Studies in History and Philosophy of Science Part C – Biological and Biomedical Science (in press)

Modeling Systems-Level Dynamics: Understanding without Mechanistic Explanation in Integrative Systems Biology

Miles MacLeod

University of Helsinki

miles.macleod@helsinki.fi

Nancy J. Nersessian

Harvard University

nancynersessian@fas.harvard.edu

Abstract

In this paper we draw upon rich ethnographic data of two systems biology labs to explore the roles of explanation and understanding in large-scale systems modelling. We illustrate practices that depart from the goal of dynamic mechanistic explanation for the sake of more limited modeling goals. These processes use abstract mathematical formulations of bio-molecular interactions and data fitting techniques which we call top-down abstraction to trade away accurate mechanistic accounts of large-scale systems for specific information about aspects of those systems. We characterize these practices as pragmatic responses to the constraints many modelers of large-scale systems face, which in turn generate more limited pragmatic non-mechanistic forms of understanding of systems. These forms aim at knowledge of how to predict system responses in order to manipulate and control some aspects of them. We propose that this analysis of understanding provides a way to interpret what many systems biologists are aiming for in practice when they talk about the objective of a "systems-level understanding."

Systems biology has provided an important motivation and material for extending philosophical conceptions of mechanistic explanation to dynamic systems (see Bechtel, 2011; Bechtel & Abrahamsen, 2005, 2010; Brigandt, 2013; Levy & Bechtel, 2013; Fagan 2012). Our aim in this paper is to explore typical cases of model-building in systems biology research that depart from standards of dynamic mechanistic explanation. These departures are practically and pragmatically motivated by the many constraints systems biologists face. Studying these departures opens up a window on a very rich largely unexplored set of practices afforded by the power of modern computational and mathematical techniques of approximation and abstraction, particularly in connection with parameter-fixing. These practices allow modelers to extract reliable information and achieve some form of understanding of systems despite errors and inaccuracies in their representations. In this paper we investigate how these practices operate, how to situate them with respect to mechanistic explanation, and what kind of understanding they produce.

Our analysis will draw upon cases from a 4-year ethnographic investigation of labs working in the bottom-up stream which self-identify as doing "integrative systems biology," a form of systems biology that aims to integrate computational and mathematical methods with experimental biology. These labs focus both primarily on modeling metabolic networks and to a lesser extent on gene regulatory networks and inter-cellular interaction. We conducted unstructured interviews and field observations, followed by semi-structured interviews to focus on specific issues, and attended lab meetings, which varied in frequency depending on the lab. Researchers in these labs are mostly graduate students and mostly come from engineering backgrounds. These labs are nonetheless diverse in other ways. The first, Lab G, is composed only of modelers (engineers of various kinds and applied mathematicians), some of whom model biological pathways for various purposes, while others work on generating mathematical methods for parameter-fixing or structure identification. These modelers work in collaboration with molecular biologists from outside the lab. The second, Lab C, comprises mostly researchers who do both modeling and experimentation in the service of building their models, although usually with some collaborative support from molecular biologists outside the lab for theoretical guidance. Our claims about modeling in this paper derive from our analysis of ethnographic interviews around the model-building processes and practices of the modelers in both these labs, as well as the literature in this stream of systems biology.[1]

We find that in practice 1) systems biologists often frame limited predictive and explanatory goals with respect to complex target systems, which can be met without dynamic mechanistic explanations of how these systems function; and 2) these goals prescribe problems that are tractable using mathematical and computational techniques of parameter-fitting. These techniques rely upon layers

[1] This research is funded by the US National Science Foundation which requires following human subjects' regulations, so we use numbers to name the researchers of each lab. Also, many of the researchers are non-native English speakers, which accounts for the grammatical errors in the quotes.

of top-down abstraction and approximation that draw out desired relationships from a mechanism but compromise the accuracy of the resulting simulation model with respect to other aspects of the mechanism. These processes thus tend to render opaque the role that the underlying mechanism of parts and interactions plays in producing the phenomena of interest.

If mechanistic explanations are not the target, then there is important question to be asked about what it is this kind of systems biologists is aiming for. We will suggest that such modelers aim for understanding in terms of how to intervene on certain network elements in order to influence network dynamics. They aim for predictively accurate or robust mathematical representations of the dynamic relationships between a limited selection of variables from a system. This form of understanding is pragmatic in the sense that it aims for control (Lenhard, 2007), but is also pragmatic in the sense that it is a response to the significant constraints modelers face. It trades away the accuracy and detail of a mechanistic account, which would demonstrate the links between parts, organization, and system-level properties or phenomena, for limited but achievable information on the relations between specific key parts and the over-all dynamics of the system. This information serves the expressed goal of systems biology to achieve mathematical accounts of system-level dynamics, and indeed can be used to interpret what systems biologists mean in practice when they say they aim – in contrast to molecular biologists – at a "systems-level understanding."

This paper proceeds as follows. First, we outline the goals and practices of systems biology in general to provide some insight into the problem-solving context. Second, we briefly detail recent developments in the philosophical discussion of dynamic mechanistic explanation in relation to systems biology. Third, we examine with the aid of our own case studies how the attitudes of modelers and model-building practices in systems biology diverge from the objective of mechanistic explanation. Finally, we argue that nonetheless, these systems biologists often claim their mathematical models provide a form of non-explanatory understanding of dynamic relationships amongst variables, which is essential to their ability to construct models and handle complex systems.

1. Systems Biology

As many commentators have pointed out systems biology is a diverse field characterized by different methods and practices (Calvert & Fujimura, 2011; Clark, 2008; O'Malley & Dupré, 2005). One can best understand what unifies these practices as a shared commitment to model complex biological systems using computational and mathematical resources and to an often loosely specified idea of a "systems approach." Although dynamical systems theory and even the notion of "systems biology" have longer historical roots, the modern incarnation is about 20 years old, born of the widespread availability of adequate computational power, developments in mathematical and algorithmic techniques, and the development of mass data production technologies. (Kitano, 2002; Krohs & Callebaut, 2007; O'Malley & Dupré, 2005). Such technologies include in particular high-throughput data technologies which collect dense dynamic information from a system. One of the prime

methodological innovations of modern systems biology has been to integrate engineering with biology by integration engineering concepts of "system," and its mathematical methods for analyzing large-scale systems, with modern-day computational processing and high-throughput technologies for data collection.

Systems biology positions itself against traditional biological fields like molecular biology. The latter apply experimental techniques to measure molecular properties, often *in vitro*, discover interactions, and build pathways. Rather than studying molecules and their interactions systems biology studies the dynamic behavior and properties of the systems they form. The need for a "systems approach" is supported with several philosophical claims in this regard. Firstly one of the central claims of systems biology is that properties and biological functions of components are dependent on their participation within systems (Kitano, 2002; Westerhoff & Kell, 2007). Secondly since parts and operations are typically determined and modified within the bounds of large-scale systems, only at this scale can predictively accurate models be constructed to guide control and intervention on systems (e.g., for medical purposes). Operating with small-scale representations and smaller pathways risks neglecting many important interactions these pathways have with external elements and networks that control and ultimately determine their dynamics. Because of the complexity and sensitivity of large-scale networks (arising from the many interacting components and nonlinearities in the form of feedback relations), they need to be represented mathematically. Only mathematical analysis can draw out the dynamic features and the control relationships of such networks. The qualitative approaches of molecular biology simply cannot be effective in obtaining this information. All systems biologists more or less, thus, share a commitment to modeling system dynamics mathematically (see also O'Malley and Dupré, 2005; 1273).

That said, many quite distinct pursuits take the name "systems biology," ranging over different biological subject matter from genes up to ecological systems and over different methodological approaches, such as "top-down" and "bottom-up streams" approaches (Krohs & Callebaut, 2007; Westerhoff & Kell, 2007). The former aim to reverse-engineer system structure from dense data sets and the latter, to build models from lower-level data on pathway structure and component interactions. Thus, these streams represent quite distinct methodological approaches. For modelers in both streams, the predominant standard for a good model is often first and foremost predictive accuracy. Data are often left out of the model-building process to then test ("validate") the model and build faith in its robustness, although other arguments are often used to establish robustness as we will see. If a model can be validated this way, then inferences can be put forward regarding intervention and control based on analysis of the model.

A significant component of systems biology, however, is concerned not with large-scale modeling, but discovering design motifs or small sub-networks that exhibit a particular function that is reproduced and reused across different biological systems and different species. (Alon, 2007; Green &

Wolkenhauer, 2013; Levy & Bechtel, 2013; Wolkenhauer & Green, 2013). As Levy and Bechtel argue such models have mechanistic representation directly in mind rather than predictive accuracy.

It is because of approaches in systems biology that Bechtel and Abrahamsen (2005), Levy and Bechtel (2013) and Brigandt (2013) have sought to extend the concept of mechanistic explanation to complex dynamical systems.

2. Systems Biology and Mechanistic Explanation

Initial conceptions of mechanistic explanation in biology were arguably geared towards explanation in molecular biology and other relatively reductionistic biological fields such as neuroscience. For example, the original Machamer, Darden, and Craver (2000) paper characterized mechanisms as "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions." (3) This language served to require sequential ordering of mechanisms which suited well mechanisms like protein synthesis or cell-division that have sequential steps, and clear beginning and end points. However although this account of explanation is to some extent compatible with practices that focus on fairly linear processes and small scale systems, it does not lend itself well to systems-orientated fields, that build models of nonlinear, large-scale systems. The mechanisms underlying a system do not necessarily have beginning and termination points, but often exhibit cyclic and highly non-sequential organization. Feedback and feed-forward relationships drive the dynamic variation of parts and their properties generating complex nonlinear behavior. In biological systems these dynamic internal variations are governed also by external environmental constraints on a system, which serve to regulate and control internal operations. Nonetheless at the core of these patterns and operations are parts and their organization. Hence to construe systems as mechanisms requires more sophisticated notions of mechanism that can incorporate more dynamically interconnected behaviors.

Bechtel and Abrahamsen (2005) have, accordingly, formulated a notion of mechanistic explanation that reflects the role of modes of organization amongst parts in producing the behavior observed by researchers. According to their account, "[a] mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization." (423) A mechanistic explanation in these contexts demonstrates how parts and operations are "orchestrated" through particular modes of organization to produce the phenomena under examination. It is the extra information embodied in the internal organization and external control of the system that enables the mechanism to function as it does in any particular instance. This information cannot be left out. Bechtel and Abrahamsen term this type of explanation *dynamic mechanistic explanation* since such mechanisms depend on the dynamic variability of their parts and operations. As such, dynamic mechanistic explanation places standards of accuracy on the correspondence of parts, relations, and organization that will ensure the patterns of orchestration are exhibited accurately. Dynamic

mechanistic explanations are meant to explicate how a mechanism functions (given external input and constraints) to produce a particular phenomenon.

Brigandt (2013) and Bechtel and Abrahamsen (2005) both argue that computational simulation of a mathematical representation of a dynamical system is essential to being able to show that a proposed mechanism does in fact produce the phenomena under investigation. Brigandt argues that in systems biology, particularly where qualitative emergent phenomena are the subject of explanation (such as threshold behavior or oscillatory patterns), mathematics is necessary for mechanistic explanation. Only mathematical methods are able to demonstrate that such phenomena can emerge from a candidate mechanism. As he puts it dynamic properties "cannot be foreseen by consideration of the basic structure of the causal network, but only by the mathematical modeling of the network's quantitative interactions and the resulting dynamics, so that without such equations (and their quantitative parameters) the *explanandum* does not follow…" (4) Thus, for Brigandt the mathematical equations add essential detail about the behavior of the mechanism over and above the details of parts and interactions. Mathematical equations produce information not accessible in purely linguistic or pictorial form. As such Brigandt takes issue with Craver (2006, 2007), who argues that the mathematical formulation of a mechanism is simply representational, while the explanatory content resides only with the components and their interactions.

Brigandt (2013) and Levy and Bechtel (2013) argue against Craver (in press – cited in Brigandt 2013) further, that abstraction can be indispensable to the explanatory force of a mechanistic explanation. Levy and Bechtel, using cases of design motifs in systems biology, argue that the relevance of abstraction has been neglected in many discussions of mechanistic explanation. They argue that in fact abstraction plays a very important role in the ability of researchers to identify the relevant mechanistic elements and details actually responsible for a given phenomenon by peeling away complexities. The relevance of abstraction goes beyond issues of representation and serves a legitimate explanatory preference for general mechanistic accounts over detailed specific ones. Detail is not always a regulative ideal. Brigandt argues that what does not contribute to the explanatory content of a model should be left out on the principle that explanations are explanatory only in virtue of those details that contribute to the production of a phenomenon. If a functional description of the role of certain components suffices for this then there is no need to unpack those components into their own parts. Although these authors do not provide an account of abstraction, their analyses focus on a specific form, akin to what mathematicians call "generalization" and Nersessian (2008) calls "generic abstraction" (in order to differentiate it from generalization in logic). Nersessian argues that there are several forms of abstractive reasoning that enter into model-building processes. The most common and widely recognized forms are: idealization, approximation, limiting case, simplification, and generic abstraction. All of these play a role in eliminating complexity and provide a means of fitting data to mathematical representations. It would take us too far afield to attempt a taxonomy that specifies the distinct features of each form of reasoning, and the differences in representational

change they achieve can be subtle. We single out "generic abstraction" here since it relates to Levy and Bechtel's point about the importance of abstraction in creating design motifs. Generic abstraction selectively suppresses information of the specific features of a case so as to make inferences that pertain only to the generic ("in general") case, such as inferences about triangles-in-general as opposed to an equilateral triangle. Generality in representation is achieved by interpreting the components of a representation as standing for types, rather than specific instances of these. Design motifs provide instances of generic abstractions in systems biology.

Our purpose here is not to deny the value of mechanistic explanation or to critique existing accounts, but to add to the mix of objectives and outcomes of systems biology research an awareness of wide-spread, important practices that do not have mechanistic explanation as an objective. These practices provide significant understanding of the behavior of large-scale dynamical systems and afford the possibility of prediction, manipulation, and control of their behavior. We take the accounts of dynamic mechanistic explanation as our starting point for our discussion of practices in the labs we investigated. These accounts all agree that dynamic mechanistic explanation involves 1) demonstrating that a mechanism is represented by a correct set of parts and 2) building a mathematical model that captures to a sufficient extent the interactions between those parts, that when simulated, adequately reproduces the phenomena of interest. These criteria provide specific constraints on what counts as a dynamic mechanistic explanation. However, as we noted above, each analysis focuses on particular classes of problems and practices in systems biology. For instance Brigandt picks out models built for the purposes of explaining what he calls qualitative phenomena which are often termed "emergent" by scientists themselves. The dual stable states of the apoptosis system in cells, for instance, cannot be understood by breaking the system into components and studying them in isolation. The target in these cases is clearly explanation, viz., to explain how these phenomena arise by capturing the mechanism that produces them. Levy and Bechtel concentrate on design motif research, a project that aims principally to explain general biological function by extracting general principles of mechanistic organization. Finally Bechtel and Abrahamsen focus on circadian models that were developed over successive periods of investigation, some closely attendant on molecular biological investigations. The tight linkage of the empirical research with these investigations helped to develop rigorous mechanistic details for modeling. All these cases are instances of successful well-developed models that do seem to function as mechanistic explanations. None of these authors would claim that dynamic mechanistic explanation should account for all modeling in systems biology. Rather, they have used systems biology to develop a more expansive account of mechanistic explanation. We, too, will be examining modeling specific kinds of systems. In the labs we have investigated modelers deal with large-scale dynamic systems different from those that have contributed to the dynamic mechanistic account. Building dynamic mechanistic explanations is not pursued, and not necessarily desirable given complexity and other constraints of modeling these systems. Nevertheless, having this notion of dynamic mechanistic explanation in hand provides us a base-line, as we will see, for understanding the choices and trade-offs many modelers

make in practice, particularly since at the start these modelers rely on quite detailed pathway representations of their systems that capture much mechanistic information.

3. Model- Building in Integrative Systems Biology: Insights from an ethnographic study

To understand the nature of the complex relationship in practice between mechanisms, explanation, and understanding in integrative systems biology, we need first to have some understanding of the processes of problem-solving and model-building in this bottom-up stream of systems biology.

Despite the diversity of goals in modeling large-scale systems, there are three canonical steps in the model-building process (see Voit 2000; 2013). In the first place a biological pathway is constructed of the relevant interacting network of metabolites and genes involved in the process or system. Sometimes the experimental collaborator provides a piece of a pathway, but modelers always need to build-out this pathway from the literature and databases in addition to whatever data the experimentalist might have shared with them. Secondly this pathway is represented mathematically. Thirdly unknown parameters are fixed using the available data. The model is then tested predictively to check its validity and robustness. This process might involve a collaborator conducting experiments or the modeler might reserve some data independently of the parameter-fixing process and use that to test the model. If the model does not replicate the data, parameters need to be refit.

Model building practices, of course, are not bound to and, in our investigation, are rarely able to follow such schema rigorously, and divergences can be quite substantial. Model-builders face highly complex problem-solving tasks trying to build workable models which limit their ability to perform each step independently. In practice these steps are run together, as modelers adapt the pathways in interaction with the developing model to fit what is workable in terms of parameter-fixing and adapt parameter- fixing methods and mathematical representations to fit the structures and data they have. Figure 1 provides a more accurate representation of the model-building process. We constructed it from our data and then verified its accuracy with our participants.

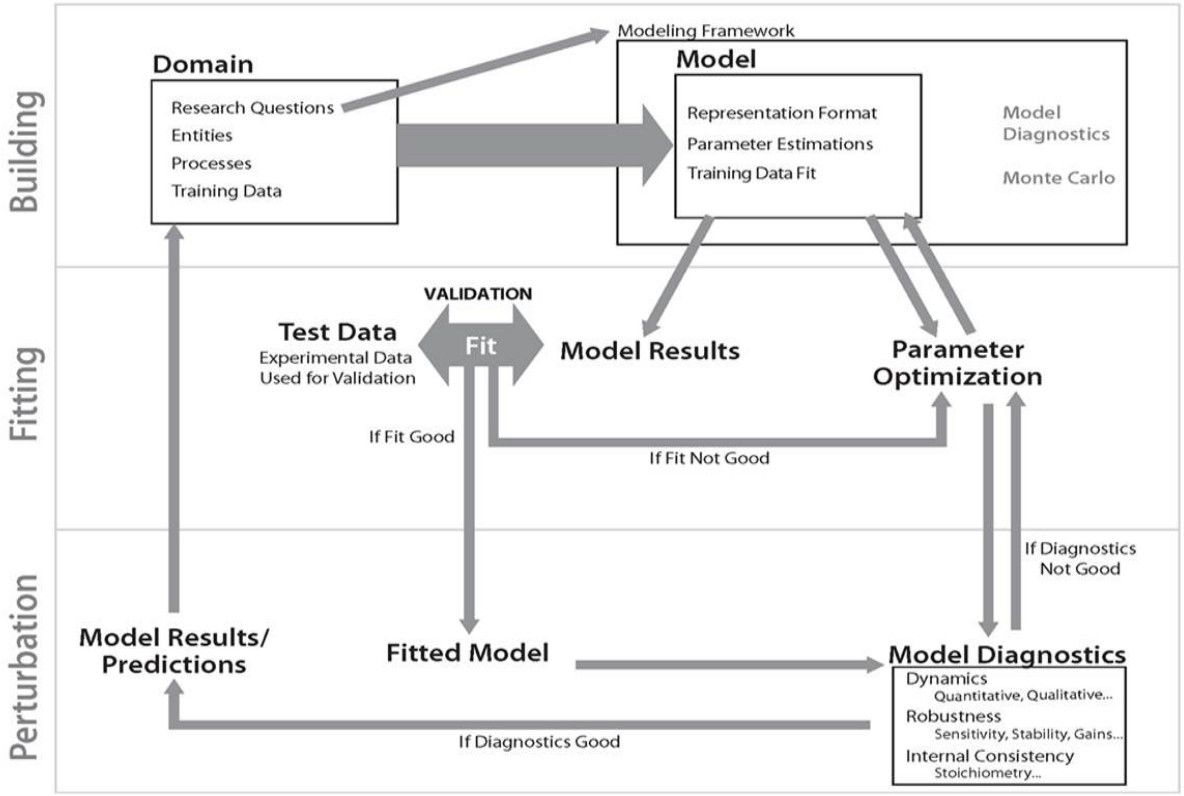

Figure 1. Our map of the modeling process. In practice, model-building is complex and highly recursive.

There are three factors that make a linear process of model-building more or less impossible. Firstly the biological systems themselves are extremely complex. Modelers almost always have to coordinate between pathway representation, mathematical representation, and parameter sets to adapt a problem so as to make tractable the production of a reasonably robust or predictively successful model, at least within a given range. Secondly the data available to modelers are almost always inadequate and rarely time-series. Much of the data is foraged from the literature and online databases by modelers, but these data are rarely sufficient or are only available in a form that makes dynamic modeling tricky, such as steady-state data. Very few parameters required for a dynamic model have been measured experimentally beforehand. High-throughput time series data – the gold-standard for the ordinary differential equation modeling that our researches conduct – from which parameter values can be derived easily are available rarely (MacLeod and Nersessian 2013a, 2013b). Information might be limited not only on parameter values but often also on pathway structure. All of these factors and more add up to a situation where modelers often have to adapt their problem-solving approach to the specific problem, including having to design and tailor algorithms for parameter-fixing that fit the situation at hand. As we will see in the next section, abstraction strategies that can handle the constraints for building large-scale models are not always directed at

identifying the system mechanism for explanatory purposes, especially when the goals are not specifically explanatory themselves.

4. Non-explanatory goals in pathway modeling

Rarely is "mechanistic explanation" of a system or system-level phenomena given as an explicit goal by our researchers. For the most part, our modelers have more limited and specific goals in mind. These at least partially serve the aim of systems biology to capture mathematically system-level properties and relationships. We look at two goals in particular that represent a broader approach in both our labs. These focus on just specific relationships between particular dynamics of systems and particular parts.

1) Modeling a system in order to discover robust mathematical relationships among particular "input" and "output" variables in a system (in order to intervene on them to manipulate real-world phenomena).

2) Modelling a system in order to infer the (possible) role of a particular molecular or component process in a network and its interactions; then using this information to predict the effect of manipulations of this molecule or process on system dynamics.

In lab G, in particular, these goals are often connected to requests by experimental collaborators to use the data they have to help generate hypotheses about what is missing or where the data might be wrong or to discover some new relationships in the data. With this information they can better direct their experiments. These goals are not explicitly tied to having dynamic mechanistic explanations of the systems involved and can conceivably be satisfied by more limited mathematical representations that capture only specific relationships. Let's consider two cases.

One researcher "G10" was studying lignin synthesis with a goal of the first type above. Lignin is a natural polymer that helps harden plant cell walls. The hardening provides the plant with structural rigidity, and thus supports growth. While this hardening property is biologically useful for the plant, it is a problem for the biofuel industry, because lignin is difficult to breakdown (exhibits what they call "recalcitrance") when biomass is processed into fermentable sugars (using enzymes/microbes). This recalcitrance of lignin makes the extraction of sugars from biomass difficult and costly. G10's task was to model the lignin synthesis pathway in order for the experimentalists to optimize current transgenic biomass producing species to break down lignin. His chief target for doing this was to find mathematical relationships that would connect concentration levels of the building blocks of lignin (monomers H,G,S) and the key enzymes over which control might be exerted with the S/G ratio, which was known to be a factor in the resistance of converting woody biomass to biofuel.

G10 built two models for this, one focusing on synthesis in the poplar tree and another in alfalfa, but both pursued similar strategies. In each case the first step was to isolate from the literature the

relevant interconnected pathways in the production of H, G and S lignins which meant deciphering the dominant components and interactions contributing to the production of these lignins, but also hypothesizing relations and components missing from the literature. (see Figure 2)

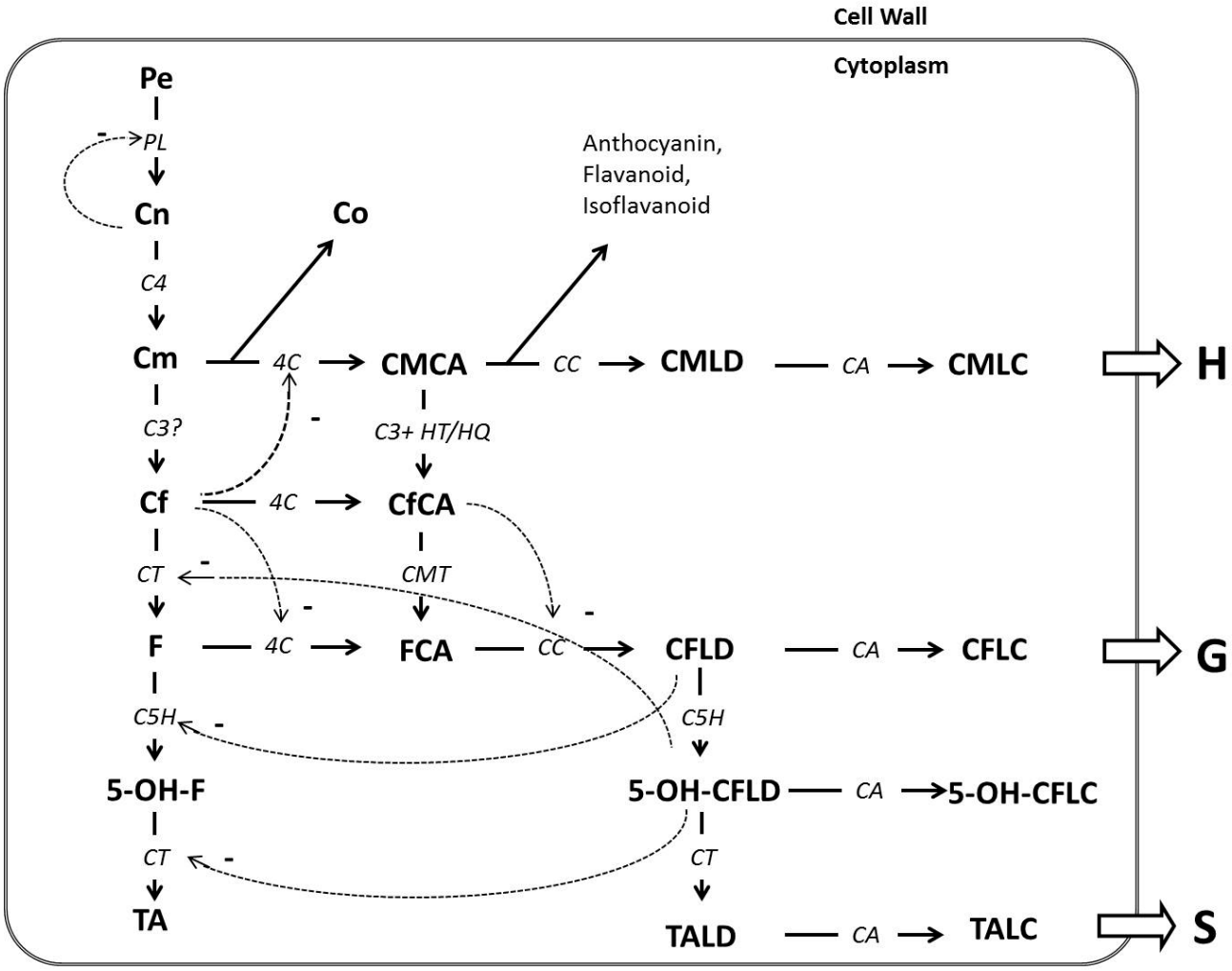

Figure 2. G10's lignin production pathway model for the poplar tree built from the published experimental results. The chemical names of the reactants are not relevant for our purposes and have been substituted by us. H,G and S are the outputs.

G10 had very little data with which to build his models and complex systems. He used some mechanistic information about the system to produce a model, but at same time employed several abstraction strategies. To begin with he used a static constraint-based approach (modeling the system at steady state) to estimate rate constants leaving out the effect of potential nonlinear interactions on these. Following that, in his words, "the optimization were done only for a few significant parameters,

and we set the remaining parameters to the values that were deemed biologically reasonable." Interestingly, these significant parameters were determined statistically, by calculating Pearson correlation coefficients for each with the S and G ratio. This required running the model many times across different parameter sets and collecting data points. It helped fine-tune the model for the sole purpose of its ability to predict and control effects on the S/G ratio. These important parameters were finally fit algorithmically to data sets and tested for predictive accuracy against data held back from the fitting process.

In the end five optimized models were found to test well, and each gave similar predictions (with no guarantee a global optimal was found) providing convergent results that G10 could rely upon. As such he was able to claim in his dissertation: "Now, since we had a valid model, we could use it to find out which enzymes should be targeted to, for example, decrease the S/G ratio." G10 built his model task and strategies towards extracting dominant dependencies and using this to determine how to control the network by targeting particular points in it. For instance he was able to make the following computational estimations: "So, if you knock down just one enzyme, you could get a 25% decrease in the S/G ratio. But if you're allowed to change the expression of as many as three enzymes, you could get a 38% decrease in the S/G ratio, which is huge compared to the natural variation that is about 5 to 10%."

Another researcher, "G12," was given her task by a collaborating experimenter who approached the lab director with a set of data looking for some modeling to help him further his experimentation: "He [the collaborator] has some pathway… they found some novel mechanosensitive genes that can activate this Nox, this enzyme, leading to this inflammation response. So they want to collaborate with us to build up a model." These data unfortunately could not provide much information for dynamic modeling, and mostly provided G12 with a pathway map for the inflammation response but few useful parameter measurements. One of her tasks was to try to model the dynamic response of the Nox system to external stimuli, meaning to try to work out how Nox was interacting and regulating the inflammation system. Like G10 she built a mathematical model of the overall system again with the limited data from her collaborators with which to test various hypotheses about how Nox might be interconnected within the network. She tested the alternatives according to which gave the best response across a large multivariate parameter space using Monte-Carlo sampling of parameter values. The testing did not result in an integrated mechanistic explanation of the system's function. She, nonetheless, had a platform for reasoning about the likely nature of Nox interaction within the network.

With this information on regulatory roles of Nox she could move to the next step, which was to model the dynamic response of interventions on the network (through an inflammation activating agent AngII). This process required additional hypotheses about the signaling mechanism (in turn, requiring additional components for the Nox model) through which this intervention could occur. She

researched three alternative signaling mechanisms and again applied Monte-Carlo testing of how her Nox model performed with each addition across large parameter spaces, picking the best performing model. This left her with one best overall choice of additions, although with many possible alternative parameterizations for the upgraded Nox model. However looking at common features of all these model possibilities she was able to compute relationships among particular molecular elements and dynamic features of the network and make some predictions about which particular signaling elements had most impact on her network dynamics.

In both these cases G10 and G12 fashioned limited goals to extract only specific information from their systems, and they constructed model-based arguments to argue the results were reliable for their purposes, despite a lack of data. We find in general that our researchers select their goals based on the nature of the data they have and what they think they can do with data of this sort. Often discovering what can be done with a particular data set is part of the model-building process itself. What might be a successful project is unclear to researchers at the outset and only emerges after extensive trial and error with different projects and goals. Many of these research projects do, nonetheless, seek new mechanistic information for experimenters or make new causal explanatory claims in terms of the regulatory roles particular elements and processes might play in a given system. However, the goals do not depend explicitly on being able to provide mechanistic accounts of the systems involved. In turn there is room to be selective about what is represented and what counts as a satisfactory outcome. Nothing we have said so far however shows that these kinds of model-building practices produce results that could not at least approximate mechanistic explanations, only that the aims are not explicitly mechanistic explanations. However having goals which are not explicitly tied to mechanistic explanation frees these researchers up to apply abstraction strategies that do not draw out explanatorily relevant details. Rather, these strategies serve particular goals by finding tractable model-building pathways for testing and extracting particular dynamical relationships, often at the expense of explanatorily relevant details.

5. Top-down Abstraction Strategies

As we noted above, there are many forms of abstraction that scientists use in constructing models. Abstraction in systems biology modeling facilitates mathematization and computational simulation by selectively discarding information so that the model instantiates only specific elements deemed most relevant for the system at hand given the specific goals. As discussed, Brigandt (2013) and Levy and Bechtel (2013) have argued that model-building in systems biology that has the goal of mechanistic explanation discards information deemed irrelevant to the explanation in order to draw out only explanatorily relevant features and create generalizable explanations for that kind of system. However, the kinds of goals we discussed in the previous section that modelers of large-scale systems pursue, due to the complexity of these systems and the lack of data and its noisiness, often favor abstraction strategies that obscure the relationship between the underlying mechanism and the

modeling outcome. We label these strategies "top-down" because they principally operate to find mathematically tractable ways to fix uncertainties in particular parameters in a model by best fit to the phenomena. Here we discuss these abstraction strategies in terms of what they achieve for parameter fixing rather than in terms of how they map to a taxonomy of general abstractive techniques.

The epistemic significance of parameter fixing (or parameter estimation) in model building has not been addressed sufficiently by philosophers of science (notable exceptions are Parker (2010a, 2010b) and to some extent also Winsberg (2010)). In systems biology it is a strong determinant of the model-building process and of the form resulting models take, and is often at the center of modeler discussions about what any particular model achieves. Parameter estimation is widely acknowledged by modelers to be the most difficult task they have to face, given the limited data and computational constraints with which they are usually work. Typically it is the part that requires most ingenuity as well as developing what our modelers call a "feel for the model." Modelers of large-scale dynamical systems can typically have upwards of 30 unknown parameters for a particular system. The complex nonlinearity of the models themselves often make it difficult to estimate in what parameter ranges best fit solutions are likely to be found. The fact that global maxima are hard to find and be assured of means that researchers often revert to strategies based on "best performance" across parameter ranges or convergence among multiple models with different parameter values as we saw in the cases of G10 and G12 above.

Top-down fitting processes generally rely on three features: canonical mathematical frameworks that help cut down mathematical complexity required for parameter fixing, various parameter simplifying and fixing processes that help simplify the parameter space, and global processes that fit remaining uncertainties simultaneously.

5.1 Canonical Mathematical Frameworks

The first aspect of model-building commonly used in our labs involves abstracting the relationships among components from the details of the molecular biology into a tractable mathematical form that can be fitted to the data. This strategy is one of the most important problem-solving moves that systems biologists make, since it serves to reduce what would otherwise be an intractable degree of complexity. For example the Michaelis-Menten relations are a common representation of enzyme kinetics. These represent specific models of catalytic reactions considered canonical mechanistic representations, particularly among molecular biologists.[2] Other formulae also exist that are more

[2] Michaelis-Menten kinetics is a specific model of enzyme kinetics or the rate at which enzymes catalyze a particular reaction. Michaelis-Menten postulates that the reaction rate $v = V_{max}[S]/(K_m + [S])$ where $[S]$ is the concentration of the substrate molecules (the molecules acted upon by the enzyme), V_{max} is the maximum reaction rate, and K_m is the concentration at which v is half of V_{max} measured empirically. This model's most important property is its prediction of a saturation rate for $v = V_{max.}$

detailed or cover different types of enzymatic mechanism. However the requisite data are rarely available for modelers to easily incorporate Michaelis-Menten. Since the Michaelis-Menton factors involve rational functions, they are harder to simulate, and parameter spaces are harder and more time consuming to explore, which create often intractable parameter-fixing problems for large-scale models. Another option is to use more abstract representations of interactions that require less knowledge or data of the mechanistic interactions. Power laws and lin-log representations provide such kinds of abstractions, and in theory these representations are flexible enough to mathematically approximate many kinds of interaction relationships, including highly nonlinear ones, between metabolites over a wide range of conditions within the degrees of freedom set by their parameters (Voit, 2000, 2013). According to Biochemical Systems Theory, for instance, which utilizes power law relationships, these power laws cover the variety of relationships commonly observed in biochemical systems. BST posits that the processes of a given metabolic/signaling network can be modeled dynamically as power law functions of the form $\gamma_{ik} X_j^{f_{ikj}}$ (Savageau, 1969; Sorribas & Savageau, 1989). There are two types of canonical representation, Generalized Mass Action (GMA) and S-System formats (Voit, 2000). Firstly GMA uses ordinary differential equations of this type:

$$\dot{X}_i = \sum_{k=1}^{T_i} \left(\pm \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ikj}} \right)$$

Where T_i is the number of terms in the *i*th equation. The X_i's represent the concentrations of the various molecular elements in the network. Each separate flux in and out is represented by an individual power law.

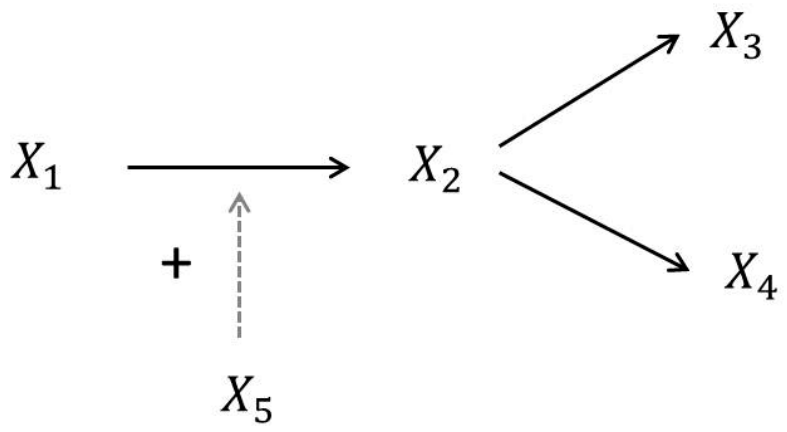

Figure 3. A simple network model with a feed-forward relation.

For example, in the case of the system above (Figure 3) we model $\dot{X}_2$ as:

$$\dot{X}_2 = \gamma_{21} X_1^{f_{211}} X_5^{f_{215}} - \gamma_{23} X_2^{f_{232}} - \gamma_{24} X_2^{f_{242}}$$

The S-system format, on the other hand, pools all the effects of ingoing and outgoing flux into one term:

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}$$

In this case;

$$\dot{X}_2 = \alpha_2 X_1^{g_{21}} X_5^{g_{25}} - \beta_2 X_3^{h_{23}} X_4^{h_{24}}$$

Applying these canonical models is a matter of filling in the network details (which X_i's are interacting with which others), and the parameters with which such interactions take place; namely the kinetic orders (the *h,g*, and *f*'s) and the rate constants (the α, β, γ's). These terms are relatively simple mathematically compared to other options that more closely model the mechanisms and involve relatively fewer parameters.

Many modelers, including in our labs, apply this methodology. However, in using these abstraction methods the details of the mechanism of interaction are not represented, but deliberately "black-boxed" for the sake of tractable mathematics. These power-law terms represent only mathematical relationships between flux-in and flux-out. What they do provide, however, are representations that can be fixed to the data through manageable algorithmic processes and which provide easily analyzable structures for obtaining information on dominant dynamical relationships and network control structures. Both G10 and G12 relied on the affordances of this canonical system. G10, for instance, relied sparingly on knowledge of the details of enzymatic or other interactions and, in fact, had little data available to represent these interactions. He applied the GMA system and was able through various parameter-fixing techniques to discover the important dynamical relationships for lignin synthesis by using the mathematical affordances of the BST systems to analyze the network for its sensitive parameters. The rest were set arbitrarily in order to reduce the parameter-fixing task. BST is designed to facilitate just this kind of parameter-fitting.

5.2 Shrinking and Simplifying the Parameter Space

Fixing parameters is rarely just an automated final step. Our modelers often have to modify data sets iteratively to get a more tractable data fitting problem as they better understand the problems they are facing These modifications aim to shrink and simplify the parameter space through combinations of abstraction techniques including off-lining data, using data sets that represent a simpler case,

linearizing the data, optimizing the data to biological constraints, numerically simplifying the data to say integer values, and smoothing the data. These moves in turn facilitate the introduction of simplified mechanisms that are thought capable of reproducing these simplified data in simulations.

For example G10 had 27 unknown parameters in his poplar model, a brief account of which we gave above. He used several abstraction techniques to determine them. Firstly he started from the non-dynamic steady-state state (wild-type data) available from his collaborators. He used flux balance analysis, which assumes a steady state net input and output to the network and thus linearizes flux relationships, to estimate rate constants for the network. He then optimized just a few significant kinetic order parameters while more or less guessing the rest based on what was biologically plausible. However, as G10 reported in his dissertation defense, linearizing the model for flux balance analysis to fix the steady-state data in this way served to eliminate much mechanistic information, such as the ”kinetic information of individual enzymes” and “nonlinearities like the competition between substrates, or the allosteric regulation of enzymes by activators or inhibitors.” Doing so created a simplified structure that could be fit easily to the data but left it unclear as to what extent the model represented the underlying mechanism. Nonetheless as we saw the result enabled him to make some assertions about the dynamic relationship between the S/G ratio and particular enzymes and enabled predictions that were borne out by his experimental collaborators.

These decisions are made principally by our researchers to make the parameter search problem tractable rather than to reduce mechanistic information in way that helps account for how the mechanism works, as we saw with the cases used for the dynamic mechanistic explanation account. As such it becomes increasingly difficult with layers of top-down driven abstraction choices to recover a mechanistic account. Choices of mechanistic detail to include are, in fact, sometimes made on the basis of what facilitates the best fit across a range of possible unknown parameters. G12 modeling atherosclerosis, for instance, had a very difficult parameter-fixing challenge when trying to hypothesize how the Nox network responded to AngII. She required a number of abstractions. Firstly because of lack of data and the complexity of the system, she had to off-line various variables in the system (disconnecting them from external dynamical influences) and hand feed values for them into the model in the process of searching for a good fit. She used sensitivity analysis to reduce the set of parameters she would need to only those that dominated network activity. However the most difficult problem as we saw was that she had three possible candidate signaling mechanisms. This generated a multidimensional parameter fixing problem, which required evaluating the multiple alternatives she hypothesized across different parameter sets. Her only hope was to find ways of shrinking the parameter space further. For instance she approximated one subsystem at steady state. This allowed her to generate linear flux conditions that reduced the number of unknown parameters. She was in the end required to test 3 different mechanism candidates using 6 different performance conditions derived from the data, which she did by approximating 4 critical parameter ratios as simple 1’s and 0’s (or high or low states). The 16 combinations of 1 and 0’s resulted in 48 possible

combinations to test, 16 for each model. Her final recourse for such complexity, as is often the case, was testing through Monte Carlo sampling. This process distinguished one of these candidates as giving a distinctly better performance for the overall model over the parameter space of unfixed parameters than the others. As a result of using a large-scale model of the inflammation response system, she managed to extract a preferred mechanism describing how AngII could trigger the inflammation network. However the process involved so many abstractions directed principally at transforming her problem to one that could be solved through a fitting process, she would not claim to have provided a mechanistic explanation of the inflammation-response process with this model.

5.3 Global Fixing

Modelers often try out a range of different methods to try to fix their remaining parameters globally, including basic grid searches, linear and nonlinear regression, genetic algorithms, or simulated annealing. Typically these do not work on first iterations due the complexity of the model and noise in the data. They need to be combined with further mathematical approximations as well as simplifications of the pathway representation in order to move toward a parameter problem that can be solved, thus, giving-up potential representational accuracy. Often a modeler can end up with alternatives that fit the data with large degrees of uncertainty. Substantial work in systems biology is directed towards coming up with better algorithms for the parameter-fixing process.

As we have noted, parameter-fixing is in essence a process of abstraction in that it fixes from above (or top-down) rather than from the mechanistic details. Finding a best-fit is about fitting the structure you have to the data you have, in the hope that the mathematical structure of the dynamics is approximated within the degrees of freedom of the mathematical representation and in the hope that this approximation will be recovered through the parameter fitting process. But as every parameter-fixer knows, with a high dimensional parameter fixing problem of complex nonlinear networks the best fit is rarely found, and what one seeks is an efficient technique to find solutions that are robust over an important range. Often this means, as we saw in the case of G10 and G12, settling for a family of models and deriving conclusions from those rather than one model in itself.

5.4 Obscuring Mechanisms

The result of this process of mathematically fixing parameters through its layers of abstraction is to obscure the relation of the final mathematical representations to the underlying mechanistic processes of the system. Parameters throughout the process are left out or given numerical dummy values to shrink the parameter space in the hope that the fixing process will compensate for any errors introduced. Global fixing processes do not fuss about the accuracy of one parameter: but seek out combinations that render the overall model accurate. The result is often systematic distortion. Parameters can compensate for the errors of one another and many different parameterizations can yield exactly the same residual error. In both G10 and G12 cases this was precisely the situation with

which they were left. Actual global optimal solutions are rarely found, and there is rarely an opportunity to check fixed parameters experimentally. The result is that although the “parts” of the model might to a large degree correspond with parts of the mechanism, through for instance the pathway diagrams, the ”operations” diverge from those of the mechanism. This means that the dynamics of the models are never strictly accurate, but rather tuned to fit. In such circumstances it cannot be expected that the final model captures the actual “orchestration” at work in the mechanism or even a simplified but accurate representation of it.

The kind of modeling witnessed in our labs can make partial contributions to knowledge of the causal mechanisms of the systems being modeled and can also discover causal molecular elements previously unknown during the process of model-building. These contributions can serve to help advance models of the mechanisms. But because the top-down processes of abstraction principally aim to transform a problem into one that is mathematically tractable for fitting, the resulting mathematical models bear an unclear explanatory relation to the underlying mechanism. They cannot be used to give an accurate account of how the entire mechanism of a system is composed and orchestrates the production of phenomena.

6. Understanding without mechanistic explanation

6.1 “Somewhere in between” mechanistic and phenomenalistic

The account we have developed from our investigation reflects the attitudes of our participants – modelers of large-scale dynamical systems – towards what they produce. Our researchers do express some understanding that their work has a complex relationship with explanation and with the nature of the underlying mechanisms at work in the phenomena they study. For them, however, mechanistic explanation is principally what molecular biologists do, and indeed they often acknowledge that biologists are critical of their models precisely because they see the models as not providing an account of the mechanisms. As one modeler reported, “*They [experimenters] think of [a model] as something that's… just hooked up to – to, you know, match figures [graphs]. So, for them, it's just like, you're using your data and then, you know, you're plugging in some numbers to fit the output of your model to that.”* Accordingly, modelers often attribute to biologists a lack of understanding of how mathematical techniques can produce reliable outcomes through top-down processes, which leads biologists to be dismissive of the value of modeling. In an interview where we probed around the issues we address here, the director of Lab G described what modelers produce as lying “somewhere in between” a fully mechanistic explanation and a black box or “phenomenal account” as he calls it. When discussing the explanatory nature of G10’s Lignin model for instance G4 told us,

“Well, for me, if you can trace out a causal pathway then it’s an explanatory model even though you may not know every single detail of it. But every explanatory model somewhere has some regression

in there or some association in there. Again, it's not pure. But you have a certain level of explanation…. something causal you didn't know before."

Clearly G10 provided a causal account of the relationship between the S/G ratio and various input variables. His model, which is able to make counterfactual predictions about these relationships, satisfies Woodward's account of a causal explanation (Woodward, 2003). But the overall result is not "pure," as G4 put it, meaning that the model has a complex and unclear relation to the underlying mechanism through abstraction processes such as regression and association. The actual molecular mechanisms producing S/G behavior are not represented precisely. The kind of explanation thus attributable to such models rests on the causal-explanatory relationships extracted by them and on the new elements they might add as causal factors. However, irrespective of whatever the novel causal-explanatory content of a model might be, what is usually emphasized by researchers is how these models help shed light or "understanding" on the control and regulatory relationships that drive a system's dynamics.

6.2 Understanding at the systems-level

Our researchers often claim explicitly that what they achieve, and what they aim to achieve, is a measure of understanding of the dynamical behavior of their systems. As mechanistic explanation is not the source of this understanding, there is a legitimate question to be asked as to what sort of understanding such mathematical models provide. We cannot give a detailed answer to that question, but we can at least show that there are potential options for unpacking the notion of understanding at work in many cases in systems biology. What Lenhard (2006) has called (with respect to cases with similar features in physics-based simulation modeling) a "pragmatic" notion of understanding seems a reasonable option in the circumstances we have been discussing.

The notion of "systems-level understanding" that is mentioned widely in the literature on systems biology is unfortunately ambiguous. It is not clear, for instance, whether the notion as used by systems biologists might be unpacked through concepts of mechanistic explanation alone. For example Brigandt's cases of qualitative explanation demonstrate how a mechanistic account, through the aid of a mathematical formulation, can explain higher level system properties such as bi-stable switches. In these cases there does not seem any need to treat a systems-level understanding as anything more than dynamic mechanistic explanation.

However what complicates this option is that those that use the term "systems-level understanding" often seem to want to draw a contrast between mechanistic understanding and a systems-level understanding. For instance, according to G2, "So we are not necessarily dealing with mechanistic detail. We want a more system-level perspective." In C9's words, "You don't necessarily need to understand every single interaction of every single molecule in every single second to sort of understand the overall dynamics." It is this understanding of the "overall dynamics" that seems to

suggest that understanding the system dynamics has a different target than understanding how the parts and constraints coordinate to produce the whole and is a distinct and important goal.

Leading advocates for systems biology, such as Kitano (2002), also seem to encourage a distinction between mechanistic explanation/understanding and a system-level understanding, for which mechanisms are means but not the ends to achieving the latter. Kitano (2002), fittingly titled "Looking beyond the details," seeks to add clarity to the notion of a systems-level understanding which he labels "ambiguous." He argues that "system-level knowledge should be grounded in such a way that the system is composed of molecules; and molecules follow the laws of physics. However, how a system operates can only be described by a set of theories that focus on system-level behaviors" (2). Kitano employs a well-known traffic analogy to elaborate this point:

"This raises an important question 'what does it mean to understand a system?' A system is an abstract concept by itself. It is basically an assembly of components or static structures. It is basically an assembly of components in a particular way that exhibits certain behaviors. Obviously, it is not sufficient to draw a diagram of all gene and protein interactions. Such a diagram is necessary and provides insights on how the system works. But, it is like a road map where we wish to understand traffic patterns and their dynamics."(2)

The traffic analogy, although much used, does not provide much insight since traffic is, after, all a very homogenous system of identical parts that are unlikely to have specific functions. Traffic is unlike biological systems in many ways. In explaining "systems-level understanding," Kitano also talks about "dynamic properties" of a system including phase transition boundaries and other nonlinear behaviors. Such properties emerge through nonlinear dynamic analysis, including bifurcation analysis and other dynamic analytical methods, which "provide means to understand the dynamic properties of biological systems" (4). One, thus, might interpret Kitano as asserting that the content of this kind of understanding should be unpacked in terms of the dynamic relationships and structural properties that mathematical analysis can derive from a system representation and the predictions that can be made for intervening on systems by virtue of them.

The role of the parts in a systems-level understanding is nonetheless left unclear in Kitano's account. At the very least mechanisms have a role in model construction, but the essence of the understanding seems to reside in system properties and their relationships rather than the relationships of the parts to the whole. As Kitano intimates, along with other systems biologists, the ultimate goal of systems biology is to produce theories of biological systems. These theories would put systems biology on a par with theoretical physics in terms of the role it would play in directing and governing experimentation (Westerhoff and Kell, 2007). Systems biology theory should capture general mathematical features and properties of systems and facilitate the construction of theory-based models of individual systems. The use of "theory" in analogy with physics suggests that a systems-level understanding is not meant pragmatically, as in a capacity for control and manipulation alone,

but as a genuine theoretical or mathematical form of understanding from which control and manipulability would follow.

Of course, authors such as Kitano are articulating ideal scenarios and advocating for what systems biology might achieve or should aim for, rather than discussing what in practice it actually does. What our modelers are trying to achieve suggests that a more limited concept of understanding is applicable for describing the current state of play – at least as we witnesses in their activities. Such a form would be pragmatic in motivation and in content. It is pragmatic in motivation in the sense that although these modelers set robust and predictively successful mathematical description of system-level dynamics as a goal, per Kitano, that enable hypotheses about control and intervention, the complexity of the systems and the constraints of modeling force them to pursue more limited goals with respect to what they can learn about their systems. They make a pragmatic decision to pursue less robust and less predictively accurate models valid for just certain elements of systems. Mechanistic accuracy is the trade-off. The understanding these models provide is pragmatic also in content in that it is neither a higher-level mathematical/theoretical understanding nor a mechanistic explanation. Rather, this kind of modeling can be considered as making a kind of pragmatic move often associated with phenomenalistic representations, which relate variables and patterns in the data mathematically to an extent sufficient for manipulating a system reliably or at least making hypothesis about how specific changes might affect the system.

A useful analogy can be made to Lenhard's account of understanding in complex dynamical modeling in physics-based fields such as nanoscience. Lenhard (2006) argues that computational models and simulation are facilitating new modes of understanding that are different from the traditional mode of understanding as law-based explanations of phenomena. He argues that computation has given researchers an ability to build predictive models that, through abstractive techniques, render the operation of the laws from which the models are built "epistemically opaque," (a term Lenhard borrows from Humphreys (2004; 148)). Epistemically opaque models begin from theory and depend on theory, but equations produced for these complex phenomena are impossible to solve analytically. Instead "simulations squeeze out the consequences in an often unintelligible and opaque way" (Lenhard, 2006; 612), by applying abstracting and averaging techniques that fit the equations to the phenomena, as well as numerical methods that render equations computable. Although such simulation models of complex phenomena do not provide explanatory understanding, he argues that they nonetheless do provide the potential for intervention, manipulation, and control, and thus "pragmatic understanding" (see also Chandrasekharan & Nersessian forthcoming). Lenhard's claims about explanation with simulation models in physics-based science have some lessons for our cases in systems biology, since there we also encounter "epistemic opacity." In our case it is not the operation of laws that is rendered opaque for the sake of prediction and control, but the mechanism, or at least its representation, which is hidden beneath top-down abstraction techniques of parameter-fitting.

Modelers such as G10 start from fairly detailed pathway representations of mechanisms and use these to produce hopefully robust mathematical relationships between particular system variables. These are developed to hold over a range of manipulations since they are designed not just to model system behavior, but also make interventions on these systems. These results fit Lenhard's notion of pragmatic understanding. They reflect some knowledge of the dynamics of a system in terms of how variables relate and capture causal relations. But they are limited not only because only partial aspects of a system are correctly accounted for but also because the mechanistic details that produce the relationships are obscured. G12's derivation of the likely causal role of Nox in the inflammation system was also extracted at the expense of an accurate mechanistic representation of the system. However the result enabled her at least to make some hypotheses about how that element might contribute to particular patterns exhibited by the system (particular sets of output variables) or predictions about how it might be manipulated to direct the system in specific ways.

6. Conclusion

Philosophers often focus on cases of modeling that serve to exemplify clear epistemic principles that can be well formulated and rationally understood, as is the case in recent discussions aimed at extending the notion of mechanistic explanation to systems biology. Our intention has been to draw attention to cases that do not clearly exemplify such principles, yet nonetheless form a substantial part of knowledge-making practices in systems biology. We have illustrated practices that depart from the goal of dynamic mechanistic explanation for the sake of more limited modeling goals, using two representative cases from our ethnographic studies of model-building in systems biology. We have shown the centrality of parameter-fixing processes in modeling large-scale dynamical systems and the important role that abstraction techniques play in these departures. We characterized the practices as a pragmatic response to the constraints modelers of large-scale systems face, manifested in the pragmatic forms of understanding they settle for, especially as related to predictions about how to manipulate and control their systems. We hope our analysis will contribute to the development of a more nuanced understanding of what the widely used notion of a systems-level understanding might mean in practical contexts.

Systems biology itself is but one field in which computational and mathematical techniques are opening up the possibility for more pragmatic model-building through parameter-fixing processes and other abstractions that favor specific mathematical or computational results over representational accuracy of the models used to achieve these results. Our analysis aims to motivate investigation across computational science of these "somewhere in between cases." One thing that is needed is a broader and more detailed study of the affordances of parameter-fixing. Although parameter-fixing is often interpreted as a liability, it seems in certain circumstance to provide a flexible basis for extracting specific results from otherwise highly complex situations, as we have demonstrated.

Nothing we have said in the paper should be taken to contradict the normative value of mechanistic explanation. Concepts of dynamic mechanistic explanation, for instance, play an important role in our conceptualization of what our bottom-up modelers do by enabling us to frame their practices as departures from this standard of explanation and thus as a particular compromise between this standard and the broader goals of systems biology. It is reasonable to expect that more mechanistically accurate accounts should produce better more accurate and robust results. Our modelers often do consider their models as platforms for eventually building better mechanistic, and in turn more predictively accurate, accounts. That said one should not construe these model-building efforts as simply partial attempts on the way to mechanistic accounts. Systems biology is driven by specific mathematical goals and, as we have seen, these can be achieved to some extent through means other than accurate mechanistic representations. There is an undoubted tension between normative ideals and actual practices, and our modelers find themselves in the middle. What is of philosophical importance for those of us trying to understand systems biology and computational science more generally is to determine the precise nature of modeling practices and to account for how they can work to produce scientific understanding, even if they do not meet our current normative standards.

Acknowledgements

We appreciate the support of the US National Science Foundation in conducting this research (DRL097394084). Miles MacLeod's participation was also supported by a postdoctoral fellowship at the TINT Center, University of Helsinki. We thank the directors of Lab C and Lab G for welcoming us into the lab and the lab-members of those labs for granting us numerous interviews. We thank the members of our research group for contributing valuable insights, especially Vrishali Subramanhian, Lisa Osbeck, Sanjay Chandrasekharan, and Wendy Newstetter. We also thank the reviewers for their helpful advice and very thorough comments.

Alon, Uri. 2007. Network motifs: theory and experimental approaches. *Nature Reviews Genetics 8*(6): 450-461.

Bechtel, William. 2011. Mechanism and Biological Explanation. *Philosophy of science 78*(4): 533-557.

Bechtel, William, and Adele Abrahamsen. 2005. Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 36*(2): 421-441.

Bechtel, William, and Adele Abrahamsen. 2010. Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A 41*(3): 321-333.

Brigandt, Ingo. 2013. Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*. *44*(4), 477-492.

Calvert, Jane, and Joan H. Fujimura. 2011. Calculating life? Duelling discourses in interdisciplinary systems biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 42(2): 155-163

Chandraskeharan, Sanjay and Nersessian, Nancy J. forthcoming. Beyond correspondence: How the process of constructing models leads to discoveries and transfer in the engineering sciences. Special Issue: Modes, Media and Formats of Scientific Representation, T. Knuuttila and M. Vorms, editors. *Erkenntnis.*

Craver, Carl. F. 2006. When mechanistic models explain. *Synthese 153*(3): 355-376.

Craver, Carl. F. 2007. Explaining the brain: What a science of the mind-brain could be. New York: Oxford University Press.

Craver, Carl. F. (forthcoming). Explanation: The ontic conception. In eds. Andreas Hütteman & Marie Kaiser, *Explanation in the biological and historical sciences*. Berlin: Springer.

Fagan, Melinda. B. 2012. The joint account of mechanistic explanation. *Philosophy of Science 79*(4): 448-472.

Green, Sara, and Olaf Wolkenhauer. 2013. Tracing organizing principles-learning from the history of systems biology. *History and philosophy of the life sciences*. *35*(4), 553-576

Hood, Leroy, James. R. Heath, Michael. E. Phelps, and Biaoying Lin. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science Signaling: 306*(5696), 640.

Humphreys, Paul. 2004. *Extending ourselves: Computational science, empiricism, and scientific method*. New York: Oxford University Press.

Kitano, Hiraoki. 2002. Looking Beyond the Details: A Rise in System-Oriented Approaches in Genetics and Molecular Biology. *Current genetics 41*(1): 1-10.

Krohs, Ulrich., and Werner Callebaut. 2007. Data without Models merging with Models without Data. In *Systems Biology: Philosophical Foundations,* ed*.* Fred Boogerd, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyer, Hans V. Westerhoff, 181-213. Amsterdam: Elsevier

Lenhard, Johannes. 2006. Surprised by a nanowire: Simulation, control, and understanding. *Philosophy of Science 73*(5): 605-616.

Levy, Arnon, and William Bechtel. 2013. Abstraction and the Organization of Mechanisms. *Philosophy of Science 80*(2): 241-261.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of science* 67(1): 1-25.

MacLeod, M., & Nersessian, N. J. (2013a). Building Simulations from the Ground-Up: Modeling and Theory in Systems Biology. *Philosophy of Science, 80*(4), 533–556.

MacLeod, M., & Nersessian, N. J. (2013b). Coupling Simulation and Experiment: The Bimodal Strategy in Integrative Systems Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 572-584.

O'Malley, Maureen. A., and John Dupré. 2005. Fundamental Issues in Systems Biology. *BioEssays 27*(12): 1270-1276.

Parker, Wendy. S. 2010a. Predicting Weather and Climate: Uncertainty, Ensembles and Probability. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics 41*(3): 263-272.

Parker, Wendy. S. 2010b. Whose Probabilities? Predicting Climate Change with Ensembles of Models. *Philosophy of Science 77*(5): 985-997.

Savageau, Michael. A. 1969. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of theoretical Biology 25*(3): 365-369.

Sorribas, Albert., & Michael A. Savageau. 1989. Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. *Mathematical biosciences* 94(2): 239-269.

Westerhoff, Han. V., and Douglas. B. Kell. 2007. The Methodologies of Systems Biology. In *Systems Biology: Philosophical Foundations,* eds*.* Fred Boogerd, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyer, Hans V. Westerhoff, 23-70. Amsterdam: Elsevier.

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.
Voit, Eberhard. O. 2000. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge: Cambridge University Press.
Voit, Eberhard. O. 2013. *A First Course in Systems Biology*. New York: Garland Science.
Wolkenhauer, O., & Green, S. (2013). The search for organizing principles as a cure against reductionism in systems medicine. *FEBS Journal*. *280*(23), 5938-5948.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*:. Oxford: Oxford University Press.