Final version to be published in Studies of History and Philosophy of Biological and Biomedical Sciences

Published online: http://www.sciencedirect.com/science/article/pii/S1369848613000265

When one model is not enough:

Combining epistemic tools in systems biology

Sara Green

Centre for Science Studies, Department of Physics and Astronomy, Ny Munkegade 120, Bld. 1520, Aarhus University, Denmark

sarag@ivs.au.dk

Abstract

In recent years, the philosophical focus of the modeling literature has shifted from descriptions of general properties of models to an interest in different model functions. It has been argued that the diversity of models and their correspondingly different epistemic goals are important for developing intelligible scientific theories (Levins, 2006; Leonelli, 2007). However, more knowledge is needed on how a combination of different epistemic means can generate and stabilize new entities in science. This paper will draw on Rheinberger's practice-oriented account of knowledge production. The conceptual repertoire of Rheinberger's historical epistemology offers important insights for an analysis of the modelling practice. I illustrate this with a case study on network modeling in systems biology where engineering approaches are applied to the study of biological systems. I shall argue that the use of multiple means of representations is an essential part of the dynamic of knowledge generation. It is because of – rather than in spite of – the diversity of constraints of different models that the interlocking use of different epistemic means creates a potential for knowledge production.

Keywords: modeling, reverse engineering, network motifs, Rheinberger, engineering analogies, historical epistemology.

1. The role of models in knowledge generation

This paper focuses on the use of models in scientific practice and reflects on knowledge generation through the integrative use of multiple models and epistemic frameworks. Despite the diversity of models - from mathematical representations to living model organisms - the categorization under the same name has relatively recently been justified by their common function as investigative instruments that mediate between theory and the world (Morgan & Morrison, 1999). The broad definition implies that anything that can be used as a facilitator of scientific reasoning through representation of target objects or of hypothetical systems may

be called a model. However, models are involved in several complex relations, and there is no simple linear relation between target and representation. Of importance is not only the connection between models and their targets, but also the dynamic relations between different models. Rather than analyzing how models represent targets, I shall argue for the importance of understanding how and why models are combined in iterative processes. A characterization of the virtues of models that only takes into account how single models represent target objects, delineating success of a model with structural similarity, has the limitation that it cannot account for how scientists learn about the world with the use of highly idealised and unrealistic models. Instead, models may be more adequately described as epistemic tools; as artefacts constructed for the purpose of manipulation and specifically constrained by their representational means and use in a concrete scientific context (Knuuttila, 2011). I exemplify this with a case study on network modelling in systems biology where models are used in a spiral-like fashion to develop insights about organizing principles for biological networks.

I focus on how models work at the intersection of the known and the unknown. When unmediated evidence for the features of a target phenomenon is not a possibility, models can establish connections between the known and the unknown (Rheinberger, 1997). Since models serve a variety of different epistemic goals in research and signify different representational means, a pragmatic account of models must take the variety of model functions into account (Leonelli, 2007). I propose that models are primarily informative from a relational perspective where they are connected to other models. The activity of modeling thus has an inherent systemic nature in being coupled to other models and entities, and the modeling activity should be seen in the context of how different models are brought together (see also Wimsatt, 1981; Weisberg, 2006; Winther, 2006; Knuuttila & Loettgers, 2011). The consequence of this view is that it is more important how models help us gain knowledge about an unknown target system through the relations to other representations.[1]

In exploring the use of multiple models in systems biology I shall draw on the conceptual repertoire of Rheinberger's historical epistemology where modeling is understood as the activity of "shuttling back and forth between different spaces of representation" (Rheinberger, 1997, p. 108). I provide an integrated account where modeling is understood as a matter of constructing, manipulating and comparing representations in a spiral-like fashion where a whole body of models interact. I first underline the aspects of Rheinberger's conceptual framework that I find particularly useful for understanding modelling and its relationship to experimental practices. Section 2 addresses this issue with a case study from systems biology where a variety of models are used to reverse engineer a biological network in search for design principles. Section 3 examines how methods and models from engineering and biology are integrated and combined in order to gain knowledge about organizing principles. Finally, I relate the account of this paper to recent philosophical accounts and discuss what is gained by employing this framework for discussing the role of models.

[1] The traditional representationist view defines representation in this narrow sense whereas I shall adopt a broader notion and specify different roles of models as representational means (see Knuuttila, 2011 for further clarification of the different notions of representation).

1.1. **A conceptual framework for understanding modeling**

Scientific reasoning is never as clear as it is expressed later when presented in published papers. During the process of knowledge generation the relevant representations are neither known nor readymade for comparison. When modelling is used for gaining knowledge about a (partly) unknown object the process is not an approximation to something already stable. New research objects are unstable in the sense that their properties and boundaries are neither known in advance nor directly accessible for observation. Because there is no direct representation of the yet unknown target the process of stabilizing a new entity happens indirectly through relations to different bodies of already certified knowledge and by production of new representations for comparison and integration. This process is often an iterative cycle of modelling and experimentation, where a series of representations with different constraints are combined. The constraining relations of epistemic tools are historically situated in what Rheinberger calls an experimental system (Rheinberger, 1997). With this notion Rheinberger captures Bachelard's description of the instruments of modern science as "theories materialized". The idea is that the technicalities of an experimental system productively constrain scientific reasoning by determining the realm of possible representations (Rheinberger, 1997, p. 20). His notion of experimental system is broad and includes instruments, models, and material and theoretical entities with varying degrees of scientific stability. *Stability* is here an indicator of how well the scientific objects, instruments or concepts are accepted by the scientific community and is thus an epistemic, rather than an ontological, category.

The focus on stability is important for understanding the role of models in research since models occupy a middle position in the continuum between two types of objects that experimental systems consist of. These are differentiated by their different degrees of (epistemic) stability:

1. *Technical objects* provide the stable context for experimentation. They can be material objects, concepts, systems of accepted knowledge, or instruments. Rheinberger defines technical objects as "answering machines", since they work as unquestioned tools to produce answers about epistemic objects.

2. *Epistemic objects* are unstable entities existing at the boundary between the known and the unknown. These can be physical structures or processes that are objects of investigation in a specific context at a particular time. Through the embodiment of what one "does not yet know", epistemic objects thus work as "question generating machines" that drive science forward.

Whereas technical objects set the boundary conditions for the possible representations within the experimental system, epistemic objects are only vaguely present in the context of a research project. They do not (yet) appear as stable scientific facts, methods or entities but are brought into existence within the context of an experimental system that embeds and constrains them. Knowledge production is spiral-like and takes the form of a continuous transformation of epistemic entities to technical entities, if they can become sufficiently

stabilized through a process Rheinberger calls *resonance* (figure 1, resonance will be clarified below). In this process models play a key role because they are situated at the interface of technical and epistemic objects, or between the known and the unknown. On one hand, models must include sufficiently reliable and stable features to enable researchers to make sense of the comparison and re-production of representations. On the other hand, they produce new questions and thereby help to constitute objects of knowledge through the embodiment and production of what is not yet known (Rheinberger, 1997, p. 110).

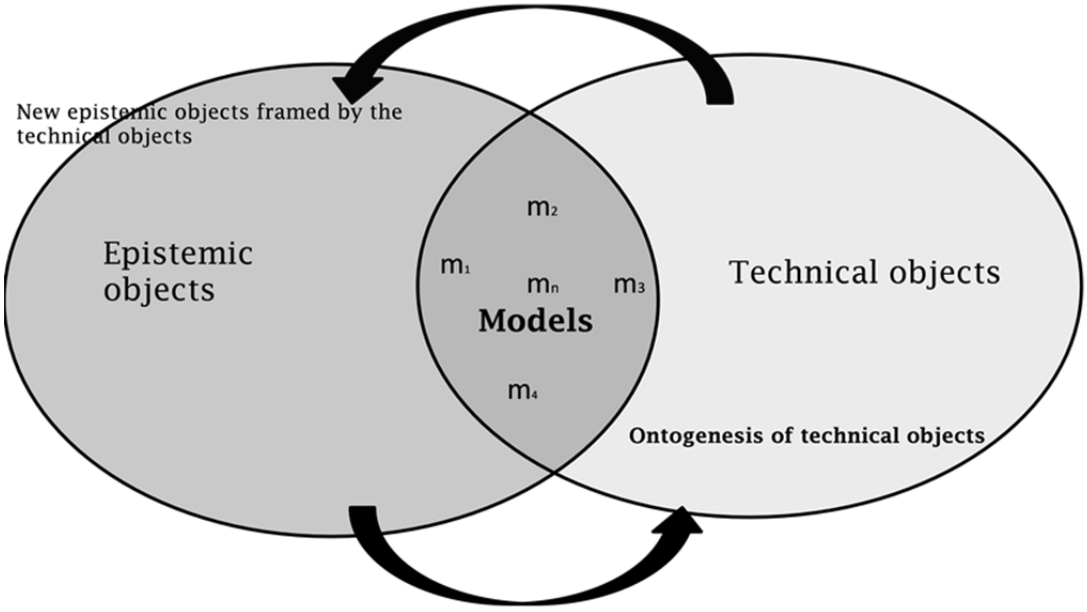

Figure 1, Knowledge generation: a self-correcting spiral of oscillations between representations. The ontogenesis of new technical objects goes from left to right as epistemic objects become sufficiently stabilized. But the production of epistemic entities goes from right to left as technical objects frame new epistemic entities. The epistemic stability and specific constraints of models depend on the historicity of the research contexts, the representational means and the epistemic goals inherent in the design and use of models.

The activity of modeling has a systemic nature. One model never addresses all relevant features of the target phenomena; if so, the model would be almost as complex as the target itself and of no use as an epistemic tool. For Rheinberger, the notion of representation relevant for understanding the role of models has nothing to do with approximations in the sense of mirroring a target. Rather, representation is a matter of producing and reproducing traces for comparison and manipulation. He distinguishes between three types of representation that will be discussed in section 3. Because of the complexity of most target phenomena, models inform by being instances of specific idealisations and manipulations that must be integrated with other models in order to be informative. Without a firm epistemic foundation to build scientific knowledge on, the scientists must rely on how well representations with different constraints can be integrated as a criterion for the soundness of the knowledge claim. Thus, scientific reasoning is a spiral-like pursuit of combining different representations to establish new scientific objects, a process Rheinberger calls *resonance* (see also section 3). The concept of resonance has several meanings. I here clarify the concept with an analogy that serves to explain why matching of different representations in this context is not a mere addition of pieces of information from different models, but a matter of generating and superposing material traces and epistemic tools with different constraints.

The concept of resonance in acoustics is defined as the tendency of a given system to oscillate with greater amplitude at some frequencies than at others. A sound will be prolonged if it is the result of a sympathetic vibration: a vibration in a body created as a response to a vibration of another body nearby with the same frequency as the 'natural frequency' of the body. At resonant frequencies the system will vibrate strongly compared to other frequencies. Knowledge production in science can be seen as analogous to the search for resonant frequencies. The materiality and different representational constraints put restrictions on what kind of representations can be made resonant. The difficult task is not only to compare representations, but to choose and produce representations that might have relevant constraints for shaping a new epistemic object. The scientist does not have the privilege of comparing his model to a "real", completely independent object but must compare the model to other models and shuttle back and forth between different representations (Rheinberger, 1997, p. 98). These differences explain why knowledge generation is possible as a spiral-like co-production of scientific objects. In Rheinberger's view, knowledge generation is thus a dynamic dialectic relation between the known and the unknown; between fact and artefact.

In the following I shall give a concrete example of how the different constraints of models, e.g. mathematical and material constraints, set the boundary conditions for the possible questions to be asked and answered by the experimental system. Rheinberger's terminology will be used in the discussion in section 3 in order to help with the difficult task of describing a process by which a new epistemic object comes into existence. This framework facilitates the recognition of how modeling can be understood as a loop between different representational spaces. However, Rheinberger does not offer a detailed account of how the specific constraints of different representational means can be used to evaluate the causal relationships and properties of the same object or process. Therefore, I shall in section 3 expand on this framework and relate his work to recent accounts in philosophy of science. Here, I emphasize that different representational means constrain in different ways and how the failure of a complete match of representations can be understood in terms of a productive tension for learning about differences in the representational domains. This discussion will draw on the case examined in the following section where multiple models and engineering approaches are applied to facilitate the study of biological networks.

2: Reverse engineering biological networks

In the wake of the new *omics* fields a large amount of data has become available for scientific analysis and provides a great potential for understanding the functional organization of living systems. However, the current approaches to realizing that potential still face many challenges such as i) the lack of sufficiently detailed data on dynamic processes with multiple variables, and ii) technical and theoretical tools to interpret and understand the data already available for analysis. A shift of perspective and the quantifiable tools of network theory have however offered unforeseen and promising possibilities for understanding functional organization in cells. Whereas many scientific achievements in molecular biology build on in-depth analyses of isolated components or pathways, the starting point for many projects in

systems biology is to study the topology of network models using mathematical tools. Rather than focusing on a specific entity, the researchers abstract from the molecular details to focus on systems properties. This choice can be based on various considerations including tractability and what kind of analysis the details of the dataset allow. More important for the epistemic aim of this strategy is, however, that the overall structural and functional organization of the network can be investigated.

General functional principles are often called organizing or design principles. The strategies of synthetic and systems biology reflect a hope that such features can be generalized across a wide variety of biological systems despite evolutionary contingencies. This research is often guided by the application of engineering principles. A key methodological framework called *reverse engineering* builds on an assumption of order in biological systems. The method starts by examining the solution - the existing structure or operation in living systems - and explores the function of this feature. The structures that appear in living systems are often tentatively assumed to represent (sub-)optimal solutions to environmental challenges. Reverse engineering is often employed in research projects with the aim of making sense of data from high-throughput analyses in e.g. transcriptomics and proteomics. In the current paper I exemplify this with the investigation of network motifs; possible design principles of biological networks defined by a group of systems biologists. The group is located at the Weizmann Institute of Science, Israel, with Uri Alon as the principal investigator. Alon is trained in physics and mathematics and joined the Department of Molecular Cell Biology at the Weizmann Institute in 1999 as a senior scientist. Alon's group uses a combined experimental and theoretical approach to decompose biological networks into functional units.

In reverse engineering biological networks, the systems biologists not only search for a function of a given structure; often the analysis begins by exploring the possibilities of relevant structures for a functional analysis of the network. For this reason Alon defines this type of reverse engineering as "mapping of an unknown network" (Alon, 2006, p. 239). Analogies from engineering can guide the search for order in biology by suggesting possible functional principles that are hypothesized to be similar within the two domains. A common analogy in both systems and synthetic biology compares cell circuits and electronic networks. Following the heuristic framework of this analogy, the group hypothesized that cells could be organized around small sub-circuits like electronic circuits built by engineers. This led them to believe that recurring circuit elements could be found that are analogous to general engineered designs like amplifiers and filters (Alon, 2006, and personal correspondence). The analogy had the advantage of focusing attention on a specific range of possibilities, namely small circuits within the network. Furthermore, the application of engineering terms also served as a rigid conceptual constraint for the formulation of possible biological functions regardless of the molecular details of the structures. However, the fact that engineering analogies are useful in the context of living systems is not obvious - while engineered systems are designed by an agent (a designer), biological systems evolve by a more complex process of evolutionary constraints where traits are highly interdependent. Therefore, the application of engineering principles has also met with critique (see section 3.1.). I return to this issue in section 3 but first I clarify how the engineering approaches were applied in this concrete project.

2.1. Searching for design principles

The aim of the group was to interpret data on transcriptional regulation in *E. coli*, available from the database RegulonDB and other sources. A now commonly used strategy is to picture the data as a computational network model that represents transcriptional interactions as a directed graph. Here, operons (one or several bacterial genes transcribed from the same mRNA) are represented as nodes, and regulatory transcriptional interactions are pictured as directed arrows, called edges. The group used this strategy to generate a network model that provided an overview of all the interactions (Alon, personal correspondence). In table 1 below this is called the "Network model of biological data". This is a graph where each edge is directed from an operon that encodes a transcription factor to another operon, representing the regulation of the target gene(s) by the transcription factor. The graph-theoretical framework provides a representational space where questions about the topology of the regulatory network can be asked without the need for molecular details. This makes it possible to model the overall interconnectivity of the network.

Name of the model	Basic features	Experimental data	Epistemic goal
1) Network model based on biological data Visual representation of transcriptional interactions on genome level as a directed graph	Overview of regulatory interactions represented as a network	Data on transcriptional regulation from RegulonDB and literature study	Basis for the search for recurring circuits in the network
2) Scanning algorithm (mathematical model)	A connectivity matrix to scan for recurring n-node sub-graphs	Model 1, transcriptional interactions represented as a graph	To find the most abundant n-node sub-graphs
3) Randomized network A collection of 1000 randomized networks (directed graphs)	Markov-chain algorithms. One type was based on the randomization of model 1. Another was constructed from an empty matrix	Type 1: data from model 1 as a starting point. No experimental data in the second type	To serve as background for comparison of networks
4. Activation profile Mathematical model	Graph displaying the input function, based on the structure of the network motif, using Boolean kinetics	The previous models serve as basis for describing the structure and possible functional dynamics of each motif.	A theoretical prediction of the functional dynamics of a network motif
5. Transgenic *E. coli* Model organism	Transgenic bacteria with fluorescent reporter plasmids controlled together with the promoters under study	The GFP activity (expression of green fluorescent protein) represents the activity of the respective promoters	To test the predictions of the temporal response time for systems controlled by different types of regulation

Table 1 The basic features of the different models examined. The first three models served to suggest key functional units based on the statistical significance of recurring patterns. Model number 4 suggested the functional dynamics of the network motifs. Model number 5 served to bring the theoretical and material representations together.

It has long been known that engineering principles such as feedback loops are common in biological systems. The analogy used by this group suggested a new analogue feature, namely that both electronic and biological networks could be decomposed into smaller

circuits due to the reuse of the same small circuit designs with specific regulatory functions (Alon, 2003; 2007b). The assumption was that if such recurring circuits could be found, it would suggest specific regulatory principles that may have been preserved through evolution because of beneficial functions. As expected from the analogy, but nevertheless a surprising finding given the differences between the networks, a high quantity of recurring loops was found in the biological network. The group defined these as network motifs.

Some of the motifs could be seen directly from the network model but to check for other important patterns, and to assure that the motifs were not just theoretical constructs, this representation had to be combined with other representations. For that purpose they compared the first model to a background model, consisting of corresponding networks for comparison but without the influences of natural selection (model 3, table 1). In order to establish a systematic automated search for network motifs, the group developed algorithms for automated pattern detection by scanning the network model for recurring 3, 4, and 5-node patterns (Shen-Orr et al. 2002). In table 1 this cluster of algorithms is referred to as the "scanning algorithm". The algorithms provided a quantitative representation of the networks as a list of numbers of n-node sub graphs, and served as the basis for a statistical representation of the network topology. The group argued that differences between biological and randomized networks point to the preservation of certain structures by natural selection against mutations that randomly change edges in the networks (Milo et al. 2002; Mangan & Alon, 2002). The comparison of networks is illustrated in figure 2.

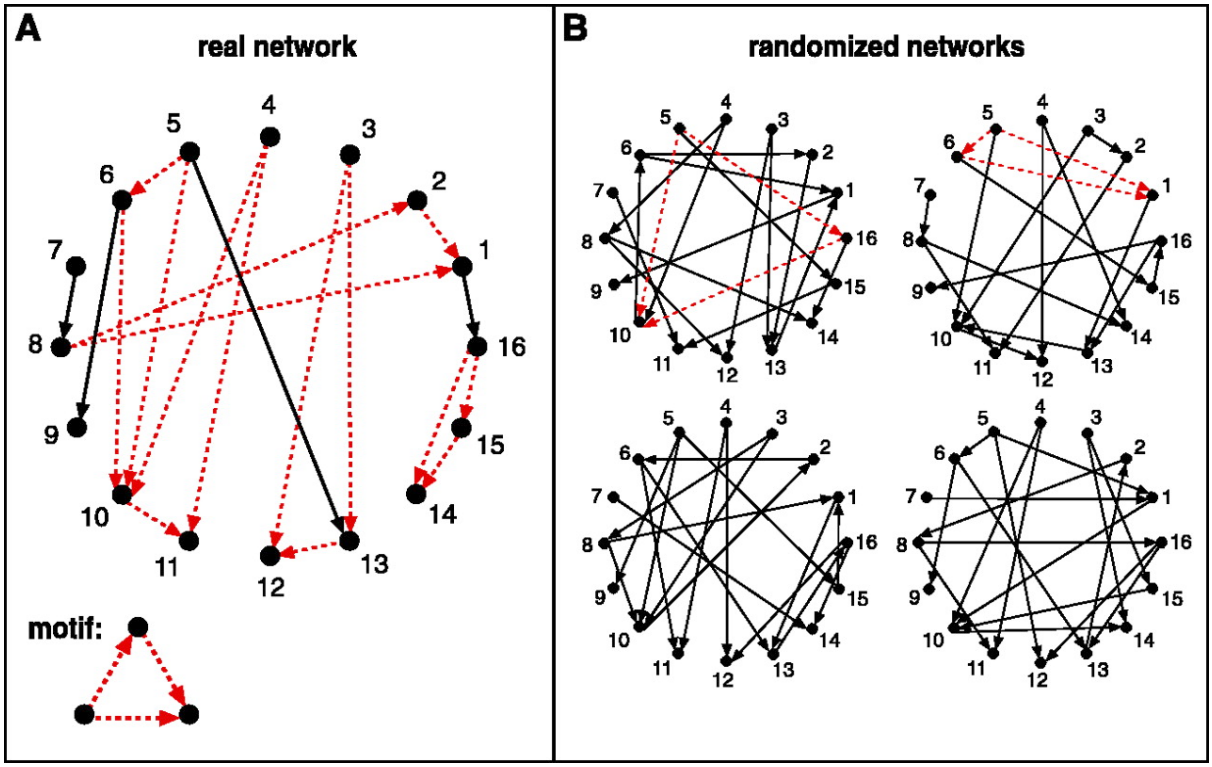

Figure 2. Comparison of networks. A "real" network here refers to the scientist's terminology for the network based on biological data. The figure shows how one motif occurs much more often in the real than in the randomized networks (dotted lines). Source: Milo et al. 2002.

The results of automated scans of both the network based on biological data and the randomized networks were compared in a statistical analysis and only statistically significant patterns were considered as network motifs. The result was that three motifs were statistically significant. These were defined as the Feedforward Loop (FFL), Single input module (SIM) motif and Dense Overlapping Regions (DOR) (see figure 3).

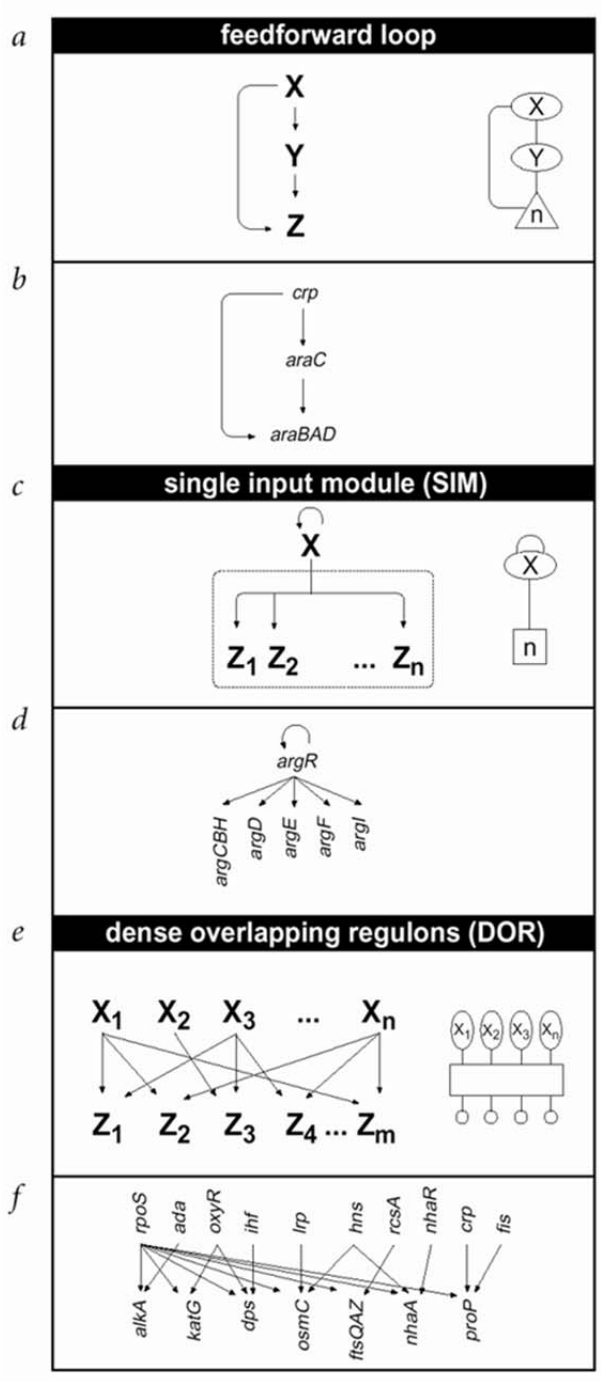

Figure 3 illustrates network motifs found in the *E. coli* transcriptional network (a, c, e) with examples from empirical studies (b, d, f). **a**, Feedforward loops are three-node patterns with two transcription factors jointly regulating target operon(s). **c**, SIM motif, where a single transcription factor regulates several operons, **e.** DOR motif, where a set of operons (Z_m) are regulated by a combination of different transcription factors (X_n). Source: Shen-Orr et al. (2002).

In the following we shall concentrate on the FFL. This is a three-node loop where a gene Z is regulated by two transcription factors, X and Y (see figure 3a), and where X simultaneously controls both Z and Y. The FFL is one of the motifs that since the first study have been found in several different species and different networks (Alon, 2006). Of interest is not only what motifs are possible, but also what motifs never or rarely appear in biological networks. These are called anti-motifs. Following the representation of the FFL in figure 3, there are 8 possible FFLs, since each of the arrows from transcription factor to operon can be either positive (activator) or negative (repressor). Four of those are called *coherent* FFLs, where the direct path from X to Z has the same overall sign (they both either activate or repress the transcription). In the other four, called *incoherent* FFLs, the direct path and the indirect path have opposing signs. Only two of these 8 possibilities are found in biological networks: a coherent FFL and an incoherent one (Alon, 2006).

The scanning algorithm and the randomized networks showed that decomposition of the network into specific sub-modules was possible, and the statistically significant motifs matched the patterns seen directly in the network model. The quantitative significance pointed to a possible functional design of specific motifs, since some motifs are statistically

significant and other motifs completely absent in biological networks (Shen-Orr et al, 2002). The next step was to investigate the functional difference between different motifs.

2.3. Exploring the dynamics of network motifs

To analyse the dynamics of the motifs, the group created a mathematical model, specific for each motif. This model displayed the regulation as a time-dependent response to different signal inputs, i.e. shifting concentrations of sugars in the environment of *E. coli.* When transcription factors bind to the promoter region of DNA they can either activate or repress the transcription of genes, and thereby increase or decrease protein synthesis. In a simplified description, transcription factors exist in an inactive (X) or active state (X*), according to the availability of a signal input, S_x. This signal input corresponds to the concentration of sugars available in the bacteria's environment.

The dynamics can be modelled as input functions of the motifs (see figure 4). An input function is a mathematical representation where the rate of production of a protein, Z, is a function of the activity of transcription factors. The concentration of the transcription factors is dependent on factors such as dilution, degradation etc. but most importantly on the concentration of the signal inputs, S_x and S_y, and the time-dependency of binding of promoters. These variables of multi-dimensional input functions can be idealized using Boolean kinetics. The result of the input function for the coherent FFL is shown in figure 4. FFLs can be regulated following an AND gate logic, where both transcription factors are needed simultaneously to activate Z, or an OR gate logic, where activation of only one transcription factor is needed. In the following I concentrate on the coherent FFL with AND gate logic. AND gate logic means that for the gene Z to be expressed, it needs to be activated by both transcription factors, X and Y, and they must in turn be activated by signal inputs (S_x and S_y respectively). With this graphical representation of the input function the researchers showed that the coherent FFL might function as a sign-sensitive delay element. The delayed response is due to the time difference between direct and indirect activation routes, and the requirement for the concentration of Y* to pass a certain threshold for persistent stimuli of both S_y and X* (via stimulation of persistent stimuli of S_x). Sign-sensitivity in this context means that the FFL can act as a persistence detector of stimuli where short pulses of S_x do not lead to activation of Z (see figure 4, time unit 3). This feature is however not displayed if S_x disappears (OFF-step pictured at time unit 15 on figure 4): if X* changes to its inactive state, the system rapidly shuts down and Z is not expressed. Thus, the proposed function of the activation profile is that the coherent FFL provides a sign-sensitive delay of gene expression as a response to ON steps but not to OFF-steps. In comparison the incoherent FFL has the opposite dynamic profile and shows a delay response to OFF-steps.

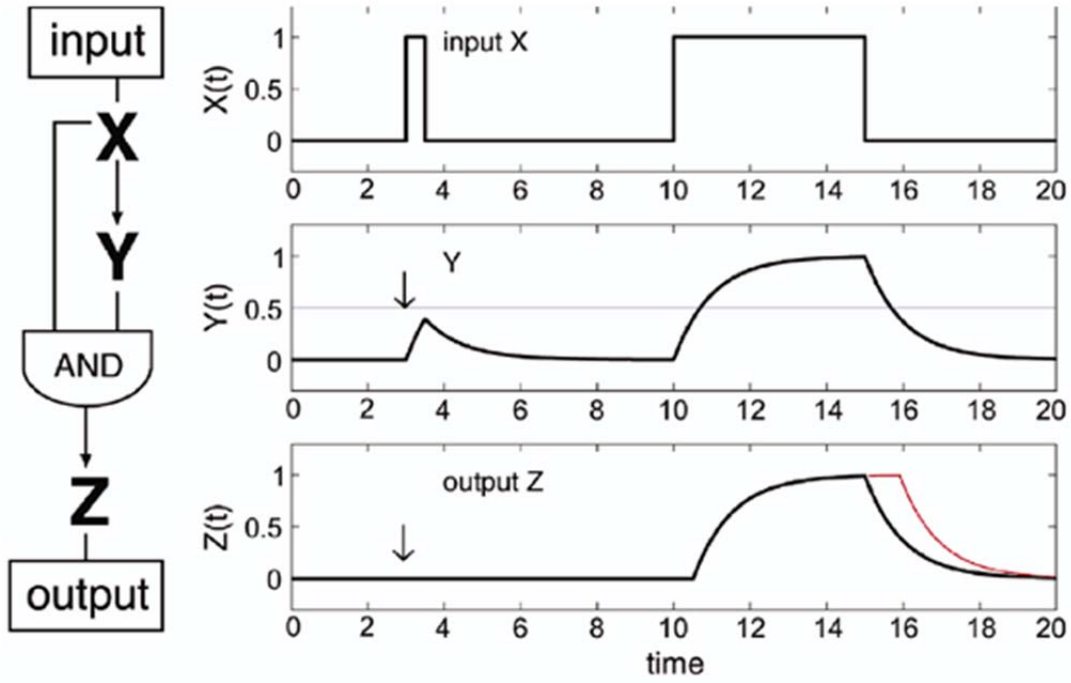

Figure 4. Activation profile of the coherent FFL. Source: Shen-Orr et al. (2002).

The activation profile represents only a possible biological function, since the constraints imposed on this analysis are primarily mathematical, involving idealizations such as the Boolean gate. Furthermore, the activation profile displays the function of motifs *in isolation*, whereas motifs in living organisms are embedded in hundreds of other interactions. Since these could strongly influence the function, it was not clear whether the dynamics predicted by the input function presented the function it displays in living organisms. The efforts of modeling therefore needed to be integrated with experiments that included material constraints of biological systems.

2.4. Combining different representations

To address the question of the function of FFLs in living cells, Alon's group chose the *L*-arabinose system in *E. coli* that was found to be regulated by the coherent FFL. This system was compared to another system with *simple* regulation, the *lac* operon (Mangan et al, 2003). The arabinose system encodes enzymes that utilize the sugar arabinose in the presence of arabinose and absence of glucose. The production of enzymes to digest arabinose is regulated according to the availability of sugars in the environment. In order to test the predictions of the model experimentally, the group measured the promoter activity for the two operons as a response to shifting concentrations of sugars in the controlled environment. Since promoter activities cannot be measured directly, the group incorporated reporter plasmids in which the *ara* and *lac* promoters also control a green fluorescent protein (*gfp* gene). The expression of green fluorescent protein (GFP activity) can be measured by automated multiwell fluorimeters. This measure indirectly represents the activity of the respective promoters because the fluorescent protein is expressed simultaneously with the ara/lac genes by the *E. coli* strains. The ara system was manipulated by controlled shifts in concentration of signal

molecules, S_y=arabinose and S_x=cAMP (a molecule produced during glucose starvation). The predicted dynamic was tested by comparing curves for response time to ON and OFF-signals for the two systems. Despite the context of other interactions, the curves created from measurements of the feedforward regulated arabinose system closely matched the overall prediction of the C1-FFL as a sign-sensitive delay element (Mangan et al, 2003). The ON response of the FFL-controlled system was significantly slower than the simple regulation, whereas the OFF response in the two systems was almost identical.

The network motifs are now a part of the basic terminology of systems biology, and the methodology developed by Alon's group and colleagues has become a widely known framework for further experimentation, e.g. the search for these and other network motifs in different networks and different species. As a consequence Alon has argued for the productivity of why-questions in systems biology, i.e. of asking why a system has a specific structure or organization (Alon, 2006). He proposes that viewing nature not only as a tinkerer but also as analogous to an engineer throws light on design principles behind biological network architecture (Alon, 2003; 2007a; 2007b). It may even throw light on questions regarding ecology. The following section describes a theoretical cost-benefit analysis called inverse ecology proposed by Alon's group. Section 3 discusses philosophical implications of the integration of methods from engineering and biology and combining multiple models.

2.5. Inverse ecology and the evolution of motifs

The discussion about the functional significance of different motifs has a dual nature. A central part of the analysis regards the function of motifs expressed in engineering terms, i.e. a function rigidly defined as a response to input signals depending on the structure of the motif in question (section 2.3.). The result of this analysis is a general function a network must have given the characteristics related to this type of motif (see also Levy & Bechtel, forthcoming). Another issue regards the possible adaptive values of motif-based functions in different types of environments. Here, the functions of network motifs are considered in relation to biological purposes of the control-mechanisms and the possible selective forces behind the origin of such design principles. When it is argued that the FFL must have been selected due to its function as a sign-sensitive delay element, it is implied that there exist strong fitness-related benefits connected to the responsiveness of transcriptional networks to changes in external stimuli. This assumption has been addressed by a comparison of different environmental characteristics.

The experiments on the arabinose system led to the question of what types of environments could facilitate the selection of specific motifs. Since sufficient information on the natural environment of many organisms - including *E. coli* - is lacking, theoretical discussions can provide suggestions to guide the experimental investigations. Mangan et al. (2003) discuss the advantages of the FFL and ask under what environmental conditions the coherent FFL would be selected over simpler gene circuits. This issue was later addressed in theoretical cost-benefit analyses for the selection of network motifs in a given environment (Dekel & Alon, 2005; Dekel et al. 2005; Kashtan & Alon, 2006; Kalisky et al. 2007). Following the implication of regulation of specific motifs, a profile of the natural environment of the organisms can be made. This is a profile of the properties of the signal

inputs (here the average duration of the signal in time) that would make each of the motifs beneficial for organisms in particular environments.

The theoretical approach is connected to an analogy of economically optimized energy budgets for organisms. This can be illustrated by the arabinose system. Since glucose is the preferred energy source for bacteria, the arabinose system should, following this analogy, only be expressed when glucose is absent. The length of pulses of glucose is important here, since short pulses of the starvation signal cAMP would have a detrimental effect on growth if Z is unnecessarily produced as a response to a brief input pulse. As the modeling and experiment showed, the coherent FFL can filter out such short pulses. In addition, whenever glucose appears or arabinose disappears, it is beneficial to immediately shut down the arabinose system - a feature the FFL shares with the simple regulation. The result of the analysis was that the C1-FFL is selected over simple regulation in environments with multi-modal pulses, where a persistence detector to filter out short input signals is beneficial. The assumption of an energy budget related to regulation by motifs thus suggests possible constraints on the environments in which the systems evolved. The cost-benefit provides a mathematically controlled comparison between different designs that can provide productive guesses about environmental features. The network motifs are here assumed to be instantiations of optimal circuit designs, representing an internal model of the environment in which they were selected. The group defined this method of deducing information about the environment from the motifs detected as *inverse ecology* (Dekel et al. 2005). This terminology has caught on as the concept ‘reverse ecology’ which now covers the use of information from genomics to gain knowledge about ecological interactions (e.g. Borenstein et al. 2008).

As we have seen, the notion of general design principles can guide the development of new hypotheses and methodological approaches in biology. However, the heuristics of optimal design in biology has also led to a discussion on the limitations of engineering principles in biology (e.g. Lynch, 2007). In the following this issue will be addressed. I first introduce a *dual notion* of constraints and show how this can be used to throw light on the attempt of applying engineering principles to the study of living systems, and to the interlocking use of multiple models.

3. Constraining new epistemic entities

In section 2 I have described how network motifs emerged as new epistemic entities from the combination of different modeling strategies and through the integration of methodologies from engineering and biology. In the following I emphasize the importance of different constraints for creating new spaces of representations. That different representations are differently constrained is implicit in Rheinberger’s definition and description of experimental systems, but the issue has recently been more explicitly addressed and contextualized by other philosophers of science (e.g. Nersessian, 2002a; 2002b; Knuuttila, 2011). A focus on constraints helps to understand how scientific reasoning is guided by a productive tension between enabling and disabling constraints. Only a selection of possibilities is considered when formulating and addressing a scientific problem in terms of selected representational tools, and this may limit the reasoning by neglecting some issues in

favor of others. However, this is at the same time what makes it possible to pick out relevant epistemic units for analysis. The combination of epistemic means with different constraints provides access to new spaces of representation unavailable without these constraints.[2] A better understanding of the micro-dynamics of knowledge-generation can be reached by analyzing the characteristic constraints of the different types of representations the modeler must combine. I first examine how the integration of multiple models and methodologies from engineering and biology can be understood. I then compare Rheinberger's framework to recent philosophical accounts addressing similar issues. I argue that these together provide an adequate view on knowledge generation through a convergent spiral-like process.

3.1. Biology meets engineering

In the case study examined an engineering analogy was employed to suggest possible principles of order in biological systems. What provides the optimism that engineered and biological networks could be similar? One possibility is that the functional similarities between the systems are stable enough to make analogy transfer productive in both domains, because both engineered devices and biological systems are constantly evaluated for their ability to respond to changing parameters inside and outside the system. This may explain why conceptualization of biological networks in terms of control theory can open a new epistemic space for understanding biological function, just like solutions in organisms have inspired engineers to develop and improve their designs. In the following I shall argue that the actual mapping of similarities is only one part of the story, but first I shall reflect on the discussion regarding similarities and dissimilarities.

The transfer of epistemic units across these domains is guided by different types of analogies. Material analogies can provide a basis for constructing formal analogies that in turn may guide the production of hypotheses about causal relations (Hesse, 2001; Nersessian, 2002a; Knuuttila & Loettgers, forthcoming). For Alon the analogy between biological networks and electronic circuits gave rise to the formal analogy of the graph theoretical framework described. This framework provides enabling constraints for analyzing biological networks with the use of mathematical models such as algorithms that scan the network for recurring subunits. Thus, the use of engineering principles affords a conceptualization of biological functions in language from control- and graph theory. Engineering terms introduce a more rigid functional language which, unlike a purely qualitative analysis, can be operationalized mathematically.[3] When network motifs are described through concepts like persistence detectors, pulse amplifiers, circuits, bi-stability etc. the functional descriptions can often be given a graphical and quantitative representation that can sharpen the predictive value of functional analysis.

The search for design principles is an attempt to find simple and general principles of order within the domain of biological complexity. Recent findings in systems and synthetic biology give some optimism that such principles could be found by tentatively viewing the "tinkerer" (nature) as an engineer (Alon 2003; 2007b). However, engineering approaches

[2] A similar argument has been advanced by Knuuttila and Loettgers (Knuuttila, 2011; Knuuttila and Loettgers, 2012).
[3] I would like to thank Pierre-Alain Braillard for pointing this out to me.

clearly have their limitations for analyzing the *origin* of structures. Behind the expectation of optimality in biological networks is often a (working) assumption of convergent evolution where natural selection is imagined to have improved the functionality of specific traits. In the case of network models problematic aspects are i) the problem of individuating network motifs as individual traits, and ii) general network construction processes that might point to the status of network motifs as by-products (or spandrels) rather than optimized design principles (Solé & Valverde, 2006). Lynch has argued that the application of engineering principles in biology tends to bring in an unwanted notion of design, if not directly through adaptationist assumptions about convergent evolution, then of purpose-laden engineering terms such as amplifiers, filters and sign-sensitive delays as we have seen in this case (e.g. Lynch, 2007). The assumption of convergent evolution has been challenged by recent work arguing for the possibility that structures such as the FFL could have evolved neutrally through processes of duplication, mutation, deletion and recombination of genes and binding cites (Cordero & Hogeweg, 2007; Knabe et al. 2008). Thus, design analogies may provide a disabling constraint when it comes to the study of the origin of biological structures, since design analogies can lead to a neglect of non-selective forces.

The question is now whether this problem leaves the integration of engineering principles unproductive? Are there scopes of representation that should not be combined because they tend to be misleading? In the following section I shall argue for the productivity of the matching of representations despite the lack of overlapping constraints. However, a stronger focus on the tensions in the constraints of e.g. engineering approaches and mathematical embedding in biology not only introduces insights into how these integrative efforts are productive but also a greater awareness of their limitation.

Rheinberger states that experimental systems operate "at the border of their breakdown" (Rheinberger, 1997, p. 135). This means that the integrative tendency of experimental systems involves a degree of instability where unpredictable results may occur. In this framework the resistance to the matching of representations is not seen as counterproductive to the goal of establishing resonant links, but as a dialectical resistance that makes the realization of new epistemic things possible: "The reality of epistemic things lies in their resistance" (Rheinberger, 1997, p. 23). Similarly, the resistance to the application of engineering principles can be interpreted as a sign of biological specificity of the epistemic thing under analysis. Knuuttila and Loettgers (this volume) give an excellent example of this. When the engineering concept of controllability is transferred to the study of biological systems, the unforeseen differences between the systems provide insights into the functional role of noise for attaining robust biological functioning. A similar example can be drawn from the current debate on network motifs and the discussion of whether network motifs can be functionally isolated in other living organisms than *E. coli*. It has been questioned whether network motifs like the FFL have a general *biological* function, like the one described in section 2.4., because network motifs are embedded in many other interactions in living systems (Mazurie et al. 2005). Further experimental evidence is needed to settle this issue, but it is possible to discuss what consequences a possible negative result would have for the view on engineering approaches. One possibility is that the result can be interpreted as a shortcoming of the engineering approach in biology – a failed attempt to establish general design principles that have similar functions despite evolutionary contingencies. However,

the negative analogy can also be seen as a productive dialectical resistance that helps to generate knowledge about what is specific for biological and engineered systems, respectively. This might provide insights into how the transcriptional networks of prokaryotes and eukaryotes differ. Alon himself finds it likely that more complex cases will be found where the function of motifs to a greater extent is affected by other interactions in the network (Alon, 2007a). Thus, the negative analogy can condition the generation of knowledge when the attempt of transferring epistemic units meets resistance.

The usefulness of engineering principles in biology cannot be foreseen, but neither would it be a complete coincidence; it depends on an open structure of the investigative process and on properties that resonate when compared on an abstract level. Rheinberger uses the term conjecture to describe the emergence of a new constellation that opens the way for what he calls unprecedented events. He prefers the term conjecture rather than 'discovery', since scientific objects never present themselves as readymade facts. Conjectures are not possible to predict, but at the same time they are not events happening by accident. The events that create conjectures he calls hybridizations of different, originally independent systems (Rheinberger, 1997, p. 135). A similar notion is integration that is increasingly discussed in philosophy of science (cf. O'Malley & Soyer, 2012). In both accounts the integrative effort implies that something new is generated when e.g. engineering and biology meet that extends the application of stable knowledge to a new context.

The openness may, however, change during a research project. The initial work on network motifs can be described as data-driven modeling, where possibilities were explored based on the engineering analogy described. Compared to the later model-driven experimentation, guided by a well-defined hypothesis, this initial research was highly exploratory. Rheinberger's approach offers an insightful description of the workings of experimental systems, but he does not offer conceptual tools to distinguish between such different modes of research. A finer grained framework is proposed by recent accounts discussing a distinction between hypothesis-driven research and 'exploratory experimentation' (e.g. Steinle, 1997; O'Malley, 2007; O'Malley & Soyer, 2012). Such a distinction is especially relevant for describing new developments within the life sciences where automated hypothesis-generation from data is currently a hot topic of debate. Rheinberger does, however, address different degrees of openness through a distinction between different degrees of representational strength in science. This distinction is useful for understanding how the network motifs gained stability from the exploratory investigation of networks to an experimental demonstration of predicted functions of network motifs. This issue will be discussed in the following section.

3.2. Epistemic objects at the intersection of constraints

During the effort to interpret, by modeling, transcriptional regulatory networks, the network motifs developed from being a concept derived from a loosely applicable engineering analogy to become a biological concept widely known in systems and synthetic biology. How can this transformation be understood? In the following I shall argue that the interlocking use of the different models (table 1) provided the network motifs with an increased representational strength. Rheinberger distinguishes between three types of model-

based representation with correspondingly different degrees of representational strength. The first is representation as an analogy where the objects or systems represented are intentionally related by different kinds of similarities. The second is representation in the form of models or simulations, and the third is representation as an experimental realization. The strength of representation differs correspondingly from "a continuum from vicarship to embodiment to realization" (Rheinberger, 1997, p. 103). In other words, there is a transformation from representations *as* to representation *of*, and to a realization where the represented transforms into a stable object.[4] The choice of the concept "vicarship" has caused some confusion, and Rheinberger does not sufficiently clarify the meaning of this term. I interpret this type of a representation as an analogue substitute, such as a material analogy like the hormonal system represented as a thermostat. In the following I shall expand on these notions to describe how the constraints of different models in the case examined transformed the status of the new epistemic thing, the network motifs. The models described (table 1) are connected through a series of common assumptions and intentions with their design and are thus not independent. However, as I shall show in the section below, what is important is that they embody different types of representational means that have different constraints.

In the case study examined, the analogy between electronic circuits and biological networks served as an analogue substitute. The creation of the network model of the dataset transferred the transcriptional interactions to a new representational space that afforded both a visual pattern-recognition and an automated scan for recurring subunits. As described in section 3.1., analogies should not only be considered productive for fixing shared features between two domains but have a productive role as tentative and unstable assumptions (see also Knuuttila & Loettgers, forthcoming). The analogy constrained this analysis, since only small subunits – similar to those of electronic circuits – were scanned for. In the first study three different patterns were classified as network motifs due to their statistical significance. The mathematical constraints of the scanning algorithm and the statistical comparison provided support to the overabundant patterns seen by eye by automated pattern recognition. The recent critique of the method employed can be interpreted as a criticism of the lack of evolutionary constraints in this comparison, since the background randomization in the models for comparison does not resemble the evolution of biological networks (see section 3.1.). However, because of, rather than despite of, this false assumption, structures were found that are general for many networks. This is an important insight even if the functionality turns out to be different in different networks, and even if the network motifs are not results of convergent evolution. The function analysed using input functions may even be especially informative because it lacks the biological constraints of other interactions and only provides information about motifs in isolation. This analysis can thus provide the basis for the next step, where knowledge about biological contexts may be gained from the deviation from this function. For the experimental realization of network motifs as *biological*

[4] The distinction as it is used here is not a matter of the accuracy of representation of a real world target, since models never directly represent and often depict hypothetical and fictional systems. The distinction refers to the way the target is represented and how stable this historical relation is – whether this takes the form of a loosely applicable analogy, as embodiment of a property, or as the realization of new knowledge. In the latter case the model may lose its role as a model and become a more direct representation of our understanding of the target.

organizing principles it was, however, important that the same function could be experimentally demonstrated. How did the researchers arrive at this stage?

During the analysis of the functional role of the motifs the strength of representation of the models shifted from being an analogue substitute - a treatment of biological networks *as if* they were analogous to an engineered network – to a more substantial embodiment of (biological) research questions regarding the functionality of the motifs. The activation profile thus embodies a motif's function. But since the constraints imposed on this analysis are primarily mathematical, the function was not realised in a biological context. In this example the target of the transcriptional regulation was treated as a Boolean gate with the number 0 or 1, denoting that the genes are always either off or maximally on. This idealization is a reduction of the complexity found in living systems since this is usually a graded response. However, the epistemic goal of using this model was not to have an accurate representation based on precise values of parameters. Rather the aim was to create an *idealized epistemic space* where the functional significance of *different* motifs could be analysed. This comparative mathematical analysis allows for a quantification of biological functions where the difference between e.g. coherent and incoherent FFLs can be given a precise measure and thus has predictive value for the following experimental designs. The important relation was foremost how the mathematical constraints of the model of one motif differ from that of another motif. Accurate representation was thus secondary to imposing simple logic models that provide mathematical constraints for limiting the number of possible functions of the network motifs.

As mentioned, the mathematical model displays the function of motifs in isolation, whereas motifs in living organisms are embedded in hundreds of other interactions that could influence the regulatory mechanism. Therefore, the efforts of modeling needed to be integrated with experiments that included material constraints of biological systems. In understanding this step toward an experimental realization, the material resistance of biological phenomena is important. It provides a challenge to purely theoretical forms of coherence, and the coupling of in vivo and in vitro approaches is thus crucial for creating strong resonant links between theoretically and materially constrained representations (Rheinberger, 1997; Knuuttila & Loettgers, 2011). The material constraints of model organisms explain why they often have a special epistemic status. Because they are made of the "same stuff" as the target object, including the same internal composition and functional organization, they provide productive, rigid constraints for the shaping of an epistemic object.[5]

The effort of combining mathematical predictions with experimental manipulations was crucial for establishing a link between the (highly idealized) mathematical representation and the biological phenomena. The crucial step here is the *re*-presentation of the epistemic entity in other contexts with different constraints. In this sense there is an iterative loop of reinserting and reproducing the emerging stable features into new experimental contexts. The experiment provided information on e.g. the specific concentrations of input signals required for activation of transcription factors. This made it possible to create a more exact dynamic

[5] In addition to this they have a special status as representatives of the whole organism and a broad representational scope of similar organisms (Ankeny & Leonelli, 2011). For an insightful analysis of the similarities and differences between modeling and experimentation, see Knuuttila & Loettgers (2012).

model for the specific regulatory system, and the experimental result can in this sense be said to provide feedback to the theoretical model in an iterative cycle. However, this experiment confirmed that the overall dynamic features do *not* depend on the details of the parameters in the input function. This gave some optimism that possible organizing principles can be investigated using a simple activation profile without the requirement of detailed information on parameter values of the specific systems they are found in. Once this method was developed, it was repeated with different motifs and different datasets to further stabilize the concept of network motif. This part of the research can be described as highly hypothesis-driven, and the fact that the network motifs were found in a variety of different organisms and in different networks (see Alon, 2006) further supported the experimental realization of the new epistemic entity. Due to the recent critique, network motifs can, however, still be considered as question-generating machines – as partly stable epistemic entities that produce new questions for further experimental analysis (Mazurie et al. 2005; Knabe et al, 2008).

To conclude I want to emphasize a strong parallel between Rheinberger's account and existing accounts of modeling within philosophy of science. Rheinberger's notion of resonance has strong similarities with what Levins and Wimsatt call robustness, allied to Campbell's notion of triangulation (see Wimsatt, 2007). Knuuttila and Loettgers draw on this framework in an integrated account of modeling where scientific reasoning builds on the triangulation of different epistemic means, including synthetic modeling and experimentation (2011; 2012, and this volume). Whereas Rheinberger has stressed the productive tension between the known and the unknown, or the dialectic between fact and artefact, Knuuttila and Loettgers stress a similar dialectic of constraints where it is the "relatedness and independence that gives combinatorial modeling its epistemic leverage leading to iterative cycles of modeling" (Knuuttila & Loettgers, this volume). In both frameworks there is an emphasis on the cyclic nature of knowledge generation through iterativity or resonance of different epistemic means. Rheinberger provides a series of interrelated categories for conceptualising transformations in science that are difficult to intellectualize because they concern something pre-theoretic; the generation of knowledge about what is not yet known. With the notion of the irreducibly vague concept of epistemic objects it is possible to speak of an object of study that is still in the process of becoming a scientific entity. Rheinberger does not, however, explicitly address the issue of how specific constraints are triangulated in the process of resonance. For this reason, I have added a dual notion of constraints of engineering principles applied to biology and of the models employed in the case study.

Conclusion

I have examined a case study in systems biology where engineering principles were applied to a biological network investigated by multiple models. Drawing on Rheinberger's terminology, I have argued that integrative activities provide a potential for understanding how new knowledge is gained from only partly stable constellations of multiple models and representational spaces brought together. On one hand, experimental systems must be sufficiently stable and constrained to allow for the comparison and reproduction of epistemic units. On the other hand, they must be unstable enough to allow for unpredictable results. I

have shown how the combination of mathematical and biological constraints in the modeling process was crucial in the case examined where network motifs emerged as new epistemic objects in systems biology. The specific properties of productive constraints for generating knowledge is, however, often only clear in retrospect. At the time when research is carried out, knowledge has to be gained not only of how well the different constraints fit the yet unknown constraints of the emerging epistemic object, but also about the specificity of the representational spaces brought together. It is in this sense that models can be described as only partly stable because they embody a share of what is not yet known. This tension between stability and instability and between different types of constraints explains what makes experimental systems and the models within them productive. A better understanding of how specific constraints are triangulated thus provides insights into how modeling is not a matter of accurately representing targets but of generating, manipulating and superposing different epistemic tools to learn about what is not yet known.

Acknowledgements

I would like to thank Hanne Andersen, Maureen O'Malley, Tarja Knuuttila, Rasmus Grønfeldt Winther, Sabina Leonelli, Brian Hepburn, and two anonymous reviewers for valuable suggestions and comments on earlier versions of this paper. I also thank Uri Alon for helpful guidance through the published work and databases of his group, for email correspondence and participation in a telephone interview. The research was funded by a grant for the project group of Philosophy of Contemporary Science in Practice by the Danish Research Council for Independent Research, Humanities.

References

Alon, U. (2003). Biological networks: The tinkerer as an engineer. *Science*, 301(5641), 1866-1867.

Alon, U. (2006). *An introduction to systems biology. Design principles of biological circuits*. Chapman and Hall/CRC Mathematical & Computational Biology.

Alon, U. (2007a). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, 8, 450-481.

Alon, U. (2007b). Simplicity in biology. *La Nature*, 446(7135), 497.

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A*, 42(2), 313-323.

Borenstein, E., Kupiec, M., Feldman, M. W., & Ruppin, E. (2008). Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of science*, 105(38), 14482-14487.

Cordero, O. X., & Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. *Molecular Biology and Evolution*, 23(10), 1931-1936.

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050) 588.

Dekel, E., Mangan, S., & Alon, U. (2005). Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Physical Biology*, 2(2) 81.

Doyle, F. J., & Stelling, J. F. J. (2006). Systems interface biology. *Journal of the Royal Society*, 3(10), 603-616.

Kalisky, T., Dekel, E., & Alon, U. (2007). Cost–benefit theory and optimal design of gene regulation functions. *Physical Biology*, 4(4), 229-245.

Kashtan, N., & Alon, U. (2006). Spontanous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences*, 102(39), 13773-13778.

Knabe, J. F., Nehaniv, C. L., & Schilstra, M. J. (2008). Do motifs reflect evolved function? -No convergent evolution of genetic regularoty network subgraph toplogies. *Biosystems*, 94, 68-74.

Knuuttila, T. (2011). Modeling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42, 262-271.

Knuuttila, T., & Loettgers, A. (2011) Causal isolation robustness analysis: The combinatorial strategy of circadian clock research. *Biology and Philosophy*, 26(5), 773-791.

Knuuttila, T., & Loettgers, A. (2012). Modeling and experimenting: The combinatorial modeling strategy of synthetic biology. *Philosophy of Scientific Experimentation: A Challenge to Philosophy of Science*, Pittsburgh (October 15-17).

Knuuttila, T., & Loettgers, A. (this volume). Basic science through engineering? Synthetic modeling and the idea of biology-inspired engineering. *Studies in History and Philosophy of Biological and Biomedical Sciences.*

Knuuttila, T., & Loettgers, A. (forthcoming). Varieties of Noise: Analogical Reasoning in Synthetic Biology. *Studies in History and Philosophy of Science*.

Kuo, D. P., Banzhaf, W., & Leier, A. (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.

Leonelli, S. (2007). What is in a model? Combining theoretical and material models to develop intelligible theories. In M. D. Laubichler, & G. B. Müller (Eds.), *Modeling biology. Structure, behaviors, evolution* (pp. 15-35). Cambridge, MA: The MIT Press.

Levins, R. (1966). The strategy of model building in population biology. *American Scienctist*, 54, 421-431.

Levins, R. (2006). Strategies of abstraction. *Biolology and Philosophy*, 21, 741-755.

Levy, A. & Bechtel (forthcoming), Abstraction and the Organization of Mechanisms.

Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*, 8, 803-813.

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Science*, 100(21) 11980.

Mazurie, A., Bottani, S., & Vergassola, M. (2005). An evolutionary and functional assessment of regulatory network motifs. *Genome Biology*, 6, 35.1-35.12.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594), pp. 824-827.

Morgan, M. S., & Morrison, M. (1999). *Models as mediators: Perspectives on natural and social science*. Cambridge: Cambridge University Press.

Nersessian, N. (2002a). Maxwell and "the Method of Physical Analogy": Model-based reasoning, generic abstraction, and conceptual change. In D. Malament (Ed.), *Essays in the History and Philosophy of Science and Mathematics* (pp. 129–166). Lasalle, Il: Open Court.

Nersessian, N. (2002b). Model-Based Reasoning in Conceptual Change, in Magnani, L., Nersessian N. and Thagard, P. (Eds.), *Model-Based Reasoning in Scientific Discovery* (pp. 5-22). New York: Kluwer Academic/Plenum Publishers.

O'Malley, M. (2007). Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case. *History and Philosophy of the Life Sciences*, 29(3), 337-360.

O'Malley, M., & Soyer, O. (2011). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, Vol 43, Issue 1, pp. 55-68.

Rheinberger, H. (1997). *Towards a history of epistemic things. Synthesizing proteins in the test tube*, Stanford: Stanford University Press.

Rheinberger, H. (2007). Experimental model systems: An epistemological apercu from the perspective of molecular biology. In M. D. Laubichler, & G. B. Müller (Eds.), *Modeling biology. Structure, behaviors, evolution* (pp. 37-46). Cambridge, MA: The MIT Press.

Rheinberger, H. (2009b). Recent science and its exploration: The case of molecular biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(1), 6-12.

Rheinberger, H. (2010). *An epistemology of the concrete. Twentieth-century histories of life*. Durham: Duke University Press.

Shen-Orr, S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, 31(1) 64.

Solé, R. V., & Valverde, S. (2006). Are network motifs the spandrels of cellular complexity? *TRENDS in Ecology and Evolution*, 21(8), 419-422.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. *Philosophy of Science*, 64, 65-74.

Weisberg, M. M. (2006). Robustness analysis. *Philosophy of Science*, 73(5), 730-742.

Wimsatt, W. (1981). Robustness, reliability, and overdetermination. In M. Brewer, & A. Hoffman (Eds.), *Scientific inquiry and the social sciences* (pp. 124-163). San Fransisco: Jossey Bass.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Winther, R. G. (2006). Parts and theories in compositional biology. *Biology & Philosophy*, 21(4), 471-499.