behavior and robustness at the systemic level.

Robustness is often presented as one of the paradigmatic examples of an emergent property; at least systems biologists frequently describe it as such. The ideas discussed in this chapter may shed some light on the reasons for this usage of the term. Philosophical accounts of emergence have mostly focused on system properties that somehow 'exceed'

the capacities of the components, or are unpredictable based on information about individual parts (e.g. Bedau 1997, Kim 1999, Boogerd et al. 2005). However, as the french philosopher Edgar Morin has noticed, a system is not only more than the sum of its parts, it is also *less* than the sum of its parts in certain respects (Morin 2008). The behavior of the components is constrained in various ways by the structure and organization of the system, which keeps them from exhibiting many of the properties that they might show in isolation or in different contexts. Robustness is thus a striking example of the reduction of possibilities. Consequently, what scientists mean by emergence might often simply be the idea that the system is *different* from the sum of its parts. Restricting ourselves to change-relating relationships may prevent us from understanding how such kinds of emergent behavior are brought about.

CONCLUSIONS

Overview of the Results

The main goal of my thesis was to understand, from a philosophical perspective, what systems biology is and how it differs from the traditional approach of molecular biology.

Chapter 1 prepared the ground for my philosophical analysis an introduced some necessary concepts and distinctions. The upshot was that the philosophically relevant differences between molecular biology and systems biology should be investigated by focusing on the discovery and development of mechanistic explanations. Moreover, I argued that it is illuminating to frame this analysis in terms of *heuristics* which I defined as strategies to reduce the epistemic complexity of a given research task.

In Chapter 2 I showed that the traditional approach of molecular biology can be characterized sufficiently well in terms of a particular set of heuristic strategies. Some of these heuristics, such as the strategies of decomposition and localization, are fairly general and not only applied in molecular biology but across a wide range of different scientific fields. Others are more specific and not likely to be found outside the current domain of molecular biology. These more specific heuristics correspond to a particular picture of the organization of living systems at the molecular level, notably on the idea of biological processes as information transmitting sequences. Central to this picture is the assumption, which goes back to Jacques Monod's idea of 'gratuity,' that the organization of biological processes is in a certain sense unconstrained by the underlying principles of biochemistry. This allows molecular biologists to investigate single steps in a mechanism independently

from each other. Another crucial and closely related assumption of traditional molecular biology is that (non-trivial) population effects can be neglected, which is implicit in the habit of molecular biologists to represent the steps in a mechanisms in terms of individual molecules or molecular complexes, even though it is clear that many of these processes involve large populations of molecules. The idea of biological organization that emerges from these assumptions licenses a focus on simple sequential processes that can be described in qualitative terms. As I have emphasized, my ambition was not to provide an exhaustive list of the research strategies utilized in molecular biology. Undoubtedly, one will find even more specific strategies when going to a more fine-grained level of describing scientific discovery, and in the end one will probably have to focus on the individual research group as the right unit of analysis. Moreover, I am aware that by not talking in any detail about the practice of working with experimental systems, I have neglected an entire dimension of scientific activity with its own strategies and related epistemic problems (e.g. Rheinberger 1997b). My primary aim, however, was to find a number of heuristics that are sufficiently general to be found in most of molecular biology, and at the same time relevant for the intended comparison with approaches in systems biology.

Heuristic strategies can create bias if some of the underlying assumptions about the system under study are not met. In Chapter 3 I looked at a number of different approaches in systems biology that all promise to remove some of the biases of the traditional approach by relaxing some of its assumptions. The systems analysis of the spindle assembly checkpoint mechanism revealed that mechanistic models proposed by molecular biologists can turn out to be inadequate once additional physical and biochemical constraints are taken into account. Systems biologists can include these constraints by formulating quantitative models of molecular processes. In general, due to their ability to incorporate more background knowledge and empirical information than traditional accounts, mathematical models can serve as sensitive tools to detect deviations between predicted and observed behavior. This suggests that one of the central roles for mathematical modeling lies in its potential to accelerate the development of mechanistic explanations. We have seen, however, that there are limits to the complexity that can be handled by mathematical models, and their efficient use relies on strong idealizations. It is important to note, therefore, that the achieved gain in analytic power usually goes along with a loss in

empirical adequacy.

Systems approaches to the study of large networks propose alternatives to the more fundamental strategies of decomposition and localization. The approach of network motifs provides a structural criterion for the decomposition of networks, whereas the attractor perspective proposes to resist decomposition and to focus on global behavior. Both of these approaches have heuristic character insofar as they rely on the idea that biological networks are *simple* in some sense. The approach of network motifs, as we have seen, relies on strong assumptions about the evolution of functional modularity in biological networks. The attractor view, on the other hand, presupposes that some properties of biological networks are typical—in the sense that they are generic properties of a statistical ensemble of networks. Both approaches suggest interesting directions for further research, but to the extent that the underlying assumptions are not warranted it might be useful to complement them with other modeling methods.

The last case study in Chapter 3 can be seen as a brute-force approach to relieving some of the potential biases introduced by decomposition and localization. I discussed the impressive work by Karr et al. (2012) who created a model that incorporates all known processes and components of the small microbe *Mycoplasma genitalium*. Even though the model achieves completeness in a certain sense, it is far from a 'realistic' representation of an organism, and its main merit lies in a clever way of integrating a set of smaller models that are formulated with different mathematical techniques. Obviously, the whole-cell model inherits the idealizations and assumptions on which the component models are built, and the strategy of integrating them relies on a strong assumption of modularity. Nevertheless, by taking into account the communication between the mechanistic modules in an organism, this model can act as a tool to test the consistency of existing mechanistic accounts, and it allows scientists to investigate behaviors that emerge at the interface of the modules. The main drawbacks of the whole-cell model are due to its size, which is likely to impede the localization of modeling errors and the assessment of possible adverse effects of the introduced simplifications. Moreover, one should keep in mind that Karr et al. were able to build a manageable model mainly because they chose the simplest organism found in nature—a parasite with only 525 genes. At present, it is not evident whether the whole-cell approach can easily be scaled up to more complex

organisms.

In Chapter 1 I introduced the distinction between 'epistemic complexity' and 'intrinsic complexity.' Epistemic complexity refers to the difficulty of a given scientific task, while intrinsic complexity is a feature of the systems that scientists study. Even though it is useful to distinguish between the two concepts, it should have become clear from my analysis of discovery that scientists often simplify their epistemic task by making assumptions about the intrinsic complexity of the system under study. Herbert Simon (1962) suggested that hierarchical organization and decomposability are the principal features that scientists project onto the systems they are trying to understand. However, it is important to see that biological systems might be simple in very different ways. The risk to underestimate the real complexity of a system is not the only danger when making use of heuristic strategies. It is perhaps equally problematic when heuristics make us overlook unexpected simplicity, and using them incautiously can be even more detrimental when underlying assumptions sneak into our conceptions and standards of scientific explanation. In Chapter 4 I have illustrated this phenomenon by discussing the concept of biological robustness. The focus on 'change-relating relationships,' although perhaps appropriate for the explication of causal relationships, can hide important aspects of the role of quantitative models when it comes to explaining the behavior of biological systems. Dynamical modeling in systems biology suggests that some higher level features that might look as if they require the presence of additional mechanisms can be explained in terms of 'non-change-relating' relationships. As I have illustrated with several concrete examples, mathematical models can often explain why particular lower level details *do not matter*. Systems biology, therefore, has the potential to justify the autonomy of higher level descriptions—not by invoking anti-reductionist arguments, but by explicitly showing how macro-simplicity can emerge from the molecular level.

The Sum of the Parts

Obviously, at the end of this discussion one might still ask 'What is systems biology?' My analysis in terms of heuristic strategies characterized systems biology negatively, by pointing to specific deviations from the traditional approach of molecular biology. My

main goal was to develop a framework that allows for the detection of such deviations and can thereby replace the vague talk of a divide between molecular biology and systems biology with a more precise account. But is there something more fundamental that different systems approaches have in common, apart from all being different from something else? It seems that the diversity of approaches, in spite of systems biology's programmatic calls for integration (e.g. Ideker et al. 2001), renders the 'epistemic landscape' of biology more heterogeneous than before.

Based on my analysis, I want to suggest that the plurality of approaches results from the fact that there are many different ways of reducing epistemic complexity. Perhaps one of the most important merits of systems biology so far has been to draw attention to such alternative ways. I have argued in Chapter 2 that the traditional approach of molecular biology was based on a very specific idea of how we can reduce epistemic complexity. The main components are the 'divide and conquer' approach of decomposition and localization plus an informational vision on how individual molecular processes are organized. This picture is well reflected in a recent article by Sidney Brenner, one of the pioneers of molecular biology, from which I permit myself to quote at some length:

> Any mammalian cell has about 20 000 active genes each producing a polypeptide chain, and we may ask how are we to understand the function of cells through these molecules and their interactions? It is unlikely that we can find a set of differential equations governing these activities and which might allow us to calculate the behaviour of the system. I have always found it advisable when confronted by such questions to analyse how the biological system itself has solved the problem. We first notice that single polypeptide chains hardly ever act alone, but are assembled with others into molecular devices that perform the function. .... If we assume that the average number of components is 10, then such assemblages immediately provide an order of magnitude reduction in complexity and allow us to deal with about 2000 devices instead of 20 000 polypeptide chains. Furthermore, the cell is not a homogeneous solution of molecular entities but is divided into compartments ..., and this provides another order of magnitude reduction in complexity. Thus, in each compartment, on average, we need to focus only on about 200 devices,

> the interactions among them and their communications with other compartments. Several features of this organization should be emphasized: firstly, we can make a distinction between strong interactions which govern the assembly of the devices and weak ones which are involved in the interactions between devices. .... The whole may therefore be pictured as a communication system, with devices transforming and passing information to each other. (Brenner 2010, 209–210)

Even though Brenner is not opposed to the use of mathematical modeling in biology, and in this respect leaves the confines of the qualitative approach of molecular biology, he seems to uphold the idea there is a natural way of exploiting the simplicity of living systems.

Brenner's attempt to find simplicity should be contrasted by the opinion of some systems biologists who suggest that the central aim of their field is to confront biological processes in their full complexity:

> With the present understanding of Life, and of the limitations that biochemical processes have, it is possible to estimate the minimum number of processes required to sustain Life. Living systems function essentially at a non-equilibrium steady state. To maintain this steady state they need to import Gibbs free energy, use some of that to drive thermodynamically uphill processes, and dissipate the rest to speed up the process rates .... An important example of such a process is the breakdown of glucose to alcohol and carbon dioxide by yeast. The solution that evolution has generated is a series of steps in a metabolic pathway that are each catalyzed by a protein. This leads to a requirement of at least 10 proteins. The information needed to specify these proteins must be stored in an information molecule, in practice requiring a nucleic acid of at least 3 kbp. The information has to be translated into protein, which requires a nucleic-acid informed protein-synthesizing enzyme system .... The nucleotides and amino acids out of which these macromolecules consist, need to be made from what is available outside the cell. The corresponding biosynthetic pathways require at least 88 additional enzymes .... It is important that all these components of Life are held together.

> The evolutionary solution for this has been a phospholipid-based membrane, adding a requirement for phospholipid synthesis and transport proteins, requiring another 20 proteins at least. With such an argumentation one readily comes to a minimum requirement for Life of more than 120 proteins, hence more than 120 genes. Genome sequencing has shown that the smallest known genome has some 450 genes . . . , subsequent knock-out experimentation suggesting that the minimum number of genes required for Life is slightly in excess of 375 . . . . All these genes are apparently necessary to maintain each other. . . . . Three hundred and seventy-five is certainly not in the realm of simplicity. . . . . Using Occam's razor we might . . . wish to explain biological formation of ATP in terms of the action of 14 proteins . . . . The above implies that this is impossible, as the 14-enzyme pathway cannot be disentangled from the functioning of 361 other gene products. (Westerhoff et al. 2009, 3884–3885)

The important point is not that Westerhoff et al.'s and Brenner's numerical exercises lead to numbers of the same order of magnitude.[6] Rather, one should appreciate their tendencies towards diametrically opposed methodological recommendations. Westerhoff et al. conclude from their reasoning that the traditional way of reducing epistemic complexity is misleading:

> With respect to Occam's razor, we propose a new paradigm, i.e. that an explanation in terms of fewer than 300 gene products is less likely to be true and complete than an explanation making a provision for the possible influence of more than 300. (Westerhoff et al. 2009, 3887)

Given this statement, it would be natural to expect that the alternative 'paradigm' of systems biology should be one that can do without any reduction of epistemic complexity. However, when considering the authors' own strategy of creating a model of metabolism in the yeast *Saccharomyces cerevisiae*, that is presented in the same article, one cannot help but notice that it is full of heuristics itself:

> To determine what is most important for the organism itself, as well as for the use mankind makes of it, we consider yeast leavening dough and yeast mak-

[6] Note, however, that Brenner talks about mammalian cells, while Westerhoff et al.'s estimate refers to minimal forms of life.

> ing wine. We simplify to an idealized growth medium. Then we anticipate that under these conditions, *S. cerevisiae* only makes use of a small part of its network. To examine which parts of the network it might use theoretically, we implemented standard flux balance analysis for this condition, which indeed led to a greatly reduced number of fluxes in the network. .... And then there is the strategy of modularization, where the hope is that intracellular networks are composed of a number of subnetworks that are heavily networked within themselves but have very few connections between them. To the extent that intracellular networks are indeed scale free ... this strategy seems unlikely to be realistic, but on the other hand the concept of pathways and elementary modes ... suggests that if one would look at dynamic pathways with the fluxes in them, then this simplification through modularization may work. (Westerhoff et al. 2009, 3888)

The main difference with respect to Brenner's account is that Westerhoff et al. do not assume that there is one natural way of simplifying the task of figuring out how living things work. Instead, they seem to draw from a whole toolkit of different heuristic strategies for the *tentative* reduction of epistemic complexity, while being aware that each of these might introduce distortions in the model. At present, systems biology cannot do without reducing epistemic complexity, in spite of the availability of large amounts of data and sophisticated mathematical tools. However, it enriches biology by providing a variety of alternative strategies. In this light, the set of heuristics of traditional molecular biology represents only one among different possible ways of approaching the study of living systems.

At first glance, this perspective on the plurality of systems biology appears to be in the spirit of those philosophers who more generally take a stance against monism and for pluralism in science (e.g. Dupré 1993, Mitchell 2003, Kellert et al. 2006). Usually, these scholars argue for pluralism by referring both to the diversity of human interests and to the complexity of the world (cf. the discussion in Meunier 2011). Here I do not want to take a strong position on science in general, but only briefly address the question of unity in systems biology. First of all, it can certainly not be denied that different biologists have different interests and are working on different problems. However, the diversity of in-

terests should not be overrated, and one should keep in mind that one overarching aim that is frequently expressed by both molecular biologists and systems biologists consists in understanding how living organisms work. I have mentioned in Chapter 1 that the application of a heuristic strategy corresponds to a transformation of the original problem. Thus it may often seem that scientists approaching the same fundamental question with different strategies are in fact working on different problems. A molecular biologist, for instance, might not share the systems biologist's interest in robustness, even though the concept might turn out to help to explain other behaviors that she is interested in. Awareness of the heuristic character of scientific discovery might reveal the relationships between seemingly unrelated problems and create room for dialogue and constructive criticism.

Regarding the issue of complexity, I have based my explanation of the plurality of systems approaches on the idea that there are different strategies of dealing with *epistemic complexity*. This does not exclude that the *intrinsic complexity* of biological systems can eventually be managed by means of a unified approach. Of course, this is impossible to know at present, especially given that biologists, as I have shown, often do not agree about the organization and the level of intrinsic complexity of living systems. Indeed, the goal of many of the heuristics that we have discussed is not only to simplify problems, but also to find patterns of hidden simplicity in reality. The concepts that are fashionable among systems biologists, such as *modularity*, *robustness*, *principles* (of *design*, *organization*, or *optimization*), or even *laws of systems biology* all have this double character of reflecting both intellectual needs and possible features of reality.

Systems biology provides a great opportunity of unifying biological knowledge, by creating formal models that make underlying heuristic assumptions explicit and facilitate the integration of different approaches. Unification seems to be an important ideal, not necessarily as an aim in itself, but definitely as a further tool for the development of adequate mechanistic accounts. With reference to the field of neuroscience, Carl Craver has spoken of a "mosaic unity" of different perspectives (Craver 2007, Chapter 7). The epistemic role of this conception of unity is that it enables scientists to use "constraints from different fields to shrink the space of plausible mechanisms" (Craver 2007, 269). In a similar way, integration in systems biology can move the field beyond the coexistence of

individual approaches since it creates the potential to exploit further constraints and to correct heuristic biases. The whole-cell model of *Mycoplasma genitalium* provides a first glimpse of what the sum of the parts might look like.

ACKNOWLEDGMENTS

First of all, I want to thank all the faculty and colleagues in the FOLSATEC program, as well as the SEMM administration staff. Special thanks goes to my supervisor Mark who took a lot of time discussing and helping me shape my thesis project, even when residing on a different continent.

Moreover, I wish to express my gratitude to Andrea Ciliberto and all the members of his lab for giving me the opportunity to be a small part of the very science that I am discussing in this thesis. In this regard, I also want to thank Luisa Lanfrancone, Giuseppe Testa, and Marco Giorgio for allowing me to spend time in their labs and to perform 'real' experiments at the bench (with mixed success).

I am indebted to Bill Wimsatt not only for agreeing to be my external supervisor, but also for opening up a whole new perspective on the philosophy of biology. He also provided me with the great opportunity to spend several months at the Minnesota Center for the Philosophy of Science. Apart from Bill and his wife Barb, notably Ken Waters, Alan Love, and the grad students of the Center made my visit in Minneapolis a both intellectually rewarding and enjoyable experience.

A big thanks goes to Pierre-Luc for reading the whole of my thesis and for providing most valuable and encouraging feedback. Other people who have provided helpful comments on parts of this work are: Robert Meunier, Jan Baedke, Sara Green, Federico Vaggi, Veli-Pekka Parkkinen, Pierre-Alain Braillard, Christophe Malaterre, and Giovanni Boniolo.

Finally, I thank the University of Milan for providing me with a scholarship that has sustained me for the past four years.

REFERENCES

Adami, Christoph. 2002. What is complexity? *BioEssays* 24 (12): 1085–1094.

Alon, Uri. 2007. *Introduction to systems biology.* Boca Raton: Chapman & Hall/CRC.

Andrews, Peter W. 2002. From teratocarcinomas to embryonic stem cells. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 357 (1420): 405–417.

Artzy-Randrup, Yael, S. J. Fleishman, N. Ben-Tal, and L. Stone. 2004. Comment on "Network motifs: simple building blocks of complex networks" and "Superfamilies of evolved and designed networks". *Science* 305 (5687): 1107.

Aubin, David. 2004. Forms of explanation in the catastrophe theory of René Thom: topology, morphogenesis, and structuralism. In M. N. Wise (Ed.), *Growing Explanations: Historical Perspectives on Recent Science*, pp. 95–130. Durham: Duke University Press.

Badii, Remo and A. Politi. 1997. *Complexity: Hierarchical structures and scaling in physics.* Cambridge: Cambridge University Press.

Barabási, Albert-László and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* 5 (2): 101–113.

Barkai, Naama and S. Leibler. 1997. Robustness in simple biochemical networks. *Nature* 387 (June): 913–917.

Batterman, Robert W. 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence.* Oxford: Oxford University Press.

Bechtel, William. 2006. *Discovering Cell Mechanisms.* Cambridge: Cambridge University Press.

———. 2011. Mechanism and Biological Explanation. *Philosophy of Science* 78 (4): 533–557.

Bechtel, William and A. Abrahamsen. 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421–441.

———. 2009. Decomposing, Recomposing, and Situating Circadian Mechanisms: Three Tasks in Developing Mechanistic Explanations. In A. Hieke and H. Leitgeb (Eds.), *Reduction: Between the Mind and the Brain*, pp. 177–190. Heusenstamm: Ontos.

———. 2010. Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A* 41 (3): 321–333.

Bechtel, William and R. C. Richardson. 1993. *Discovering Complexity*. Princeton: Princeton University Press.

Becskei, Attila and L. Serrano. 2000. Engineering stability in gene networks by autoregulation. *Nature* 405 (June): 590–593.

Bedau, Mark A. 1997. Weak Emergence. *Noûs* 31 (Supplement s11): 375–399.

Bhalla, Upinder S. and R. Iyengar. 1999. Emergent properties of networks of biological signaling pathways. *Science* 283 (January): 381–387.

Black, Max. 1993. More about metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (2nd ed.)., Chapter 2, pp. 19–41. Cambridge: Cambridge University Press.

Boogerd, Fred C., F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff. 2007. Introduction. In F. C. Boogerd, F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff (Eds.), *Systems Biology: Philosophical Foundations*, pp. 3–19. Amsterdam: Elsevier.

Boogerd, Fred C., F. J. Bruggeman, R. C. Richardson, A. Stephan, and H. V. Westerhoff. 2005. Emergence and Its Place in Nature: A Case Study of Biochemical Networks. *Synthese* 145 (1): 131–164.

Braillard, Pierre-Alain. 2008. *Enjeux philosophiques de la biologie des systèmes (Philosophical issues in systems biology)*. Ph.D. Diss., Université Paris 1 - Pantéon Sorbonne.

———. 2010. Systems biology and the mechanistic framework. *History and philosophy of the life sciences 32*: 43–62.

Bray, Dennis. 2003. Molecular networks: The Top-Down View. *Science* 301 (5641): 1864–1865.

Brenner, Sydney. 2010. Sequences and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*: 207–212.

Brigandt, Ingo. 2006. Philosophical issues in experimental biology. *Biology and Philosophy 21*: 423–435.

Brigandt, Ingo and A. C. Love. 2012. Reductionism in Biology. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*. URL = <http://plato.stanford.edu/archives/sum2012/entries/reduction-biology/>.

Brown, Kevin S., C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. 2004. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical biology* 1 (3): 184–195.

Burian, Richard M. 1993. Technique, task definition, and the transition from genetics to molecular genetics: aspects of the work on protein synthesis in the laboratories of J. Monod and P. Zamecnik. *Journal of the History of Biology* 26 (3): 387–407.

Callebaut, Werner. 2005. The Ubiquity of Modularity. In W. Callebaut and D. Rasskin-Gutman (Eds.), *Modularity: Understanding the Development and Evolution of Natural Complex Systems*, pp. 3–28. Cambridge: MIT Press.

Calvert, Jane and J. H. Fujimura. 2011. Calculating life? Duelling discourses in interdisciplinary systems biology. *Studies in History and Philosophy of Biology and Biomedical Sciences* 42 (2): 155–163.

Cannon, Walter B. 1932. *The wisdom of the body*. New York: W. W. Norton.

Carlson, J. Gordon. 1956. On the Mitotic Movements of Chromosomes. *Science* 124 (3214): 203–206.

Carlson, Jean M. and J. Doyle. 2002. Complexity and robustness. *Proceedings of the National Academy of Sciences of the United States of America* 99 (suppl. 1): 2538–2545.

Chapeville, François, F. Lipmann, G. Von Ehrenstein, B. Weisblum, W. J. Ray, Jr., and S. Benzer. 1962. On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences of the United States of America* 48 (1): 1086–1092.

Chen, Rey-Huei, J. C. Waters, E. D. Salmon, and A. W. Murray. 1996. Association of spindle assembly checkpoint component XMAD2 with unattached kinetochores. *Science* 274 (5285): 242–246.

Ciliberto, Andrea and J. V. Shah. 2009. A quantitative systems view of the spindle assembly checkpoint. *The EMBO journal* 28 (15): 2162–2173.

Craver, Carl F. 2007. *Explaining the Brain*. New York: Oxford University Press.

Craver, Carl F. and W. Bechtel. 2006. Top-down Causation Without Top-down Causes. *Biology & Philosophy* 22 (4): 547–563.

Craver, Carl F. and L. Darden. 2001. Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush, and P. McLaughlin (Eds.), *Theory and method in the neurosciences*, Chapter 6, pp. 112–137. Pittsburgh: University of Pittsburgh Press.

Crick, Francis H. 1958. On protein synthesis. *Symposia of the Society for Experimental Biology 12*: 138–163.

———. 1959. Biochemical activities of nucleic acids. The present position of the coding problem. *Brookhaven symposia in biology 12*: 35–39.

Cross, Michael A. and T. Enver. 1997. The lineage commitment of haemopoietic progenitor cells. *Current opinion in genetics & development* 7 (5): 609–613.

Daniels, Bryan C., Y.-J. Chen, J. P. Sethna, R. N. Gutenkunst, and C. R. Myers. 2008. Sloppiness, robustness, and evolvability in systems biology. *Current opinion in biotechnology* 19 (4): 389–95.

Dao Duc, Khanh and D. Holcman. 2012. Using default constraints of the spindle assembly checkpoint to estimate the associated chemical rates. *BMC biophysics* 5 (1): 1.

Darden, Lindley. 2006. *Reasoning in Biological Discoveries*. Cambridge: Cambridge University Press.

De Antoni, Anna, C. G. Pearson, D. Cimini, J. C. Canman, V. Sala, L. Nezi, M. Mapelli, L. Sironi, M. Faretta, E. D. Salmon, and A. Musacchio. 2005. The Mad1/Mad2 Complex as a Template for Mad2 Activation in the Spindle Assembly Checkpoint. *Current Biology* 15 (3): 214–225.

De Backer, Philippe, D. De Waele, and L. Van Speybroeck. 2010. Ins and Outs of Systems Biology vis-à-vis Molecular Biology : Continuation or Clear Cut ? *Acta Biotheoretica 58*: 15–49.

Delehanty, Megan. 2005. Emergent properties and the context objection to reduction. *Biology & Philosophy* 20 (4): 715–734.

Doncic, Andreas, E. Ben-Jacob, and N. Barkai. 2005. Evaluating putative mechanisms of the mitotic spindle checkpoint. *Proceedings of the National Academy of Sciences* 102 (18): 6332–6337.

Doncic, Andreas, E. Ben-Jacob, S. Einav, and N. Barkai. 2009. Reverse engineering of the spindle assembly checkpoint. *PloS one* 4 (8): e6495.

Dupré, John. 1993. *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge: Harvard University Press.

Dyson, Freeman. 2004. A meeting with Enrico Fermi: How one intuitive physicist rescued a team from fruitless research. *Nature* 427 (1): 297.

Fagan, Melinda B. 2011. Waddington redux: models and explanation in stem cell and systems biology. *Biology & Philosophy* 27 (2): 179–213.

Ferrell, James E. and E. M. Machleder. 1998. The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes. *Science 280*: 895–898.

Ferrell, James E. and W. Xiong. 2001. Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. *Chaos* 11 (1): 227–236.

Foster, David V., J. G. Foster, S. Huang, and S. A. Kauffman. 2009. A model of sequential branching in hierarchical cell fate determination. *Journal of theoretical biology* 260 (4): 589–597.

Freddolino, Peter L. and S. Tavazoie. 2012. The dawn of virtual cell biology. *Cell* 150 (2): 248–250.

Freyre-Gonzalez, Julio A. and L. G. Trevino-Quintanilla. 2010. Analyzing regulatory networks in bacteria. *Nature Education* 3 (9): 24.

Friedman, Michael. 1974. Explanation and Scientific Understanding. *The Journal of Philosophy* 71 (1): 5–19.

Gatherer, Derek. 2010. So what do we really mean when we say that systems biology is holistic? *BMC Systems Biology* 4 (22): 1–12.

Gerstein, Mark B., A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* 489 (7414): 91–100.

Gibson, Daniel G, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. a. Hutchison, H. O. Smith, and J. C. Venter. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329 (5987): 52–56.

Gilbert, Scott F. and S. Sarkar. 2000. Embracing Complexity : Organicism for the 21st Century. *Developmental Dynamics* 219 (March): 1–9.

Glennan, Stuart. 1996. Mechanisms and the Nature of Causation. *Erkenntnis* 44 (1): 49–71.

———. 2002. Rethinking Mechanistic Explanation. *Philosophy of Science* 69 (3): S342–S353.

Goldberg, Aaron D., C. D. Allis, and E. Bernstein. 2007. Epigenetics: a landscape takes shape. *Cell* 128 (4): 635–638.

Gould, Stephen J. and R. C. Lewontin. 1979. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 205 (1161): 581–598.

Greenspan, Ralph J. 2001. The flexible genome. *Nature reviews. Genetics* 2 (5): 383–7.

Gross, Fridolin. The Relevance of Irrelevance: Explanation in Systems Biology. In C. Malaterre and P.-A. Braillard (Eds.), *Explanation in Biology*. Springer. (Forthcoming).

Gunawardena, Jeremy. 2010. Models in Systems Biology: The Parameter Problem and the Meanings of Robustness. In H. M. Lodhi and S. H. Muggleton (Eds.), *Elements of Computational Systems Biology*, Chapter 2, pp. 21–47. Hoboken: John Wiley & Sons.

———. 2012a. Silicon dreams of cells into symbols. *Nature biotechnology* 30 (9): 838–840.

———. 2012b. Some lessons about models from Michaelis and Menten. *Molecular biology of the cell* 23 (4): 517–519.

Gutenkunst, Ryan N., J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. 2007. Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology* 3 (10): 1871–78.

Hartwell, Leland H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. From molecular to modular cell biology. *Nature* 402 (Supp.): C47–C52.

Hartwell, Leland H. and T. A. Weinert. 1989. Checkpoints: controls that ensure the order of cell cycle events. *Science* 246 (4930): 629–34.

Hayes, Andrew W. 1992. An Introduction to Chaos and Law. *UMKC Law Review 60*: 751–773.

He, Enuo, O. Kapuy, R. A. Oliveira, F. Uhlmann, J. J. Tyson, and B. Novák. 2011. System-level feedbacks make the anaphase switch irreversible. *Proceedings of the National Academy of Sciences of the United States of America* 108 (24): 10016–10021.

Hempel, Carl G. 1965. Aspects of Scientific Explanation. In C. G. Hempel (Ed.), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pp. 331–496. New York: Macmillan.

Hitchcock, Christopher and E. Sober. 2004. Prediction Versus Accommodation and the Risk of Overfitting. *British Journal for the Philosophy of Science* 55 (1): 1–34.

Hodgkin, Alan L. and A. F. Huxley. 1952. A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. *Journal of Physiology 117*: 500–544.

Hooker, Cliff. 2011. Introduction to Philosophy of Complex Systems: A. In C. Hooker (Ed.), *Philosophy of Complex Systems*, pp. 3–90. Amsterdam: Elsevier.

Howell, Bonnie J., D. B. Hoffman, G. Fang, A. W. Murray, and E. D. Salmon. 2000. Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. *The Journal of Cell Biology* 150 (6): 1233–1250.

Hoyt, M. Andrew, L. Totis, and B. T. Roberts. 1991. S. cerevisiae genes required for cell cycle arrest in response to loss of microtubule function. *Cell* 66 (3): 507–517.

Huang, Sui. 2000. The practical problems of post-genomic biology. *Nature biotechnology* 18 (May): 471–472.

———. 2004. Back to the biology in systems biology: what can we learn from biomolecular networks? *Briefings in functional genomics & proteomics* 2 (4): 279–97.

———. 2011. Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences 366*: 2247–2259.

Huang, Sui, G. Eichler, Y. Bar-Yam, and D. E. Ingber. 2005. Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Physical Review Letters* 94 (12): 1–4.

Huang, Sui, I. Ernberg, and S. A. Kauffman. 2009. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in cell & developmental biology* 20 (7): 869–876.

Huang, Sui, Y.-P. Guo, G. May, and T. Enver. 2007. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology* 305 (2): 695–713.

Huerta, Araceli M., H. Salgado, D. Thieffry, and J. Collado-Vides. 1998. RegulonDB: a database on transcriptional regulation in Escherichia coli. *Nucleic acids research* 26 (1): 55–59.

Hwang, Lena H., L. F. Lau, D. L. Smith, C. A. Mistrot, K. G. Hardwick, E. S. Hwang, A. Amon, and A. W. Murray. 1998. Budding Yeast Cdc20: A Target of the Spindle Checkpoint. *Science* 279 (5353): 1041–1044.

Ibrahim, Bashar, S. Diekmann, E. Schmitt, and P. Dittrich. 2008. In-silico modeling of the mitotic spindle assembly checkpoint. *PloS one* 3 (2): e1555.

Ideker, Trey, T. Galitski, and L. Hood. 2001. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics 2*: 343–372.

Isalan, Mark. 2012. A cell in a computer. *Nature* 488 (August): 9–10.

Isalan, Mark, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, and L. Serrano. 2008. Evolvability and hierarchy in rewired bacterial gene networks. *Nature* 452 (April): 840–846.

Jacob, François and J. Monod. 1963. Genetic repression, Allosteric inhibition, and Cellular Differentiation. In M. Locke (Ed.), *Cytodifferentiation and Macromolecular Synthesis*, pp. 30–64. New York: Academic Press.

Jeong, Hawoong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. 2000. The large-scale organization of metabolic networks. *Nature* 407 (6804): 651–654.

Karess, Roger. 2005. Rod-Zw10-Zwilch: a key player in the spindle checkpoint. *Trends in cell biology* 15 (7): 386–392.

Karr, Jonathan R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. 2012. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* 150 (2): 389–401.

Kauffman, Stuart A. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology* 22 (3): 437–67.

———. 1970. Articulation of Parts Explanation in Biology and the Rational Search for Them. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1970, pp. 257–272.

———. 1974. The Large Scale Structure and Dynamics of Gene Control Circuits: An Ensemble Approach. *Journal of theoretical biology* 44 (1): 167–190.

———. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press.

———. 1996. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. New York: Oxford University Press.

———. 2004. A proposal for using the ensemble approach to understand genetic regulatory networks. *Journal of theoretical biology* 230 (4): 581–590.

Kay, Lily E. 1993. *The Molecular Vision of Life: Caltech, the Rockefeller Foundation, and the Rise of the New Biology*. New York: Oxford University Press.

———. 2000. *Who wrote the book of life? A history of the genetic code*. Stanford: Stanford University Press.

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

———. 2005. Revisiting "scale-free" networks. *BioEssays* 27 (10): 1060–1068.

Kellert, Stephen H. 2008. *Borrowed Knowledge: Chaos Theory and the Challenge of Learning across Disciplines*. Chicago: University of Chicago Press.

Kellert, Stephen H., H. E. Longino, and C. K. Waters. 2006. Introduction: The Pluralist Stance. In S. H. Kellert, H. E. Longino, and C. K. Waters (Eds.), *Scientific Pluralism*, pp. vii–xxix. Minneapolis: University of Minnesota Press.

Kim, Jaegwon. 1999. Making sense of emergence. *Philosophical Studies 95*: 3–36.

Kitano, Hiroaki. 2002a. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Current genetics* 41 (1): 1–10.

———. 2002b. Systems biology: a brief overview. *Science* 295 (5560): 1662–1664.

———. 2004. Biological robustness. *Nature Reviews Genetics* 5 (11): 826–37.

Kitcher, Philip. 1981. Explanatory Unification. *Philosophy of Science* 48 (4): 507–531.

Knabe, Johannes F., C. L. Nehaniv, and M. J. Schilstra. 2008. Do motifs reflect evolved function?—No convergent evolution of genetic regulatory network subgraph topologies. *BioSystems* 94 (1–2): 68–74.

Kolmogorov, Andrey N. 1998. On tables of random numbers. *Theoretical Computer Science* 207 (2): 387–395. Reprinted from 1963. *Sankhyā: The Indian Journal of Statistics, Series A* 25 (4).

Krohs, Ulrich. 2009. The Cost of Modularity. In U. Krohs and P. Kroes (Eds.), *Functions in Biological and Artificial Worlds: Comparative Philosophical Perspectives*, Volume 276, Chapter 15, pp. 259–276. Cambridge: MIT Press.

Krohs, Ulrich and W. Callebaut. 2007. Data without models merging with models without data. In F. C. Boogerd, F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff (Eds.), *Systems Biology: Philosophical Foundations*, pp. 181–213. Amsterdam: Elsevier.

Kuhn, Thomas S. 1963. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuo, P. Dwight, W. Banzhaf, and A. Leier. 2006. Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *BioSystems* 85 (3): 177–200.

Kuorikoski, Jaakko. 2007. Explaining with equilibria. In J. Persson and P. Ylikoski (Eds.), *Rethinking Explanation*, pp. 149–162. Dordrecht: Springer.

Kyburg, Henry E. 1965. Salmon's Paper. *Philosophy of Science* 32 (2): 147–151.

Langley, Pat, H. A. Simon, G. L. Bradshaw, and J. M. Zytkow. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge: MIT Press.

Lara-Gonzalez, Pablo, F. G. Westhorpe, and S. S. Taylor. 2012. The Spindle Assembly Checkpoint. *Current Biology* 22 (22): R966–R980.

Levesque, Mitchell P. and P. N. Benfey. 2004. Quick guide Systems biology. *Current Biology* 14 (5): R179–R180.

Levins, Richard. 1966. The Strategy of Model Building in Population Biology. *American Scientist* 54 (4): 421–431.

Lew, Daniel J. and D. J. Burke. 2003. The spindle assembly and spindle position checkpoints. *Annual review of genetics 37*: 251–82.

Li, Rong and A. W. Murray. 1991. Feedback control of mitosis in budding yeast. *Cell* 66 (3): 519–531.

Li, Yong and R. Benezra. 1996. Identification of a human mitotic checkpoint gene: hsMAD2. *Science* 274 (5285): 246.

Ma, Wenzhe, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. 2009. Defining network topologies that can achieve biochemical adaptation. *Cell* 138 (4): 760–773.

Machamer, Peter, L. Darden, and C. F. Craver. 2000. Thinking About Mechanisms. *Philosophy of Science* 67 (1): 1–25.

Mazurie, Aurélien, S. Bottani, and M. Vergassola. 2005. An evolutionary and functional assessment of regulatory network motifs. *Genome biology* 6 (4): R35.1–12.

McIntosh, J. Richard. 1991. Structural and Mechanical Control of Mitotic Progression. *Cold Spring Harbor Symposia on Quantitative Biology 56*: 613–619.

McMullin, Ernan. 1985. Galilean idealization. *Studies in History and Philosophy of Science Part A* 16 (3): 247–273.

Meunier, Robert. 2011. *Thick and Thin Characters: Organismal Form and Representational Practice in Embryology and Genetics*. Ph.D. Diss., University of Milan.

Milo, Ron, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network motifs: simple building blocks of complex networks. *Science* 298 (5594): 824–827.

Minton, Allen P. 2006. How can biochemical reactions within cells differ from those in test tubes? *Journal of cell science* 119 (14): 2863–2869.

Mistry, Hitesh B., D. E. MacCallum, R. C. Jackson, M. A. J. Chaplain, and F. A. Davidson. 2008. Modeling the temporal evolution of the spindle assembly checkpoint and role of Aurora B kinase. *Proceedings of the National Academy of Sciences of the United States of America* 105 (51): 20215–20220.

Mitchell, Sandra. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Mogilner, Alex, R. Wollman, G. Civelekoglu-Scholey, and J. Scholey. 2006. Modeling mitosis. *Trends in cell biology* 16 (2): 88–96.

Monod, Jacques. 1971. *Chance and necessity: An essay on the natural philosophy of modern biology*. New York: Knopf.

Morange, Michel. 1998. *A History of Molecular Biology*. Cambridge: Harvard University Press.

Morgan, David O. 2007. *The Cell Cycle: Principles of Control*. London: New Science Press.

Morin, Edgar. 2008. *On Complexity*. New York: Hampton Press.

Murray, Andrew W. and M. W. Kirschner. 1989. Cyclin synthesis drives the early embryonic cell cycle. *Nature 339*: 275–280.

Musacchio, Andrea. 2011. Spindle assembly checkpoint: the third decade. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 366 (1584): 3595–3604.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. London: Routledge and Kegan Paul.

Noble, Denis. 2006. *The Music of Life*. Oxford: Oxford University Press.

Novak, Bela, J. J. Tyson, B. Gyorffy, and A. Csikasz-Nagy. 2007. Irreversible cell-cycle transitions are due to systems-level feedback. *Nature cell biology* 9 (7): 724–728.

Olde Scheper, Tjeerd. 2008. Why metabolic systems are rarely chaotic. *Bio Systems* 94 (1-2): 145–152.

O'Malley, Maureen A. and J. Dupré. 2005. Fundamental issues in systems biology. *BioEssays* 27 (12): 1270–1276.

O'Malley, Maureen A. and O. S. Soyer. 2012. The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biology & Biomedical Sciences* 43 (1): 58–68.

Orkin, Stuart H. 2000. Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics* 1 (1): 57–64.

Pólya, George. 1941. Heuristic Reasoning and the Theory of Probability. *The American Mathematical Monthly* 48 (7): 450–465.

Poon, Wilson C. K. 2011. Interdisciplinary reflections: the case of physics and biology. *Studies in history and philosophy of biological and biomedical sciences* 42 (2): 115–118.

Popper, Karl. 1959. *The Logic of Discovery*. London: Hutchinson.

Powell, Alexander and J. Dupré. 2009. From molecules to systems: the importance of looking both ways. *Studies in history and philosophy of biological and biomedical sciences* 40 (1): 54–64.

Psillos, Stathis. 2002. *Causation and explanation*. Stocksfield: Acumen.

Railton, Peter. 1981. Probability, Explanation, and information. *Synthese* 48 (2): 233–256.

Resnik, David. 1997. Adaptationism: Hypothesis or Heuristic? *Biology & Philosophy 12*: 39–50.

Rheinberger, Hans-Jörg. 1997a. Experimental Complexity in Biology: Some Epistemological and Historical Remarks. *Philosophy of Science* 64 (Supplement): S245–S254.

———. 1997b. *Toward a history of epistemic things: Synthesizing proteins in the test tube*. Stanford: Stanford University Press.

———. 2007. What happened to molecular biology? *B.I.F. Futura 22*: 218–223.

Rieder, Conly L., R. W. Cole, A. Khodjakov, and G. Sluder. 1995. The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *The Journal of Cell Biology* 130 (4): 941–948.

Rosenberg, Alex. 2006. *Darwinian Reductionism: Or, How to Stop Worrying and Love Molecular Biology*. Chicago: University of Chicago Press.

Rosenfeld, Nitzan, M. B. Elowitz, and U. Alon. 2002. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology* 323 (5): 785–793.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

———. 1989. *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press.

Sansom, Roger. 2008. Countering Kauffman with Connectionism: Two Views of Gene Regulation and the Fundamental Nature of Ontogeny. *The British Journal for the Philosophy of Science* 59 (2): 169–200.

Sarkar, Sahotra. 1992. Models of Reduction and Categories of Reductionism. *Synthese 91*: 167–194.

———. 1996. Philosophy, History, and Molecular Biology: Introduction. In S. Sarkar (Ed.), *The Philosophy and History of Molecular Biology: New Perspectives*, pp. 1–14. Dordrecht: Kluwer Academic Publishers.

———. 2005. *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge: MIT Press.

Savageau, Michael A. 1976. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Reading: Addison-Wesley.

———. 2001. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* 11 (1): 142–159.

Schaffner, Kenneth F. 1969. The Watson-Crick Model and Reductionism. *British Journal for the Philosophy of Science* 20 (4): 325–348.

———. 1993. *Discovery and Explanation in Biology and Medicine*. Chicago: The University of Chicago Press.

Sear, Richard P. and M. Howard. 2006. Modeling dual pathways for the metazoan spindle assembly checkpoint. *Proceedings of the National Academy of Sciences of the United States of America* 103 (45): 16758–16763.

Seoighe, C. and K. H. Wolfe. 1999. Yeast genome evolution in the post-genome era. *Current opinion in microbiology* 2 (5): 548–54.

Sethna, James P. 2006. *Entropy, Order Parameters, and Complexity*. Oxford: Oxford University Press.

Shah, Jagesh V., E. Botvinick, Z. Bonday, F. Furnari, M. Berns, and D. W. Cleveland. 2004. Dynamics of Centromere and Kinetochore Proteins: Implications for Checkpoint Signaling and Silencing. *Current Biology 14*: 942–952.

Shah, Jagesh V. and D. W. Cleveland. 2000. Waiting for Anaphase: Mad2 and the Spindle Assembly Checkpoint. *Cell* 103 (7): 997–1000.

Sharp, Phillip A. 2005. The discovery of split genes and RNA splicing. *Trends in biochemical sciences* 30 (6): 279–281.

Shen-Orr, Shai S., R. Milo, S. Mangan, and U. Alon. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics* 31 (1): 64–68.

Simon, Herbert A. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society* 106 (6): 467–482.

———. 1996. *The Sciences of the Artificial* (3rd ed.). Cambridge: MIT Press.

Simonetta, Marco, R. Manzoni, R. Mosca, M. Mapelli, L. Massimiliano, M. Vink, B. Novak, A. Musacchio, and A. Ciliberto. 2009. The influence of catalysis on mad2 activation dynamics. *PLoS biology* 7 (1): e10.

Smith, V., K. N. Chou, D. Lashkari, D. Botstein, and P. O. Brown. 1996. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science 274*: 2069–2074.

Sober, Elliott. 1983. Equilibrium Explanation. *Philosophical Studies 43*: 201–210.

Solé, Ricard V. and S. Valverde. 2006. Are network motifs the spandrels of cellular complexity? *Trends in ecology & evolution* 21 (8): 419–422.

Stephan, Achim. 1999. Varieties of Emergentism. *Evolution and Cognition* 5 (1): 49–59.

Sudakin, Valery, G. K. Chan, and T. J. Yen. 2001. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *The Journal of Cell Biology* 154 (5): 925–936.

Tomita, Masaru. 2001. Whole-cell simulation: a grand challenge of the 21st century. *Trends in biotechnology* 19 (6): 205–10.

van Regenmortel, Marc H. V. 2004. Reductionism and complexity in molecular biology. *EMBO reports 5*: 1016–1020.

von Bertalanffy, Ludwig. 1950. An Outline of General System Theory. *The British Journal for the Philosophy of Science* 1 (2): 134–165.

von Dassow, George, E. Meir, E. M. Munro, and G. M. Odell. 2000. The segment polarity network is a robust developmental module. *Nature* 406 (6792): 188–92.

Waddington, Conrad H. 1957. *The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.

Wagner, Andreas. 2005. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays 27*: 176–188.

Wall, Michael E., W. S. Hlavacek, and M. A. Savageau. 2004. Design of gene circuits: lessons from bacteria. *Nature Reviews Genetics* 5 (1): 34–42.

Wang, Jin, L. Xu, E. Wang, and S. Huang. 2010. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical journal* 99 (1): 29–39.

Waters, C. Kenneth. 1990. Why the Anti-reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1990, Volume 1, pp. 125–139.

Weaver, Warren. 1948. Science and complexity. *American Scientist 36*: 536–544.

Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Weisberg, Michael. 2007. Three kinds of idealization. *The Journal of Philosophy 104*: 639–659.

Westerhoff, Hans V and B. O. Palsson. 2004. The evolution of molecular biology into systems biology. *Nature biotechnology* 22 (10): 1249–1252.

Westerhoff, Hans V., C. Winder, H. Messiha, E. Simeonidis, M. Adamczyk, M. Verma, F. J. Bruggeman, and W. Dunn. 2009. Systems Biology: The elements and principles of Life. *FEBS Letters* 583 (24): 3882–3890.

Wimsatt, William C. 1972. Complexity and Organization. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1972, pp. 67–86.

———. 1987. False Models as Means to Truer Theories. In M. Nitecki and A. Hoffman (Eds.), *Neutral Models in Biology*, pp. 23–55. London: Oxford University Press.

———. 1997. Aggregativity : Reductive Heuristics for Finding Emergence. In *Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association*, Volume 64, pp. S372–S384.

———. 2007a. On building reliable pictures with unreliable data: An evolutionary and developmental coda for the new systems biology. In F. C. Boogerd, F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff (Eds.), *Systems Biology: Philosophical Foundations*, pp. 103–120. Amsterdam: Elsevier.

———. 2007b. *Re-engineering philosophy for limited beings: piecewise approximations to reality*. Cambridge: Harvard University Press.

Winnacker, Ernst-Ludwig. 1997. Das Gen und das Ganze (The Gene and the Whole). *Die Zeit*, 2 May.

Winther, Rasmus G. 2012. Mathematical modeling in biology: philosophy and pragmatics. *Frontiers in plant science 3*: 102.

Wolkenhauer, Olaf. 2001. Systems biology: the reincarnation of systems theory applied in biology? *Briefings in bioinformatics* 2 (3): 258–270.

Woodward, James. 2002. What Is a Mechanism? A Counterfactual Account. *Philosophy of Science 69*: S366–S378.

———. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

———. 2010. Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy* 25 (3): 287–318.

Yook, Soon-Hyung, Z. N. Oltvai, and A.-L. Barabási. 2004. Functional and topological characterization of protein interaction networks. *Proteomics 4*: 928–942.

Zamecnik, Paul C. 1962. History and speculation on protein synthesis. In R. Bellman (Ed.), *Proceedings of the Symposia on Mathematical Problems in the Biological Sciences*, Volume 14, pp. 47–53.

Zirkle, Raymond E. 1970. Ultraviolet-microbeam irradiation of newt-cell cytoplasm: spindle destruction, false anaphase, and delay of true anaphase. *Radiation Research* 41 (3): 516–537.