Modeling and experimenting:

The combinatorial strategy in synthetic biology

Tarja Knuuttila
University of Helsinki
Andrea Loettgers
California Institute of Technology

In which respects do modeling and experimenting resemble or differ from each other? We explore this question through studying in detail the combinatorial strategy in synthetic biology whereby scientists triangulate experimentation on model organisms, mathematical modeling, and synthetic modeling. We argue that this combinatorial strategy is due to the characteristic constraints of the three epistemic activities. Moreover, our case study shows that in some cases materiality clearly matters, in fact it provides the very rationale of synthetic modeling. We will show how the materialities of the different kinds of models – biological components versus mathematical symbols – in combination with their different structures – the complexity of biological organisms versus the isolated network structure and its mathematical dynamics - define the spectrum of epistemic possibilities in synthetic biology. Furthermore, our case shows that from the perspective of scientific practice the question of whether or not simulations are like or unlike experiments is often beside the point, since they are used to accomplish different kinds of things.

1. Introduction

In philosophical discussion models have been located between theories and experiments, often as some sort of go-betweens facilitating the points of contact between the two. Although the relationship between models and theories may seem closer than the one between models and experimentation – and has also traditionally been conceived so – there is a growing body of literature that focuses on the similarities and differences between modeling and experimentation. The central questions of this discussion have concerned the features shared by modeling and experimenting as well as the ways in which the inferences licensed by them are

1

justified. While this discussion has certainly proven enlightening in many respects, we think that the focus on the epistemological, methodological or other similarities and dissimilarities between models and experiments has partly served to conceal the different roles they play in scientific inquiry. Although it is not always so easy to tell modeling and experimentation apart, they can still be conceived as different kinds of epistemic activities that occupy different places in actual scientific practice. Moreover, they are typically used to inform each other in cases where the results of both activities are available.

In this paper we will study the different stands taken in the discussion on the modeling and experimentation through examining the modeling practice in the circadian clock research in synthetic biology. The circadian clock research studies the day and night rhythms of organisms. In this area experiments on model organisms, mathematical models (and their simulations), and synthetic models are being closely triangulated. Apart from that the circadian clock research provides also an interesting sight to study the interplay and characteristics of the three epistemic practices since there is no division of labor in synthetic biology: The same scientists typically engage in experimentation as well as in mathematical and synthetic modeling. Consequently, one might expect that there are good reasons for why synthetic biologists proceed in such a combinatorial manner. We will argue that this is due both to the characteristic *constraints* of each of the aforementioned activities, and to the uncertainty concerning the underlying theoretical framework.

Targeting the constraints of the three different epistemic activities, experimentation on model organisms and mathematical and synthetic modeling, provides also a novel perspective to the materiality issue. We will show how the respective constraints of the modeling approach and the experimentation gave rise to a new type of model, a synthetic model that was built from genetic material using a mathematical model as a blueprint. As an example of such a synthetic model we

will examine closer the *Repressilator*, which was one of the first models of its kind. [1] In the case of the *Repressilator* the "same stuff" was clearly relevant for the epistemic results sought for, but only in combination with a carefully engineered mechanism.

2. Modeling vs. experimentation

The discussion on the relationship of modeling and experimentation has concentrated on which grounds, if any, can the two activities be clearly distinguished from each other, and whether the inferences back to real target systems are more direct in the case of experimentation than in modeling. As regards the similarity between modeling and experimentation, there are many ways to cash it out. One way is to consider them as largely analogous operations. The idea is that both in modeling and in experimentation one aims to seal off the influence of other causal factors in order to study how a causal factor operates on its own. Whereas in experimentation this sealing off happens through experimental controls, modelers use various techniques, such as abstraction, idealization and omission as vehicles of isolation (see e.g. Cartwright 1999, Mäki 2005). Consequently, a theoretical model can be considered as an outcome of the *method of isolation*, in which a set of elements is theoretically removed from the influence of other elements through the use of a variety of unrealistic assumptions (Mäki 1992).

Although the idea of seeing both experimentation and modeling as instances of the method of isolation seems intuitively plausible, one specific property of mathematical modeling does not easily fit this view. Idealizing assumptions are often driven by the requirements of tractability rather than by those of isolation. The model assumptions do not merely neutralize the effect of the other causal factors but rather construct the modeled situation in such a way that that it can be

---

[1] Our study of the *Repressilator* is based on an extensive lab study which one of us conducted following and analyzing the modeling practice in a synthetic biology laboratory at the California Institute of Technology (Loettgers 2009).

conveniently mathematically modeled making it often unclear which assumptions are crucial for the results, or whether the results are dependent on the model as a whole (i.e. on the specific mathematical construction of the model and the abstractions used, see e.g. Cartwright 1999, Morrison 2008, Knuuttila 2009). This feature of mathematical models is further enhanced by their use of cross-disciplinary computational templates that are in the modeling process adjusted to fit the field of application (for computational templates, see Humphreys 2002, 2004, Knuuttila and Loettgers forthcoming). Such templates are often transferred from other disciplines, as is also the case in the circadian clock research, where many models, formal methods and related concepts originate from physics and engineering (e.g. the concepts of oscillator, non-conservative system, feedback mechanism, noise – see below). There are also reasons to doubt whether the method of isolation really describes the actual modeling heuristic since often models seem to depict hypothetical systems rather than being abstracted from real world systems (Weisberg 2007).

Even if one does not want to ascribe to the idea that models are results of isolation, this insight implies another important respect in which modeling and experimentation might resemble another. In experimentation some real world system is intervened on. Likewise in modeling one intervenes on the model system, although in this context one usually talks about the manipulation of a model (Morgan 2003). If one characterizes experiments on the basis of an intervention on some specific system without requiring that the system in question would be the real world target system scientists are primarily interested in, it is easy to see how also models can qualify as experiments. Writing about simulations Parker (2009) makes this move, although she specifies that it is the *computer simulation studies* that meet the criteria of experiments.[2] In computer simulation studies the concrete system intervened on is the programmed digital computer. In her analysis, moreover, targets multiply: the target can as well be the real world

---

[2] Parker defines simulations as a "type of representation" – "a time-ordered sequence of states" (2009). Consequently, also experiments may serve as simulations. This definition leaves out one important class of simulations, Monte Carlo simulations, as Winsberg (2009) points out.

material/physical system (for which also the computer qualifies) or the mathematical system specified by the theoretical model equations or their off-spring, the programmed equations.

As regards the programmed equations, the recent discussants have stressed that their relationship to the basic theoretical model underlying the simulation model is anything but straightforward (e.g. Winsberg 2003, Lenhard 2007, Barberousse et al. 2009, Parker 2009). The differential equations of the theoretical model are discretized which involves various processes of manipulation, simplification and alteration involving also "tricks of trade". Lenhard (2007) discusses one such trick, "Arakawa's computational trick" in atmospheric modeling. Arakawa (1966) dispensed the search for any "true" solution to the primitive theoretical equations using instead the so-called Arakawa operator, which involved some problematic assumptions that could not be justified theoretically. Importantly, he artificially limited the growth of the instability of the modeled atmosphere by assuming that the kinetic energy is preserved, which is not the case since kinetic energy is transformed into heat by friction. Although this trick was initially criticized on these grounds it nevertheless became later accepted as it successfully imitated the flow patterns in atmosphere. This shows, according to Lenhard, that the "explorative cooperation" between experimenting and modeling is a basic characteristic of simulation: the model adjustment and tuning is driven by the comparison of its consequences and properties directly to the data or phenomena in a repeated recursive process.

Of course, not all simulation models involve the discretization of an original intractable differential model, although this is the case that has acquired most attention by the philosophers of science so far. Fox Keller (2003) argues that such simulation methods as cellular automata, that involve the construction of models of phenomena for which no general theory yet exists, are simulations par excellence. They create "artificial universes" giving rise to a new scientific spirit aimed at producing recognizable patterns and interesting behavior in these alternative "realities". Note that that this kind of scientific activity draws modeling nearer to experimentation in that the system of interest (or target system) becomes more

clearly the artificial universe itself. Thus the simulation model functions less like a representation of the behavior of some external target. This sort of simulation modeling makes even more apparent the peculiar kind of scientific understanding that is gained by simulation modeling. Lenhard (2006) suggests that simulation methods produce "understanding by control" which is geared towards design rules and predictions. Whereas in traditional modeling one manipulates highly idealized model systems, in simulation one gains understanding through experimenting with more epistemically opaque model. This provides an interesting analogy between simulations and experiments on model organisms: In both of them one tries through control to overcome the complexity barrier. However, living organisms is radically more complex than simulation models. Synthetic modeling, as we will argue below, provides another strategy deal with the complex systems.

Already on the grounds mentioned above it is not difficult to see why especially simulation modeling has been likened to experimentation. Further features of simulation models that seem crucial from this perspective are related to the issues of data and materiality. One motivation for claiming that the practices of modeling and experimentation are similar to each other invokes the fact that both simulationists and experimentalists produce data and are dealing with data analysis and error management. Winsberg (2003) suggests that the techniques simulationists use to augment their belief in their results resemble those presented by Franklin (1986) in the context of experimentation. The kinds of data produced by simulationists and experimentalists are obviously different (see Barberousse et al. 2009), and further study is needed to see to whether the similarities between the practices of the simulationists and the experimentalists are more than superficial in this respect.

Some appeals to the similarity of simulations to experiments are related to their physicality or materiality. Obviously, analog simulations involving for example the use of scale models or other specially designed physical artifacts qualify as material things. What about computer simulations – are they non-material things as Morgan (2003) suggests? Most philosophers of science seem to agree that the fact that computer simulations are implemented on a material concrete device and involve

thus physical processes when run on it, gives them a material status (see e.g. Humphreys 1994, Hughes 1999, Barberousse et al. 2009, Parker 2009). However, the opinion is divided as to what epistemic role does this materiality play. More specifically, what role does the materiality play in simulations in justifying the inferences concerning the target systems?

The answer given by Norton and Suppe (2001) seems most straightforward. They take it for granted that the materiality of simulations likens them to experimentation. The challenge is then to show, how this can be the case given that computer simulations and their real world target systems are very different kinds of things. According Norton and Suppe simulations "embed" data models in programmed computers through "lumped models": A simulation model is a lumped model embedded into a programmed computer. Since programmed computers are real systems, observations of them can give "new knowledge of the world" precisely as experiments do. However, for these observations to give knowledge of the real world target systems the simulation is supposedly mimicking, some rather stringent mapping relations have to hold between the model of the data/ the modeled physical system itself, the lumped model, and the programmed computer. The question is whether these mappings really hold in practice. Barberousse et al. (2009) argue that in most simulations the computer's physical states cannot be said to realize the lumped model, that is, the lumped model and the programmed computer qua physical system need not share common structure. This seems to pull the rug from under the idea that "the computation's being a physical process itself should explain the 'mimicry' relation "between simulation and its real world target" (ibid., 566).

Indeed, the right kind of materiality has been claimed to be the distinguishing mark of experiments and even the reason for their epistemic superiority to simulations. Either it has been claimed that computer simulations are non-material (Morgan 2003), or that the relationship between a simulation and its target is abstract, whereas the relationship between an experimental system and its target is grounded in the same material being governed by same kinds of causes (Guala 2002). The crucial difference between modeling and experimentation, according to

this view, is that whereas in simulation one experiments with a representation of the target system, in experimentation the experimental and target systems are made of the "same stuff". This difference also explains, according to Morgan and Guala, why experiments have more epistemic leverage than simulations. Morgan points out, for instance, that the anomalous experimental findings are more likely to incur change in our theoretical commitments than unexpected model results.

Despite the intuitive appeal of the importance of the "same" materiality, it has been contested on different grounds. Morrison (2009) points out that even in the experimental contexts the causal connection with the physical systems of interest is often established via models. Consequently, according to her, materiality provides no unequivocal epistemic standard that distinguishes simulation outputs from experimental results. Parker (2009) attacks directly the alleged significance of the "same stuff". She interprets the "same stuff" meaning for instance the same fluid and points out that in traditional laboratory experiments on fluid phenomena many other things such as the depth of the fluid and the size, shape, roughness and the movement of any container holding it may matter. This leads her to claim that it is the "relevant similarities" that matter for the justified inferences about the external phenomena.

We think that Parker takes Guala and Morgan too literally, since their conception of the "same stuff" includes, apart from the same material literally understood, also same forms and causal make-up (e.g. Morgan 2005, 32; Guala 2002). Parker grants, though, that in cases in which "scientists as yet know very little about a target system their best strategy may well be to experiment on a system made of same stuff"" because it can be expected to be similar to the target system in relevant respects (2009, 494). Somehow, then, her argument seems to rest on the distinction between discovery and justification. It is as if the ways in which we came to know something had no bearing on how this knowledge is justified. More problematically, still, Parker's account seem to fall prey to the perspective of finished science. How often do we know the relevant similarities, and how do we know them? In what follows we will study this question through the practice of synthetic modeling, which is precisely based on the idea that the right

kind of materiality (covering also the causal interactions) matter. The case also shows that although modeling and experimentation share some common features, from the perspective of scientific practice they nevertheless are considered as separate practices with different constraints. This in turn explains why they are often, in cases in which this is possible, triangulated with each other.

3. Combinatorial strategy in synthetic biology

Synthetic biology is a relatively novel and highly interdisciplinary field. It is located at the interface of engineering, physics, biology, chemistry and mathematics. The research practice in this field is a combination of methods, concepts, tools, and theories from those fields. In synthetic biology itself one can distinguish between two different main research branches: an engineering oriented branch focusing on the engineering of novel biological components/ systems and a basic science oriented branch using synthetic models to gain understanding on the basic design principles underlying specific biological functions, such as the circadian clock regulating day and night rhythms in organisms.

In our study we are focusing on the later branch of synthetic biology. One of the defining strategies of this approach is the combinatorial use of mathematical models, model organisms, and synthetic models. The upper part of a diagram (Fig. 1), taken from a review by Sprinzak and Elowitz (2005) on synthetic biology, shows the basic idea of this combinatorial modeling strategy. The two authors call this approach "the synthetic biology paradigm". As the diagram suggests, the results gained from each of the three different epistemic activities inform the other ones. From the point of view of a researcher this is a very demanding task. She has to have the skills and experience to use mathematical models, to perform experiments in molecular biology lab and to know the techniques of genetic engineering. Consequently, the researcher herself has to possess a high degree of interdisciplinary knowledge and skills.
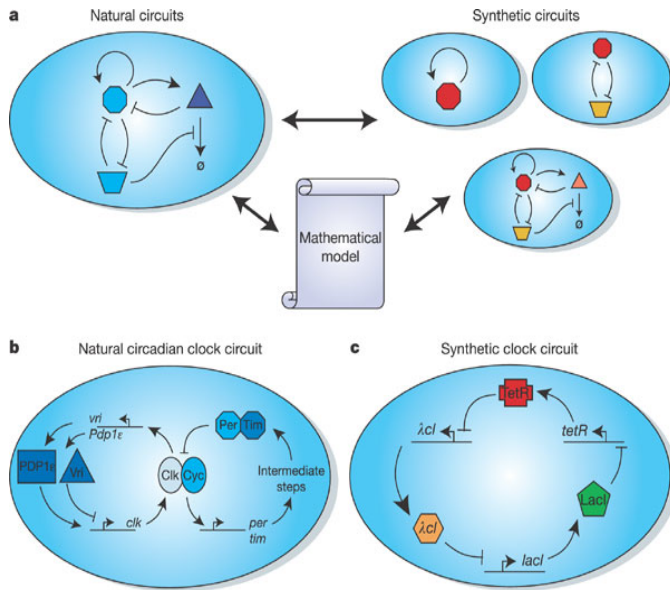
Figure 1. The diagram is taken from a review article by Elowitz and Sprinzak (Elowitz & Sprinzak 2005). The upper part of the diagram depicts the combinatorial modeling strategy, which the two authors call the "synthetic biology paradigm". The lower part compares a natural gene regulatory network and a synthetic one.

Why do researches make use of a combinatorial modeling strategy in studying the organizational principles in biology? A first clue can be found from the lower part (b) of the diagram. The left hand side of the diagram shows the "natural gene regulatory circuit" of the circadian clock in *Drosophila* consisting out of interacting genes and proteins and the right hand side a synthetic model of the circadian clock, the *Repressilator*, introduced by Elowitz and Leibler (2000). The diagram shows that the two main differences between the natural and the synthetic system are:

1. The natural system exhibits a much higher degree of complexity than the synthetic system.
2. The synthetic circuit has been designed by using different genes and proteins.

Consequently, synthetic models have the advantage of being less complex than model organisms. On the hand, in comparison with mathematical models they are of the same materiality as biological systems (although the *Repressilator* was

constructed from different genetic material than the naturally occurring circadian clocks, we will return to this below). This fact of being of the same materiality as natural systems is crucial for the epistemic value of synthetic modeling. It means that synthetic models are expected to work in the same way as biological systems. The synthetic strategy rose as a response to the constraints of the mathematical modeling and the experimentation with model organisms. There seemed to remain a gap between what could be found experimentally and studied mathematically: the experiments could not give any conclusive answer for whether the hypothetical mechanisms that were suggested by mathematical models were in fact realized by biological systems. In what follows we will discuss in more detail the specific constraints of each of the three epistemic activities.

3.1 Mathematical modeling

One of the most influential mathematical models of circadian rhythms was introduced by Brain Goodwin in his book *Temporal Organization in Cells* (1966). This book is an example of an attempt to apply concepts from engineering and physics to biology. Inspired by Jacob and Monod's (1961) operon model of gene regulation Goodwin explored the mechanism underlying the temporal organization in biological systems such as circadian rhythms, in terms of a negative feedback mechanism. Another source of inspiration for him was the work of the physicist Edward Kerner (1957). Kerner had tried to formulate a statistical mechanics for the Lotka-Volterra model, which prompted Goodwin also to attempt to introduce statistical mechanics for biological systems. These aims presented partly competing constraints on the design of Goodwin's model, which both shaped its actual formulation and the way it was supposed to be interpreted. A third constraint was due to the limitations of the mathematical tools for dealing with non-linear dynamics. These three constraints are of different character. The first constraint was largely *conceptual* but it had mathematical implications: the idea of a negative feedback mechanism borrowed from engineering provided the conceptual framework for Goodwin's work. It guided his conception of the possible mechanism

and its mathematical form. The second constraint of trying to formulate a general theory in analogy to physics was based on the assumption that biological systems should follow the basic laws of statistical mechanics. An extended theory of statistical mechanics including biological systems and their particularities as open systems was supposed to provide a general theoretical framework against which the model could be tested. This particular constraint could be called *fundamental* theory constraint. The third set of constraints was *mathematical* in character being due to the limitations of the mathematical methods, equations, and computational templates used in the formulation of the model.

The three types of constraints, conceptual, fundamental and mathematical, are of different character but what they have in common is that they belong to the toolbox available to the researcher for describing and modeling a specific system. But as a hammer comes to its limits when it is used as a screwdriver, the engineering concepts and the ideal of statistical mechanical explanation proved problematical when applied to biological systems. In the context of systems and synthetic biology the aim for a statistical mechanics for biological systems was later on replaced by the search for possible design principles of biological systems. The concept of the negative feedback mechanism was preserved, functioning as a cornerstone of subsequent research, but there was still some uneasiness about it that eventually lead to the construction of synthetic models (see below).

The tool related constraints are to be distinguished from the constraints directly related to the complexity of biological systems, which plays different roles in the three epistemic activities. Moreover, it should be noted that the tool related constraints are not static. Concepts can change their meaning especially in interdisciplinary research contexts such as synthetic biology, and new mathematical tools and computational templates can be developed or introduced, etc. The adjustments and changes in the available modeling tools form an important part of the development of the computational sciences. Furthermore, the constraints are interdependent and the task of the modeler consists in finding the right balance between them. We will exemplify these points by taking a closer look into the Goodwin model.

The basic structure of the network underlying the molecular mechanism of Goodwin's model of temporal organization is represented in the following diagram:
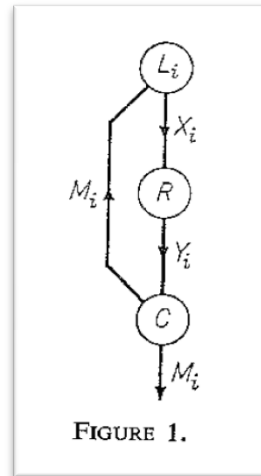

FIGURE 1.

Figure 2. The circuit diagram of the mechanism underlying the Goodwin model (1966, 6).

The main structure of the model is a negative feedback loop. It consists of a genetic locus L_i, synthesizing $mRNA$ in quantities represented by the variable X_i. The $mRNA$ leaves the nucleus and enters the ribosome, which reads out the information from the $mRNA$ and synthesizes proteins in quantities denoted by Y_i. The proteins are connected to metabolic processes. At the cellular locus C the proteins influence a metabolic state by for example enzyme action, which results in the production of metabolic species in quantity M_i. A fraction of the metabolic species is traveling back to the genetic locus L_i where it functions as a repressor.

This mechanism leads to oscillations in the protein level Y_i regulating temporal processes in the cell, such as circadian rhythm. Goodwin described the mechanism by a set of differential equations, which were due to the feedback mechanism of a non-linear character. Such systems display complicated behavior and in general no analytical solutions exist for them. Goodwin explored the dynamics described by the set of coupled non-linear differential equations by making simplifying assumptions through which he attempted to deal with two important constraints: the complexity

of the system consisting of a variety of biochemical components and processes, and the complexity given by the non-linear dynamics of the assumed mechanism. Firstly, he had to leave aside many known biochemical features of the circadian clock mechanism and, secondly, he had to make assumptions that would allow him to simplify the mathematical representation in such a way that he could use numerical methods in exploring its complicated dynamic behavior without having to solve the non-linear coupled differential equations.

Goodwin was able to show by performing very basic computer simulations that the change in the concentration of proteins Y_i and the concentration of mRNA X_i form a closed trajectory. This means that the model system is able to perform regular oscillations, such as circadian rhythms. But the oscillations were not robust. Goodwin wrote: "The oscillations which have been demonstrated to occur in the dynamic system [...] persist only because of the absence of damping terms. This is characteristic of the behavior of conservative (integrable) systems, and it is associated with what has been called weak stability." (Goodwin 1966, 53) He went on explaining that a limit-cycle dynamics would have been the desirable dynamic behavior. In this case after small disturbances the system moves back to its original trajectory. This is a characteristic of non-conservative systems. Biological systems are non-conservative systems: They are not closed systems because they exchange energy with the environment. But in order to formulate a statistical mechanics Goodwin had to concentrate on conservative systems. He wrote about this point: "What our present treatment does, in effect, is to substitute for real structure a model whose dynamic behaviour approximates to that which we suspect occurs in the cell. The approximation will be best in the neighbourhood of trajectories which are close to the limit cycles which we expect to occur in cell variables." (Goodwin 1966, 53). In order to simultaneously fulfill the aim of modeling a mechanism producing oscillatory behavior regulating the temporal organization such as circadian rhythm *and* the aim of formulating a statistical mechanics for biological systems, Goodwin had to balance different constraints ending up with a model which he believed could at best approximate the behavior he suspected to happen.

Even though Goodwin's model was not a complete success, it nevertheless provided a basic template of circadian rhythm on which following researchers based their modeling endeavors. In those modeling efforts the aim of formulating a statistical mechanics for biological systems was left behind. The feedback mechanism of circadian rhythm took center place leaving as the main constraints of the modeling approach the complexity related to the number of components and biochemical processes in natural systems, and the mathematical difficulties related to the non-linear coupled differential equations. It should be underlined that these problems were intertwined. As we saw already in the case of Goodwin, the available mathematical methods constrained the knowledge that was taken into account in modeling. However, the models constructed became more detailed as the first experiments exploring the molecular basis of the circadian rhythms became possible in the mid 1970's. These experiments, as we will show below, came with their very own constraints. While modelers were able to partly bracket the complexity of biological systems by simply assuming most of it away – though they had to deal with it nevertheless in mathematical terms – this was not possible in the experimental work.

3.2. Experimentation

The first circadian clock gene was discovered in experiments performed by Ron Konopka and Seymour Benzer (Konopka and Benzer 1971) in the beginning of the 1970's. They named the gene *period* (*per*). The experimental research on circadian rhythms in molecular biology and genetics progressed only slowly after Konopka and Benzer published their results. The reason was that the experiments were rather elaborate. A great challenge consisted in finding the protein related to the isolated gene, its function in the biochemical processes, and deciding whether the protein was part of the mechanism or an output of it regulating other functions in biological system. Only in the mid 1980's and 1990's with further advances in molecular biology and genetics did the circadian clock research start booming and more genes, proteins and possible mechanisms were discovered in experiments on

model organisms such as *Drosophila, Neurospora* and *Arabidopsis.* The complexity the scientists were facing already in these "simple" model organisms was one of the main constraints in performing experiments on model organisms. They had to deal with the *dynamic complexity* due to what in modeling terms is called the non-linear dynamics of the feedback loops and the *constitutive complexity* of the cell environment in which the assumed mechanisms were embedded (for the distinction, see Mitchell 2003). In contrast to modeling, in the experiments the scientists had to directly face the messiness due to these different forms of complexity and they also had different tools to handle them – which in turn brought with them their own constraints.

In the following we discuss the example of the discovery of the second interlocked feedback loop in the circadian clock of *Drosophila* by Paul Hardin and his colleagues (Glossop et al. 1999). At that time 5 genes of the circadian clock had been identified: *Period* (*per*), *timeless* (*tim*), *Drosophila Clock* (*dClk*), *Cycle* (*Cyc*) and *double time (dbt).* Three of these genes are rhythmically expressed: *per* mRNA and *tim* mRNA levels peak early in the evening ZT 13-16 although the associated proteins PER and TIM do not peak until late evening because of phosphorylation processes in the cytoplasm. *dclk* mRNA peak late at night to early in the morning, from ZT 23 to ZT4. ZT. i.e. the zeitgeber time, provides the environmental cues that each day (partially) reset the rhythms. In the present case light was used to entrain the rhythm of the circadian clock of *Drosophila*. At ZT 0 lights are turned on and at ZT 12 is turned off.

It was known that *dClk* played an important role in the activation of the transcription of *per* and *tim.* Little was known about the regulation of the *dClk* cycle. To find out about the regulation of *dClk* Hardin et al. performed mutation experiments. It was known that the levels of *dClk* mRNA are low in mutants per^{01} and tim^{01} lacking functioning PER and TIM proteins. This observation led to the assumption that PER and TIM activate *dClk* transcription in addition to their roles as transcriptional repressors. Three models had been introduced by Isaac Edery et al. (1994), which aimed to explain the PER-TIM dependent activation of *dClk*. In the

first two models (Figures 3a-3b) PER and TIM promote the *dClk* transcription by bringing in transcriptional activators into the nucleus or by co-activating a transcriptional complex. In the third model (Figure 3c) PER or TIM or both inhibit the activation of a transcriptional repressor.

Hardin et al. systematically explored the three models in order to distinguish between them. Based on the idea that dCLK together with CYC is necessary for the activation of *per* and *tim*, they created a clock gene mutant $dClk^{JrK}$, which led to a non-functional dCLK protein, and measured the level of *dClk* mRNA. They expected low levels of *dClk* mRNA because the concentration of PER and TIM proteins would be low.
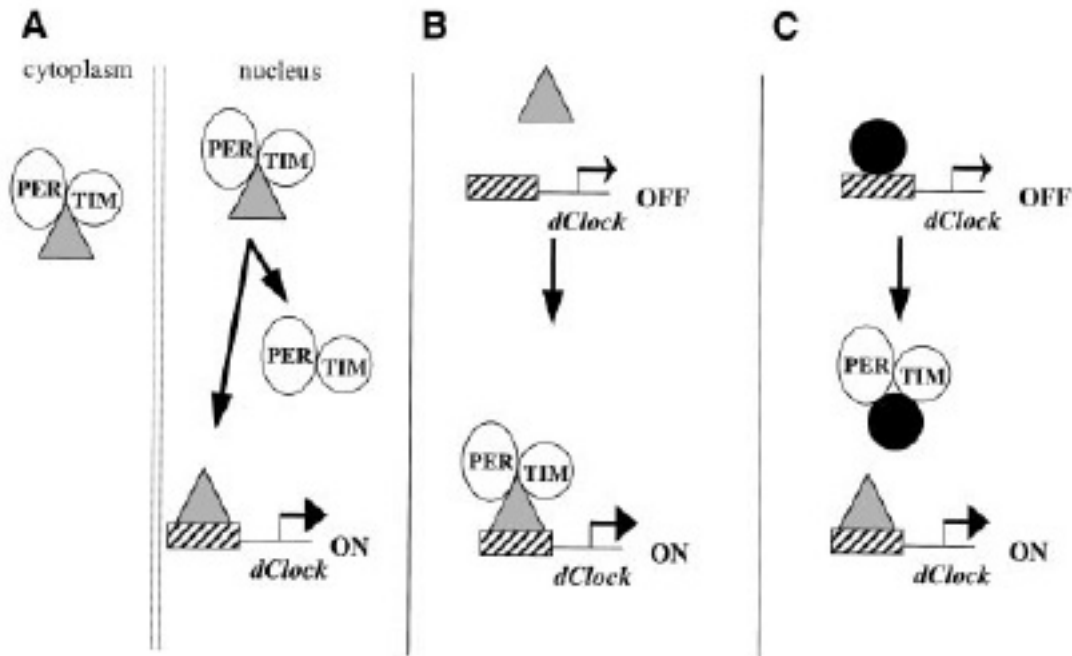

Figure 3. The three different models of *dClk* activation.

Instead they got the surprising result that the level of *dClk* mRNA of the model organisms is indistinguishable from the wild type. The only difference to the wild type was that the level of *dClk* mRNA at both times ZT1 and ZT13 was almost the same, which means a non-functional circadian clock. These results ruled out that

dClk activation is PER-TIM dependent. In further experiments, Hardin et al. tested the following double mutations $per^{01}dClk^{JrK}$ and $per^{01}Cyc^{0}$. In both cases the level of *dClk* mRNA was comparable with the level of *dClk* mRNA wild type. Only in the case of the single mutant per^{01} the level of *dClk* mRNA was low. This observation, which in the first place led to the assumption of a PER-TIM dependent activation of *dClk* and which had been ruled out by the series of experiments, had now to be explained by a different model. In this model (Fig.4), the binding of PER-TIM dimers to the dCLK-CYC dimer releases dCLK-CYC dependent repression of dCLK, thus enabling *dClk* transcription. How does this mechanism explain the observed results in the experiments? Only in the case where dCLK and CYC are present, as in the case of period gene mutant per^{01} *dClk* is repressed, and because no PER-TIM dimers are available to release the dCLK-CYC dimers, *dClk* remains repressed. In the case of the double mutants $per^{01}dClk^{JrK}$ and $per^{01}Cyc^{0}$ the absent of an active dCLK/CYC prevents the repression of *dClk* so *dClk* mRNA is produced even if it leads to a non-functional protein, as in the case of the double mutant $per^{01}dClk^{JrK}$ and $per^{01}Cyc^{0}$.
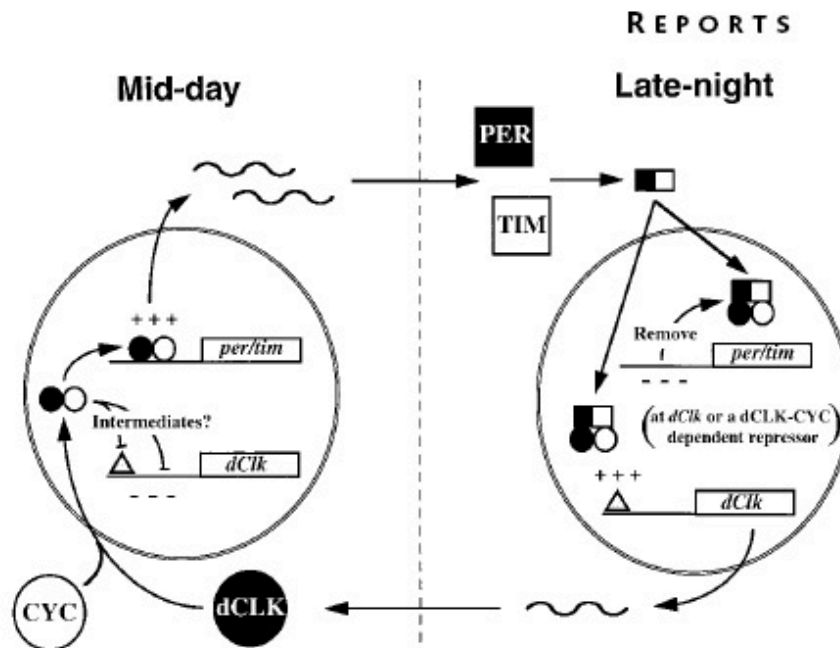

Figure 4. The interlocked feedback loop discovered by Hardin et al.

As a result Hardin et al. suggested that the mechanism underlying the circadian rhythm consisted of two interlocked feedbacks of the following form: The first loop, the so-called *per/tim* loop, is activated by dCLK-CYC and repressed by PER-TIM. The second loop, the *dclk* loop, is repressed by dCLK-CYC and de-repressed by PER-TIM.

This short discussion of how the second feedback mechanism in the circadian clock in *Drosophila* was found gives some insights into the strategies and methods used by molecular biologists in the isolation of biochemical mechanisms. The mutation experiments as they were used here can be best characterized as a variation of parameters by which different possible mechanisms are tried out. The constraints related to this particular strategy are of technical nature related to the available methods from molecular biology and genetics such as creating the mutations and measuring the effects of the mutations. But even though in the case of Hardin et al. it was possible to find a mechanism or parts of a mechanism, it always remains the haunting question whether all the components of the supposed mechanism have been found i.e. whether the network is complete, or are some components and/or interactions missing. Or, whether the hypothetical mechanism is the only one by which the observations could be explained – the traditional problem of underdetermination. Or, going a step further, whether the mechanism works in isolation, and if not, how is it interacting with the rest of the cell environment, and what role stochastic fluctuations ("noise") play in the dynamics of the network.

3.3. Synthetic modeling

Although mathematical modeling and their simulations, and experimentation informed each other in circadian clock research, the mathematical modeling suggesting possible mechanism templates and experimentation probing them and providing more biochemical detail, there remained a gap between them. The modeling effort was based on rather schematic templates and related concepts

originating often from other fields of inquiry. It was unclear whether biological organisms really functioned in the way suggested by them i.e. whether they could be realized by biological organisms. Synthetic strategy suggests a way to deal with this problem due to the specific construction of synthetic models. They are mixed, double-faced entities: On the one hand, they are of the same materiality as model organisms since they are made of biological material, such as genes and proteins. On the other hand, they differ from model organisms in that they are not results of any evolutionary process, being instead designed on the basis of mathematical models. In the construction of a synthetic model the mathematical model is used as a blueprint: it specifies the structure and the dynamics giving rise to particular behaviors or functions. Thus the synthetic model has its origin in the mathematical model, but it is not bound by the same constraints: The model is constructed from the "same stuff" (Morgan 2003) as the biological genetic regulatory networks and it even works in the cell environment. Consequently, even if it is not understood in all its details, it provides a kind of simulation device of the same natural kind as its real world counterparts.

3.3.1 The *Repressilator*

The *Repressilator* provides a compelling example of a synthetic model. It is an oscillatory genetic network, which was introduced in 2000 by Michael Elowitz and Stanislas Leibler (2000). The first step in constructing the *Repressilator* consisted in the design of a mathematical model, which was used to explore the known basic biochemical parameters and their interactions. Having constructed a mathematical model of a genetic regulatory network Elowitz and Leibler performed computer simulations on the basis of the model. They showed that there were two possible types of solutions: "The system may converge toward a stable steady state, or the steady state may become unstable, leading to sustained limit-cycle oscillations" (Elowitz and Leibler 2000, 336). Furthermore, the numerical analysis of the model gave insights into the experimental parameters relevant for constructing the synthetic model in showing that "[...] oscillations are favoured by strong promoters

coupled to efficient ribosome binding sites, tight transcriptional repression (low 'leakiness'), cooperative repression characteristics, and comparable protein and mRNA decay rates" (ibid, 336). The later point helped in choosing the three genes used in the design of the network. Elowitz and Leibler also explored the continuous as well as the stochastic dynamics of the model in order to analyze the role of internal noise in the mechanism. Internal noise in biological systems is caused by the low number of molecules in the cell.

The mathematical model functioned as a blueprint for the engineering of the biological system. The structure of the *Repressilator* is depicted in the following diagram:
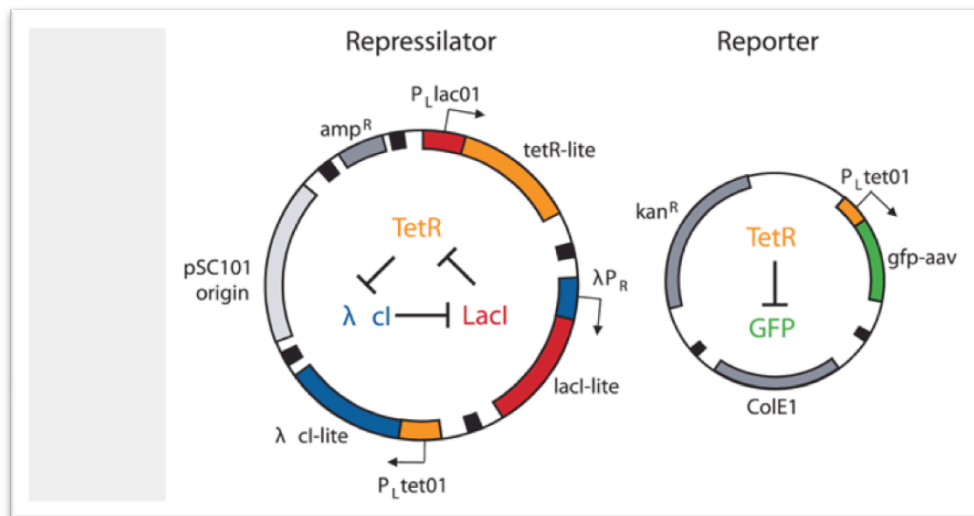

Figure 5. The main components of the Repressilator (left hand side) and the Reporter (right hand side) (Elowitz and Leibler 2000, p. 336).

In the diagram the synthetic genetic regulatory network, the *Repressilator,* is shown on the left hand side and it consists of two parts. The outer part is an illustration of the plasmid constructed by Elowitz and Leibler. The plasmid is an extra-chromosomal DNA molecule integrating the three genes of the *Repressilator*. Plasmids occur naturally in bacteria. In the state of competence, bacteria are able to take up extra chromosomal DNA from the environment. In the case of the

Repressilator, this property allowed the integration of the specific designed plasmid into E.coli bacteria. The inner part of the illustration represents the dynamics between the three genes, $TetR$, $LacI$, and λcl. The three genes are connected by negative feedback loops. The left hand side of the diagram shows the *Reporter* consisting of a gene expressing a green florescent protein (GFP), which is fused to one of the three genes of the *Repressilator*. The GFP oscillations in the protein level made visible the behavior of transformed cells allowing researchers to study them over time by using fluorescence microscopy.

The construction of the *Repressilator* was enabled by the development of new methods and technologies, such as the construction of plasmids and Polymerase Chain Reactions (PCR). As the discussion on the *Repressilator* shows, with the formation of synthetic biology a novel tool was introduced into the research on the organizational principles of biological systems: the possibility of constructing novel engineered genetic networks specially designed for answering certain kinds of theoretical questions. Their construction has so far been limited to simple networks such as the *Repressilator* and the construction components are chosen in view of what would be optimal for the behavior under study. This means that such networks need not be part of any naturally occurring system. For example the genes used in the *Repressilator* do not occur in such a combination in any known biological system but are chosen and tuned on the basis of the simulations of the underlying mathematical model and other background knowledge in such a way that the resulting mechanism would allow for (stable) oscillations. These technical constraints imply a constraint on what can be explored by such synthetic models: possible design principles of biological systems. The search for the possible design principles has replaced in systems and synthetic biology the aim of formulating a fundamental theory for biological systems such as statistical mechanics.

In conclusion, we have suggested that the way mathematical models, model organisms, and synthetic models are used in the study of the temporal organization in cells shows that the three epistemic activities are constrained both by the complexity of biological systems, and their characteristics tools, which constraints are intertwined in the actual scientific practice in different ways.

4. How and why materiality matters in synthetic biology

Coming back to the philosophical discussion on modeling, let us first note that the mathematical models of the circadian clock are difficult to consider as experiments in the sense of being abstractions of real world systems that permit the study of the effects of isolated causal factors. In the circadian clock research the modeling endeavor started before the experimentation, and the circadian clock phenomena was probed with model templates and concepts that were borrowed from other disciplines and subjects. As already discussed, it is often not possible to decompose a model into its assumptions as the isolation account of modeling suggests. The idealizations, omissions and approximations follow frequently from the model template used. The assumptions involving them are typically linked to certain mathematical abstractions, and there are limited ways in which they can be relaxed or corrected, constrained by the available mathematical methods. In the case of circadian clock oscillations, the oscillatory phenomena had already been studied by physicists for a long time and there are many well-established ways of mathematically creating them. There exists a whole body of literature on this topic such as Steven Strogatz's book *Nonlinear Dynamics and Chaos* (Strogatz 1994). Consequently, the modelers had already a toolbox for modeling oscillations. But the problem was that the alternative models were too general and underdetermined by available data.

Consequently, although modelers were able to produce the kind of phenomena sought for, that is, robust oscillations, the problem was whether the possible network designs proposed by mathematical models were really the ones that work in biological organisms. This problem was aggravated by the fact that the model templates, methods and concepts were not originally devised biological organisms in mind. In fact, a big part of the modelers in the field of systems and synthetic biology have a physics background. Moreover, this problem could not be conclusively settled by experimentation since the work with model organisms had to deal with the immense complexity of even such simple organisms as E.coli. Thus

even though the empirical research was able to find genes and proteins involved in the circadian clock phenomena, the results were still inconclusive as regards the basic mechanism underlying it.

Synthetic models fill (partly) the gap between mathematical modeling and experimentation on model organisms by offering a tool for identifying possible network design principles, and showing whether they might be realizable in biological organisms.[3] Their epistemic functioning is due to their mixed nature: On the one hand they share with mathematical models their tightly constrained nature, which makes it possible study certain theoretical questions in a regimented manner (see Knuuttila and Voutilainen 2003). This shows that constraints do not just limit reasoning, but instead also afford it. On the other hand, they are constructed of the same kinds of components as the natural systems, being implemented, moreover, in the natural cell environment. In the case of the *Repressilator* the natural genetic regulatory network was replaced by the engineered one, which gave more control for the researchers on the system under study than in the case of domesticated natural organisms (such as model organisms). Being of a lower degree of complexity compared to natural systems and at the same time being exposed to the same kind of *biological* constraints as natural biological systems is regarded as an important advantage of synthetic modeling. By reducing the degree of complexity the system becomes more tractable and easier to manipulate. In the case of the *Repressilator* the goal was not to imitate the natural circuit but to find the *sufficient components* and interactions of the supposed mechanism producing a specific behavior, such as oscillations in protein levels. The components of the network were chosen to get the most optimal behavior and not to get as close as possible to a naturally evolved genetic network. This explains why the genetic network of the *Repressilator* is

---

[3] Circadian clock research has been used recently by Bechtel and Abrahamsen (forthcoming a, b) as an exemplary case of mechanistic research that has successfully combined the decompositional approach of finding the basic components of mechanisms to the study of their interactions by modeling. On our account this "recomposition" between experimental results and modeling was far from seamless and synthetic models were partly designed to fill the gap between these two activities (cf. Loettgers 2007).

comprised of a different constellation of components than any naturally occurring network. Moreover, by implementing the synthetic genetic network into a cell it was exposed to the constraints of natural biological systems. Those are in general not known in all their details but knowing those details seem not always necessary as long as they are there. The cell provides a simulation environment, which is of the same "natural kind" as the system under investigation. This is a feature commonly associated with experiments. They are considered to be particularly useful in the contexts in which one has an imperfect understanding of the causal mechanisms at work. The control over the "same", often domesticated or technologically altered, natural kind, is expected to give some handle on the causal mechanism(s) (cf. Guala 2002).

Another feature of experiments displayed by the *Repressilator* is due to the way the researchers received the undesired results. In the case of experimentation anomalous or unexpected results are commonly taken more seriously, whereas if the model does not produce what is expected from it, the modelers usually try to devise a better model. The *Repressilator* in turn sparked a new line of research precisely due to its limited success. In contrast to the mathematical model underlying it, the *Repressilator* did not show the expected behavior: regular oscillations. Instead the oscillations turned out to be noisy. Computer simulations taking into account stochastic fluctuations did show that such fluctuations could be the cause of its noisy behavior. But the noisy behavior could have also been caused by external noise coming from the cell environment. A new line of research emerged exploring the different sources of noise and their effects on biological systems. In the context of this research, and in line with the results gained by studying complex systems in physics and in neuroscience, noise based on stochastic fluctuations gained a functional status: Molecular mechanisms in biological systems make use of internal noise, for example in decision processes such as cell division (see Loettgers 2009).

Consequently, synthetic models such as the *Repressilator* muddle the border between modeling and experimentation even more than simulations do. It can be considered both as a model and an experiment, and it serves to show that in

experimentation the objects of experiments can be located at different distances from the targets they are supposed to give knowledge about. A common misunderstanding concerning experimentation is that in experiments one is intervening the target directly (see e.g. Barberousse et al. 2009). This is certainly not the case even when it comes to the experimentation with model organisms, which can be considered as domesticated natural kinds. However, even if they allow rather direct back inferences to the corresponding wild natural kinds, most often they are used give knowledge of other different natural kinds. They might even be used to give knowledge of the general design principles of biological organisms, as our case shows. Consequently, the object and the target of experiment are not the same and the security of the inferences concerning the targets are partly based on the nature and the relative "length" of the distance between the object of experiment and its target. This distance between the object and the target that also experimentation involves makes apparent the structure of surrogate reasoning that it shares with modeling (e.g. Swoyer 1991, Suárez 2004). Both in modeling and in experimentation one studies surrogates that stand in/for other systems, and scientists use diverse inferential strategies to close the gap between the two. Sometimes the phenomenon is so well known and the theoretical framework so established that we even do not need experiments, simulations will do the job. Other times some highly intricate strategies such as synthetic modeling, are developed which involve the same materiality and the like natural kinds.

References

Arakawa, A. (1966), computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. Part I." J. Computational Physics 1: 119-43 (reprinted J. Comp. Phys. (1997) 135:103-14).

Barberousse, A., Franceschelli, S., and Imbert C (2009) Computer simulations as experiments. Synthese 169: 557-574.

Bechtel, W., Abrahamsen, A. (forthcoming a), Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. Studies in History and Philosophy of Science Part A .

Bechtel, W., Abrahamsen, A. (forthcoming b), Complex biological mechanism: Cyclic, oscillatory, and autonomous. In: Hooker CA (ed) Philosophy of complex systems. Handbook of the philosophy of science, Volume 10. New York: Elsevier.

Cartwright, N. D. (1999), The vanity of rigour in economics: Theoretical models and galilean experiments. (L. L. Economics, Ed.) Discussion paper series 43/99 .

Edery, I., Zwiebel, LJ., Dembinska, ME., Rosbach, M. (1994), Temporal phosphorylation of the Drosophila period protein. Proc. Natl. Acad. Sci. USA, 91(6), 2260-4.

Elowitz, M., Leibler, S. (2000), A synthetic oscillatory network of transcriptional regulators. Nature , 403 (6767), 335-8.

Fox Keller, E. (2003), Models, simulation, and computer experiments. In H. Radder (Ed.), The philosophy of scientific experimentation (pp. 198–216).

Franklin, A. (1986), The neglect of experiment. Cambridge: Cambridge University Press.

Glossop, NJR., Lyons, LC., Hardin, EP. (1999), Interlocked feedback loops within Drosophila circadian oscillations. Science, 286, 766-768.

Goodwin, B. (1966), temporal organization in cells, Academic Press: London and New York.

Guala, F. (2002), Models, simulations, and experiments. In L. Magnani & N. Nersessian (Eds.), Model-based reasoning: Science, technology, values (pp. 59–74). New York: Kluwer.

Hughes, R.I.G. (1999), The Ising model, computer simulation, and universal physics. In Mary Morgan and Margaret Morrison (eds.) Models as Mediators. Perspectives on Natural and Social Science Cambridge: Cambridge University Press, 97-146.

Humphreys, P. (1994), Numerical experimentation. In P. Humphreys (Ed.), Patrick Suppes: Scientific philosopher (Vol. 2, pp. 103–121). Boston: Kluwer.

Humphreys, P. (2002), Computational models, Philosophy of Science 69: 1-11.

Humphreys, P. (2004), Extending ourselves: Computational science, empiricism, and scientific method. Oxford: Oxford University Press.

Jacob, F., and Monod, J. (1961), Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Kerner, E. (1957), A statistical mechanics of interacting biological species. Bulletin of Mathematical Biophysics 19, 121-146.

Knuuttila, T. (2009), Isolating representations versus credible constructions? Economic modeling in theory and practice, Erkenntnis 70: 59-80.

Knuuttila, T., and Loettgers, A. (forthcoming), The productive tension: Mechanisms vs. templates in modeling the phenomena. In: Humphreys P, Imbert C (eds) Representations, Models, and Simulations. Routledge.

Knuuttila, T., and Voutilainen, A. (2003), A parser as an epistemic artefact: a material view on models. Philosophy of Science , 70(5): 1484-95.

Konopka, R.J., and Benzer, S. (1971), Clock mutants of drosophila melanogaster. Proc. Nat. Acad. Sci. 68(9): 2112-2116.

Lehhard, J. (2006), Surprised by nanowire: Simulation, control, and understanding. Philosophy of Science 73:605-616.

Lenhard, J. (2007), Computer simulation: The cooperation between experimenting and modeling. Philosophy of Science 74: 176-194.

Loettgers, A. (2007), Model organisms and mathematical models and synthetic models to explore gene regulation mechanism. Biological Theory 2(2): 134-42.

Loettgers, A. (2009), Synthetic biology and the emergence of a dual meaning of noise. Biological Theory 4(4):340-56.

Mäki, U. (1992), On the method of isolation in economics. Poznań Studies in the Philosophy of Science and Humanities 26, 316-351.

Mäki, U. (2005), Models are experiments, experiments are models. Journal of Economic Methodology 12:303-15.

Mitchell, S. (2003), Biological complexity and integrative pluralism. Cambridge: Cambridge University Press.

Morgan, M. (2003), Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. In H. Radder (Ed.), The philosophy of scientific experimentation (pp. 216–235). Pittsburgh: University of Pittsburgh Press.

Morgan, M. (2005), Experiments versus models: New phenomena, inference, and surprise. Journal of Economic Methodology, 12(2): 317–329.

Morrison, M. (2008), Fictions, representations, and reality. In M. Suárez (Ed.), Fictions in Science: Philosophical essays on modeling and idealization (pp. 110-135). New York & London: Routledge.

Morrison, M. (2009), Models, measurement and computer simulations: the changing face of experimentation. Philosophical Studies 143, 33-57.

Norton, S., & Suppe, F. (2001), Why atmospheric modeling is good science. In C. Miller, & P. N. Edward (Eds.), Changing the atmosphere: expert knowledge and environmental governance (pp. 67-105). Cambridge: MIT Press.

Parker, W. (2009), Does matter really matter? Computer simulations, experiments and materiality. Synthese 169:483-96.

Sprinzak, D., and Elowitz, M. (2005), Reconstruction of genetic circuits. Nature, 438, 442-448.

Strogatz, S. (1994), Nonlinear dynamics and chaos, Perseus Books Publiching, LLC.

Suárez, M. (2004), An inferential conception of scientific representation, Philosophy of Science (Symposia) 71: 767-779.

Swoyer, C. (1991), Structural representation and surrogative reasoning, Synthese, 87: 449–508.

Weisberg, M. (2007), Who is a modeler? British Journal for the Philosophy of Science 58: 207-233.

Winsberg, E. (2003), Simulated experiments: methodology for a virtual world. Philosophy of Science , 70, 105-25.

Winsberg, E. (2009), A tale of two methods. Synthese 169: 575-592.