sciendo

Causal Concepts Guiding Model Specification in Systems Biology

Dana Matthiessen
University of Pittsburgh

BIBLID [0873-626X (2017) 47; pp. 499–527]
DOI: 10.1515/disp-2017-0016

Abstract
In this paper I analyze the process by which modelers in systems biology arrive at an adequate representation of the biological structures thought to underlie data gathered from high-throughput experiments. Contrary to views that causal claims and explanations are rare in systems biology, I argue that in many studies of gene regulatory networks modelers aim at a representation of causal structure. In addressing modeling challenges, they draw on assumptions informed by theory and pragmatic considerations in a manner that is guided by an interventionist conception of causal structure. While doubts have been raised about the applicability of this notion of causality to complex biological systems, it is here seen to be an adequate guide to inquiry.

Keywords
Causation, systems biology, causal models, cellular networks, underdetermination

1 Introduction

Over the last two decades, theories and techniques of data-driven modeling, including causal modeling (cf. Spirtes et al. 1993, Pearl 2000), have become integrated into the study of complex biological systems. The growth of high-throughput data collection has made it necessary to develop sophisticated computational and statistical methods to illuminate patterns and underlying structures in a newfound wealth of information. In particular, researchers have developed algorithmic processes to infer networks of interaction among the components of a biological system. This approach to modeling

Disputatio, Vol. IX, No. 47, December 2017
Received: 05/09/2017 Accepted: 02/11/2017

biological networks has been characterized as a "top-down" approach, being opposed to a "bottom-up" approach that builds up a functional understanding of cells from a study of the interactions of constituent molecules (Westerhoff and Kell 2007).[1]

Some authors have responded to the proliferation of mathematical modeling in systems biology by arguing that much of the field is rooted in a general, non-causal and non-mechanistic form of understanding (Wouters 2007, Braillard 2010), but there are reasons to doubt the generality of such claims. For one, much of the study of cellular networks—a significant research program in systems biology—can be broadly understood within the framework of mechanistic science, as I have argued elsewhere (Matthiessen 2015). Second, the investigation of network structures is very often motivated and guided by a specific conception of their causal structure that accords with mechanistic inquiry (as described, for example, in Woodward 2013).

In what follows, I aim to analyze the strategies by which systems biology researchers refine and specify models in a highly data-driven context so as to extract informational structures designed to produce reliable predictions with respect to some phenomenon. I do not intend to show that all modeling efforts found under the wide-ranging banner of systems biology are fully compatible with the goals of causal and mechanistic explanations,[2] but instead to describe how methods and assumptions routinely employed in these data-driven contexts demonstrate a clear concern with capturing causal structure. Researchers explicitly interpret these models as bearing information about the causal structure of their target systems, and it is evident that a specific conception of causality is built into these interpretations—one that roughly corresponds to interventionist notions, which themselves might be thought to dovetail nicely with mechanistic inquiry (cf. Craver 2007, Woodward 2010). These assumptions in pursuit of specific causal information play an integral role in *model specification*, that is, the process by which researchers arrive at a model of a particular phenomenon or its underlying structure that

[1] This distinction cross-cuts with another useful distinction: that between molecular systems biology and systems theoretic systems biology (cf. De Backer et al. 2010).

[2] For a recent challenge to this claim that is highly attentive to the modeling practices of systems biologists, see MacLeod and Nersessian 2015.

includes a satisfactory amount of detail to aid in the explanation and prediction of experimental data.[3,4]

In section 2, I describe the aims of modeling in systems biology and present a basic sequence of stages of modeling specification through which these aims are realized. In section 3, I show how assumptions and complications arise in ways that are specific to each stage. In section 4, I describe the use of causal concepts in this process. Accounting for the stages of model specification is a fruitful way to examine the various modeling decisions and accompanying instances of inductive risk balancing[5] encountered by systems biologists along with the concepts that provide pragmatic footholds for such decisions, and I believe a comparable process can be observed in other scientific fields as well—in the investigation of the electronic structures of molecules and materials, for instance. With this in mind, I will conclude with some remarks on aspects of these strategies that serve to characterize the general epistemology of modeling, at least as it figures in data-driven contexts.

2 The aims and stages of modeling in "top-down" systems biology

There are many things that may count as a biological network. For the purposes of this paper, I will focus on models of what are called regulatory or signaling networks in individual cells. These are complex networks of interactions between various forms of macromolecules—primarily genes, proteins, transcribed RNA, and metabolites—that maintain the stability of a cell in response to its internal and external environment. In order to understand how cell networks function, researchers must first generate data. In one common technique, mRNA samples, often from of a single-cell organism like *E. coli* or *S. cerevisiae*, are extracted from cells exposed to experimental

[3] What counts as satisfactory is of course determined in some respects by modelers' purposes.

[4] This notion of model specification is partially inspired by the progressive concretization of modeling constructs described by McMullin (1985: Section 4).

[5] Cf. Douglas 2000.

conditions and combined with a luminous protein. These are placed onto microarrays—glass, plastic, or silicon chips that contain thousands of probes designed to detect specific mRNA sequences. Each site in the microarray contains a DNA sequence corresponding to a specific gene of the organism.[6] These are molecular complements to the distinct mRNA in the sample, which thus accumulate at the site of the gene from which they are transcribed. Robotics measure the luminosity of the mRNA at each site on a microarray, thereby obtaining a measure of the activity of their corresponding genes. Multiple parallel experiments may be carried out at once, yielding large quantities of data. In fact, so much data is produced that curated databases are used to store the results, but it is important to note that difficulties in accuracy attend to this process. Most databases are not designed to account for context-sensitive gene activity; high-throughput analyses often fail to detect rare events or unstable interactions; and the data available for model organisms usually address a small number of cell processes and experimental conditions (De Backer et al. 2010). There is a sense, then, in which systems biology is both data rich and data poor. The challenge for researchers is finding an appropriate way to infer an adequate network of interactions from data drawn from experiments, mined from databases, or sourced from extant publications.

Data generation, such as that described above, may be viewed as the first of a sequence of stages by which systems biologists arrive at a reliable representation of the phenomenon of interest. Figure 1 gives a highly schematic illustration of this sequence:

[6] In the case of *E. coli* and *S. cerevisiae*, the entire genome of the organism is known and so a complete measurement of gene expression is available by including all genes on the microarray.

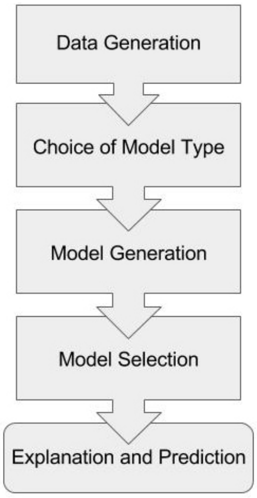

Figure 1: The stages of model specification in systems biology.

Each of these stages involves the incorporation of different assumptions drawn from theory or modeling practices, which aid in the eventual determination of a single representation of the network responsible for the data. Theoretical assumptions[7] tend to provide general guidelines and restrictions on the modeling practice, for instance by providing grounds for drawing inferences about a feature of a model from particular features of the data, or by supplying reasons for rejecting certain model types or tokens deemed to be biologically impossible. Modeling assumptions pertain more closely to the specific type of representation chosen, frequently reflecting pragmatic decisions made in the face of computational obstacles. As much of the recent literature on modeling and simulations has noted, modeling assumptions do not always acquire their warrant from theoretical commitments, but may instead be sanctioned on the basis of data-fitting calibrations and interventions, exploratory aims, or, as I mentioned, pragmatic decisions made in the face of limitations such as computational intractability (Cf. Cartwright et al. 1995, Morgan

[7] I'm not using 'assumptions' here in the sense of propositions that are entirely lacking in empirical or theoretical support. I only meant to indicate the way that they arise in the modeling context, that is, as constraints that are built into the model and so are assumed in its results.

and Morrison 1999, Winsberg 2010). Note, finally, that the progression through stages depicted here may not be passed through in a perfectly linear order: for instance, the models generated on the basis of a particular model choice may conflict so much with background knowledge or with model diagnostics that they force a return to a prior stage.

A substantial portion of the modeling strategies in systems biology and beyond consist of trade-offs between the computational opportunities afforded by particular assumptions or techniques and the inductive risks that accompany them.[8] For instance, one very broad assumption is built into the data generating technique described above: it is assumed that the presence of transcription factors such as mRNA bear a functional relationship to gene expression, and so measurements of mRNA are indirect measurements of gene activity. In a review of network modeling techniques, He et al. (2009) state that questions remain regarding the overall trustworthiness of this assumption. Thus the practice of modeling networks is carried out in the face of a number of uncertainties about its ultimate validity. Whether it ultimately stands or falls will depend on the extent to which its modeling assumptions are justified by background knowledge and mechanistic understanding of the biological underpinnings of cellular networks.

3 Stage-specific assumptions and trade-offs

3.1 Choice of model type

Having generated their data, researchers are first tasked with choosing a model type, that is, a general means of processing the data and representing the complex of biological mechanisms that give rise to it. In the case of gene expression profiling, a number of core methods available for this task involve building point-and-line graphs in which the nodes represent genes and the edges represent some form of dependency relation (derived algorithmically from the data) between these genes. Available dependency relations include differ-

[8] For a classic account of modeling trade-offs, see Levins 1966.

ential equations describing relationships between gene expression rates, or statistical methods such as measurements of the correlation coefficient between two genes, measurements of their mutual information, or measurements of the "similarity" (defined in one of numerous ways) between their expression patterns.

Any choice incorporates different biological assumptions that may limit the informativeness of the resulting graphical model. For instance, simple graphs called *co-expression networks* are built using the statistical correlation coefficient, which can measure the degree to which variable quantities change with one another, but is insensitive to non-linear dependencies. If the expression rate of one gene is actually a non-linear function of another, then this relationship will be missed by algorithms that build edges by means of correlation. In cellular networks with many interacting components, non-linear feedback relations are often encountered that render simple correlation-based models potentially unreliable. On the other hand, the use of Gaussian probability distributions to represent the state of a node—which is the main source of insensitivity to non-linearities—allows for the representation of expression levels as continuous values. Abandoning them for the sake of higher representational fidelity with respect to one network feature therefore requires coarse-graining another feature.

In an early proposal to move beyond co-expression networks, Friedman et al. write, 'Such analysis has proven to be useful in discovering genes that are co-regulated and/or have similar function. A more ambitious goal for analysis is to reveal the structure of the transcriptional regulation process' (Friedman et al. 2000: 602). Part of the motivation for finding alternatives to co-expression networks is due to the fact that they are underdetermined with respect to representations of regulatory interactions between genes, as exemplified by Figure 2:

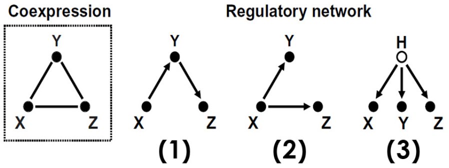

Figure 2: Correlational underdetermination of regulatory interactions.

The use of directed edges in Figure 2 is to show a direct regulatory effect of one gene on another. The figure illustrates that there are multiple possible regulatory relations explaining each co-expression measurement. With the undirected correlation graph on the left, one cannot distinguish between several possible regulatory networks that may have given rise to it; similar patterns in data could arise (1) from chaining, (2) from one gene regulating multiple others, or (3) from a common 'hidden' regulator. Thus a single co-expression network will give rise to a number of hypothetical regulatory structures that grows exponentially with the number of nodes.

Determining the actual regulatory network from expression data is highly non-trivial. In one of the most favored approaches, Bayesian network algorithms are used to encode further network structure and overcome some basic forms of model underdetermination.[9] Since only direct regulatory relationships are assumed to result, this technique allows for the construction of graphs with directed edges (i.e., arrows) showing determinate pathways of regulatory influence. The use of Bayesian networks is motivated in part by background knowledge of the structure of complex biological systems. For example, Sachs et al. write,

> There are several attractive properties of Bayesian networks for the inference of signaling pathways from biological data sets. Bayesian networks can represent complex stochastic nonlinear relationships among multiple interacting molecules, and their probabilistic nature can accommodate noise that is inherent to biologically derived data. They can describe direct molecular interactions as well as indirect influences that proceed through additional unobserved components, a property crucial for discovering previously unknown effects and unknown components. Therefore, very complex relationships that

[9] I will give special focus to Bayesian networks in the following sections. Here nodes X_i and X_j are only connected by an edge if their genes' activity is correlated *and*, knowing the behavior of all other genes and subsets of genes in the system, the behavior of X_i still yields additional information about X_j. That is, the condition for drawing an edge between nodes representing genes X_i and X_j is that they are correlated and:

$\sim(X_i \perp Xj) \mid X_S$ for all $S \subseteq V \backslash \{i, j\}$

where V is the complete set of nodes, S is a subset of nodes, and X_S is the collective activity of this subset.

> likely exist in signaling pathway architectures can be modeled and discovered (Sachs et al. 2005: 523–4).

Having a higher-resolution model of the regulatory interactions within a gene network clearly provides a more reliable tool for the prediction and discovery of further dependency relations within a cell. As the authors note, Bayesian networks are capable of representing this information in a way that accommodates additional understanding of the noisiness of biological measurements, the incompleteness of current knowledge of network components, and so on. The choice of model type is thus intertwined with the predictive aims of researchers and their established understanding of their subject matter drawn from related and overlapping research programs.

3.2 Model generation

Directed graph models of the entire network are generated by means of computational procedures designed to score them in terms of how likely they are to fit the data. A typical way of scoring the likelihood of a model G compares (i) the probability that one would observe the data set under consideration given the network topology[10] of G to (ii) the probability of seeing this data averaged over all possible models. For even the simplest metrics, the global problem of finding the best-fitting graph is NP-hard (Chickering 1996; with Heckerman and Meek 2004). Instead, researchers must employ search-and-score heuristics. Perhaps the simplest heuristic is a greedy search: starting from a prior graph representing minimal biological knowledge, different graphs that are 'nearby in search space' are tested by adding or removing single edges at different locations in the graph.[11] Each of these graphs is scored, the highest-scoring of them is selected as the new prior, and the process repeats. Each heuristic involves different trade-offs in false positives and false negatives, and their accuracy can be measured and compared by simulating data through an artifi-

[10] A network topology describes the general spatial characteristics of a graph—the average number of edges connected to a given node, whether it is fully connected or if there are disconnected sub-networks, etc.

[11] More complex concepts of neighborhood in search space can be employed instead, but difference by a single edge is perhaps the simplest and most intuitive option.

cial network and seeing how well each heuristic reproduces it (as in Yu et al. 2004). Finally, heuristics are also subjected to robustness analysis, where parameters such as quantity of data and data discretization are varied. In one such study, Bayesian network inference and scoring algorithms were found to perform best when the quantity of data greatly exceeds the number of genes being modeled, whereas information theoretic approaches perform better with fewer experiments, but are more prone to false positives in other cases (Bansal et al. 2007). Choice of model type will therefore depend not only on the type of information researchers seek to represent, but also likelihood of generating reliable models based on constraints such as the amount of available data. When resources do not permit a large number of experiments, Bayesian network algorithms may not provide the best results.

3.3 Model selection

Despite the high resolution of Bayesian networks and these scoring procedures, a vast number of data-accommodating networks can still be generated. For the majority of cases, the choice between high-scoring models is computationally underdetermined; for a given data set, available algorithms will not be able to decide between multiple regulatory structures. One reason for this is the aforementioned requirement of large amounts of data. In most cases, the number of genes is usually several orders of magnitude higher than the number of measurements taken to sample the data. This problem is commonly approached by drawing on further assumptions or pre-established knowledge of regulatory systems in order to compare models and narrow down the solution space; models may be 'filtered by making plausible assumptions on the objectives of the underlying system, such as economy of regulation (reflected by having the fewest edges that satisfy the conditions) or maximal biomass production' (Albert 2007: 3332). Over-fitting the data with an excessively powerful model is avoided by search algorithms that invoke a statistical form of Occam's razor. These favor less complex models that effectively predict limited ranges of data as opposed to highly complicated models that predict a wider range of data, but with lower accuracy (MacKay 1992). Such a process is supported by incorporating the belief that biological networks possess *sparseness*, meaning target genes can only be regulated by a limited number of transcription factors. By accounting for this and other properties such as scale-freeness, 'even an underdetermined system can be transformed into an over-determined one' (He et al. 2009: 200).

Authors attempt to give an even more accurate rendering of the actual network by integrating additional data about cell structures, such as analyses of gene location. These determine the DNA binding sites of proteins, providing physical evidence for regulatory relations between a gene that produces a given protein and those genes bound by the protein. Such information can be incorporated into model selection by selecting particular structure priors, or by giving no weight to models that fail to include edges required by location data. 'By fusing expression data with location data, the constrained search is able to consider statistical dependencies in the expression data that are consistent with the physical relationships already identified in the location data' (Hartemink 2002: 448).[12]

As with the choice of model type, the decisions encountered by researchers in the model generation and selection stages are sensitive to the general research context—the purposes of the researcher, the background knowledge available—and particular stage of model specification in which they arise. In generating candidate models, limitations in both computing power and the availability of experimental data require different choices to be made with regard to the search-and-score heuristics employed and their attendant modeling assumptions. In the selection stage, the space of possible models is narrowed down by drawing on theoretical assumptions informed, once again, by background knowledge of constituent mechanisms and processes established by 'nearby' fields such as cell biology, molecular biology, and biophysics.

This multi-stage process shares an important feature with other cases in which inductive risk balancing figures heavily: due to the sorts of limitations that accompany the data and model generating process, different purposes may cause researchers to make decisions that result in different end models. Between two researchers who begin with the same data, one who is highly cautious about false

[12] Here we see a merging of top-down and bottom-up approaches.

positives (say, because she is trying to figure out the functional role of a specific gene) will likely end up with a different graph than someone who is only interested in locating clustered 'modules' of genes that heavily regulate each other and interact much less with 'out-cluster' genes. Of course this does not entail that two graphs that disagree on whether some set of nodes are connected are both correct; there is little reason to assume that actual regulatory relations fluctuate as much as researchers' intentions. However, this serves to highlight the manner in which the data-heavy modeling of complex systems is accompanied by significant uncertainty: the definitive network structure, whatever it is, is buried beneath a compounding series of modeling assumptions, many of which enable researchers to gain traction in seeking reliable answers to certain questions while obscuring the answers to others. Often the most robust method for determining network structure as a whole involves finding effective ways to combine the results of multiple analyses—correlational, information theoretic, Bayesian probabilistic, and those based in differential equations (Le Novère 2015).

4 The use of causal concepts

4.1 Deriving causal structure from a Bayesian network

It is standard practice for modelers of biological networks to interpret directed graph edges as causal relations, where a given 'parent' node (at the origin of a directed edge) has a direct causal influence on its connected 'children'. This is seen in publications with titles that mention 'Bayesian inference for generating causal networks from observational biological data' (Yu et al. 2004) and direct claims like 'The subgraph consisting of all directed edges constitutes the inferred causal network' (Opgen-Rhein and Strimmer 2007). Indeed, one of the main features that Friedman et al. cite as an advantage of Bayesian networks over correlational graphs is the idea that 'Bayesian networks provide models of causal influence' (Friedman et al. 2000: 602).

Bayesian networks are primarily used for the purpose of modeling causal relations. This is in part because they incorporate assumptions that reflect certain intuitions about causality. The principle of *d-separation* is a prime example. A set of nodes Y is said to d-separate X from Z if and only if a node in Y intercepts every path from a node in X to a node in Z (Pearl 2000: 17). For example, if we take the chain graph (1) in Figure 2 to describe a causal network, then Y d-separates, or 'screens off', the causal relation between X and Z; any change in the value of X that affects the value of Z must flow through Y, and so learning about the value of Y renders knowledge of X irrelevant to determining the value of Z. Y appears to be a more direct influence on Z, which mediates the influence of X. To use a concrete example, suppose that the air conditioning in a room is connected to a thermostat device that turns the A/C on when the thermostat reaches a value over x^o, and that the A/C running makes the room become cool. In this case, if you notice that the A/C is running then finding out that the thermostat is over x^o tells you nothing *more* about the room becoming cool. The air conditioning mediates the influence of the thermostat on the room's temperature. In this way the interlinked conditional dependencies between the states of entities in our environment can be thought to reflect a structure of causal relations between them.

This relation between causal structure and conditional dependencies is most clearly captured in an assumption called the Causal Markov condition or CMC (Spirtes et al. 1993). A concise definition of this assumption is given by Woodward (2003):

> **(CMC)** For all Y distinct from X, if X does not cause Y,
> then $P(X|\text{Parents}(X)) = P(X|\text{Parents}(X) \cdot Y)$

In other words, the conditional independence relation in which the parents of a node d-separate it from all other predecessors is taken to be *entailed by* an underlying causal relation or lack thereof. Whether or not it is stated explicitly, researchers that understand the results of Bayesian network inference in causal terms must be taking this assumption on board. Here Friedman et al. are unambiguous:

> To learn about causality, we need to make several assumptions. The first assumption is a modeling assumption: we assume that the (unknown) causal structure of the domain satisfies the Causal Markov Assumption. Thus we assume that causal networks can provide a reasonable model of the domain [...] The second assumption is that there

are no *latent* or hidden variables that affect several of the observable variables (Friedman et al. 2000: 606).

Bayesian probabilities are typically understood in terms of an ideal epistemic agent's degrees of belief in some state of affairs. The joint probability distribution represented by a Bayesian network can thus be thought of as a model for how an ideal agents' beliefs about the states of components of a biological system should be interrelated. It is not clear that these networks license us to think that the system's behavior is inherently probabilistic; they do not clearly warrant the further step of a realistic interpretation of probabilities. When paired with the CMC, however, researchers can construe the properties of the network to correspond to some structural features of the biological system; that is, the CMC implies that there is some overlap between the structure of the joint probability distribution, and the structure of the actual system, taken to be causal and capable of generating probabilistic relationships in the data.

It may not be necessary for modelers to interpret causality in terms of a full-blown metaphysical realism, but the notion that target systems bear a causal structure that network modeling aims to identify at least serves as a kind of representational ideal for the practice. In this way, the modeling techniques seen in network inference may be understood as employing what Michael Weisberg calls minimalist idealization, a practice with a representational ideal that 'instructs the theorist to include in the model only the core or primary causal factors that give rise to the phenomenon of interest' (Weisberg 2013: 107). Levy and Bechtel likewise identify this ideal in network modeling, noting that abstract graph theoretic diagrams often help in determining the contributions of causal organization to system-level behaviors. They write, 'abstract models, such as models of connectivity [...] highlight the features of that specific system that make a difference in it—namely, its pattern of internal causal connections' (Levy and Bechtel 2013: 259).

The notion that cellular networks consist of an underlying structure of difference-making causes is seen to play a direct guiding role at one crucial stage of model selection for Bayesian networks. Regardless of the supplementary edge-pruning assumptions borrowed from background knowledge, Bayesian network selection suffers from an insurmountable form of underdetermination known as *Markov equivalence*. Markov equivalent networks share the same underlying graph, but the direction of their edges may differ (Figure 3).

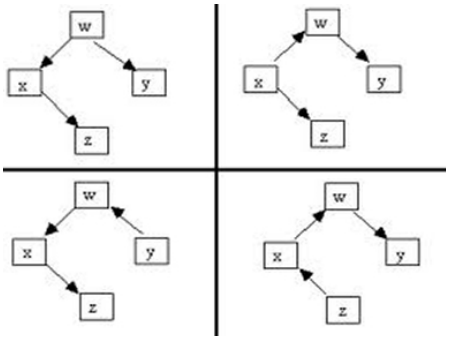

Figure 3: A set of Markov equivalent graphs.

Equipped only with observational data and Bayesian network inference, there is no principled reason on the basis of statistics to choose one of these graphs over the other. To a Bayesian algorithm, Markov equivalent networks are indistinguishable, which means the task of Bayesian network inference is best framed as a 'search for an equivalence class of networks that best matches [data] D' (Friedman et al. 2000: 604). Note, for example, that in all four graphs pictured, conditioning on X renders Z and W independent, but the precise reasons for this, taking causal relations into account, differ in each case. No matter what search heuristic is used, they will be unable to find a unique causal model; this underdetermination is strictly mathematico-computational in the sense that it is built into the algorithm of Bayesian network inference. Markowetz and Spang account for this as follows:

> Markov equivalence poses a theoretical limit on structure learning from data: even with infinitely many samples, we cannot resolve the structures in an equivalence class. In biological terms this means: even if we find two genes to be related it may not be clear which one is the regulator and which one is the regulatee. Without perturbation experiments this situation cannot be further resolved (Markowetz and Spang 2007).

In other words, obtaining a better representation of the causal structure of the system requires active intervention on the system.

There is a crucial connection between the CMC and the idea of intervention as a means of working around Markov equivalence. Friedman et al. comment:

> A causal network models not only the distribution of the observations, but also the effects of *interventions*. If X causes Y, then manipulating the value of X affects the value of Y. On the other hand, if Y is a cause of X, then manipulating X will not affect Y. Thus, although X→Y and Y→X are equivalent Bayesian networks, they are not equivalent causal networks (Friedman et al. 2000: 606).

In effect, it is the interpretation of Bayesian networks as representations of underlying causal structure licensed by the CMC that enables researchers to view statistically equivalent graphs as causally distinct; the directed edges of the network are taken to implicitly encode counterfactual information about the consequences of interventions on the system. Take, for instance, the graphs shown in Figure 3: if we interpret the directed edges to encode such information, then an intervention that only changes the value of Z will allow one to discern whether the bottom-right graph is the causal structure underlying the data, in which case we expect the value of X to change as well. So suppose the value of Z is altered, and that this results in no change in the value of X. Then the bottom-right graph is eliminated as a candidate for the underlying causal structure, whereas the remaining three must be narrowed down through interventions on other sites.

Markowetz and Spang (2007) cite studies showing that such 'interventions are critical for effective inference, particularly to establish directionality of the connections' in biological systems. An example of such an intervention in the context of regulatory networks is the use of gene knockout experiments. In one instance, Sachs et al. (2005) used small interfering RNA (siRNA) to target and silence the expression of a specific gene designated Erk in their regulatory network model. As indicated by the edges in Figure 4, their model predicted that this intervention would alter the expression of Akt but not PKA, and the result was seen to confirm these expectations, thereby validating the directionality of the edges inferred on the basis of prior data.

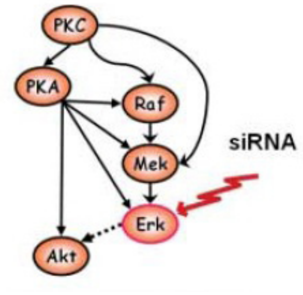

Figure 4: Intervention on a gene regulatory network.

Bayesian networks are interpreted as bearing interventional information on the basis of the CMC. Researchers employ interventions as a means of eliminating a subset of a group of equivalent graphs through the comparison of the results of interventions with those predicted by different causal structures. They can thus be seen to adopt a specific conception of causality, one that corresponds to the concept defended by Woodward:

> **(M)** X causes Y iff there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y or the probability distribution of Y would change (Woodward 2010: 290).

The meaning of 'cause' as it occurs in the antecedent of the CMC (see beginning of this section) is here elaborated in terms of shifts in the values of variables resulting from interventions on the system. Researchers use interventional relations between the expression levels of genes as a central indicator of causal relations between them. This definition does not give a reduction of the concept of cause to interventional relations because the notion of intervention is not clearly shorn of causal implications. However, the definition captures a basic operational interpretation of causality that can be productively employed in order to specify the structure of a given cell network. This interpretation is widespread in the literature on cellular networks; authors regularly view graph models constructed on the basis of interventional techniques as implicitly containing 'causal propositions that can be used to predict what is not yet known and that can be tested by experiment' (Peter and Davidson 2015: 265).

It's worth dwelling on the extent to which models of gene regulatory networks conform to a Woodward-style causal interpretation. Take, for instance, the description given by Isabelle Peter and Eric Davidson, who describe network graphs as

> literal representations of causal interactions among the regulatory genes in a network. These maps consist of the relevant regulatory genes (nodes or vertices), and they show explicitly the regulatory function these genes execute, i.e. they show for each gene how its outputs serve as inputs into other genes (linkages among genes, or "edges") (Peter and Davidson 2015: 267).

The use of 'literal' should not be taken to imply that edges in regulatory network graphs simply refer to the presence of a continuous physical process connecting two genes. Rather, they indicate the fact that alterations in the expression behavior of one gene make a difference in that of another. Regarding these edges, Peter and Davidson note that 'direct causal evidence is required to demonstrate the existence of a functional GRN [gene regulatory network] linkage' (Peter and Davidson 2015: 45), where causal evidence is primarily arrived at through experimental perturbations to a system. Possible perturbations include targeted gene mutations, knockouts, or expression silencing (as in the siRNA technique described above) that remove a transcription factor by which one regulatory gene affects another downstream. If the removal of a factor silences or otherwise alters the expression of a downstream gene, this constitutes 'direct causal evidence' of a relation between the two. Follow-up analyses, such as observing whether the factor in question is required at the site of the target gene in order for successful transcription to take place, can then be used to establish whether the causal relation between the relevant genes is direct (in which case a direct link between graph nodes is warranted) or indirect. Once again, interventions such as perturbations are held to be crucial in establishing the causal relations represented by the edges in a gene regulatory network model.

But researchers' confidence in the use of perturbative interventions for these purposes also reveals a deeper commitment to a conception of causality that accords with Woodward 2003. There, causal relations are subject to the further requirements of *stability* and *modularity*. Stability requires that the dependencies between nodes in a graph are invariant under some range of changes in background conditions. Modularity, on the other hand, requires that the disruption or alteration of the causal relation between a pair of nodes in the graph does not result in a reorganization of the causal relations between other nodes. Confidence in the results of a perturbation analysis assumes both; if a given perturbation were believed to alter background conditions in such a way that the causal relations between genes were significantly altered, either by changing their functional relationship (say, from exciting to repressing) or by triggering a reorganization of network components, then there would be no reason to think that perturbation experiments could yield information about the normal functioning of the network. Without stability and modularity, each perturbation would potentially give rise to a completely different organization among regulatory genes, and the goal of inferring how genes interact in the absence of such perturbations would be rendered nearly impossible for large networks.

4.2 Worries for those employing the interventionist framework

While the assumptions required to carry out causal modeling for cell networks are informed by biological background knowledge, they still carry the risk of glossing over important features of these systems. Perhaps more worrisome is the possibility that modelers are simply working with a deficient notion of 'cause', which causes them to systematically ignore relevant causal relations. In short, a number of theoretical and practical challenges confront this modeling paradigm, some of which have received significant attention within the philosophy of causality and philosophy of science. Although an attempt to fully respond to each issue is beyond the scope of this article, there are reasons to think that biologists are not misguided in their use of the above techniques and assumptions.

On the theoretical end, causal modelers are faced with stances critical of probabilistic causal theories as a whole. A number of philosophers, most notably Nancy Cartwright (1993, 2002, 2007, and more) have denied that the Causal Markov Condition is necessary for inferring causal relations. Cartwright has presented two main cases that violate the CMC: one involving the probabilistic decay of one particle into a particle pair, another involving the probabilistic production of by-products alongside the products of a chemical plant. Conditioning on the state of the particle or plant does not render the relation between the particle pair or chemical by-products statistically independent, despite the prior state's role as their apparent common cause. In the quantum case, this may be grounds for accepting that common notions of causal relations simply do not apply at the level of fundamental physics. This does not mean that such concepts are inapplicable or invalid at other scales such as those relevant to biological systems, just that they describe patterns that are not 'fundamental' or are not always found in certain lower-scale domains like quantum physics. The case of the chemical plant can then be considered independently. Without going into great detail, the argumentative success of this case can be seen to depend on the assumption that a finer-grained account of the production process would not be capable of locating a component that successfully screens off the correlation between the products. This assumption is at the very least questionable, and the reader is referred to the exchange between Cartwright and Hausman and Woodward (1999, 2004) for further details. For our current purposes, the important question is not whether the CMC is necessary to discover any and all causal relations, but whether it makes sense for modelers of gene regulatory networks to assume it, and this depends on whether it allows researchers to reliably make predictions about and explain the behavior of the interaction systems under scrutiny. A definitive judgment on this matter is premature, but the continued use of Bayesian networks among systems biologists suggests that the CMC continues to bear fruit, and so there is reason to believe it holds in the systems under study.

Even if the CMC is a necessary criterion for the discovery of causal relations, it may fail to realistically apply as an assumption about the structure of cellular networks. In fact, there is a reason to think it is ill-applied. The CMC gives an interpretation of the directed acyclic graphs arrived at through Bayesian modeling algorithms, but cyclical interactions between components are incredibly common in regulatory networks; many include numerous network motifs such as feedback loops that help maintain the system in a steady state against external or internal perturbations (cf. Alon 2006). It is possible, however, to work around this issue. One way that the problem of cyclical interactions can be addressed is through the use of temporal data, which provides measures of the expression rates of genes over time (He et al. 2009). These data can be used to construct dynamic Bayesian networks, where each node stands for an expression rate at a specified time. Cyclical interactions between genes will then be represented by a reversal of directed arrows between their corresponding nodes, where these nodes bear successive time stamps. In this way, cyclical interactions in the data can be 'unfolded' into acyclical graphs in a way that allows for the retention of the CMC and may even provide more detailed information such as the rates at which different processes feed back.

Another challenge for modelers working under this framework are possible violations of modularity. Sandra Mitchell (2008) has raised questions regarding the applicability of the modularity condition to biological networks, noting that they may be organized in such a way that, when the activity of a given component C is disabled, alternate components are able to compensate for its absence and produce the same effect E that was originally attributed to the absent component. Such a self-reorganizing network would appear to violate the modularity condition, and thus the interventionist's difference-making criterion for the claim that C causes E. Likewise, Markowetz and Spang (2007) cite compensatory network activity and uncertainty about the exact size of perturbation effects as obstacles to Pearl's notion of single-variable manipulation—a notion that closely resembles Woodward-style intervention. But again, these problems are not strictly insurmountable: Markowetz and Spang also note various techniques being developed to overcome such difficulties, including what they call 'soft interventions,' analyses of gene knock-out data, and the reverse engineering of regulatory pathway structure through the observation of nesting patterns in the results of interventions.

Where interventions on single genes are unreliable or result in the sort of reorganization that Mitchell warns of, they may still be accurately approximated by adopting coarser-grained notions of modularity, that is, by shifting the level of description at which stable causal relations are found and allowing for perturbations that may affect multiple genes. Researchers can detect particular sub-networks that are strongly connected, allowing for a distinction between in-cluster (nodes that can influence the sub-network without being influenced by it) and out-cluster (the converse). 'Nodes of each of these subsets tend to have a shared task; for example, in signal transduction networks, the nodes of the in-cluster tend to be involved in ligand-receptor binding; the nodes of the strongly connected cluster form a central signaling subnetwork; and the nodes of the out-cluster are responsible for the transcription of target genes and for phenotypic changes' (Albert 2007: 3332). Building on this approach, Bansal et al. claim that Bayesian network inference is useful for 'identifying functional modules, that is, identifying the subset of genes that regulate each other with multiple (indirect) interactions, but have few regulations to other genes outside the subset' and for 'predicting the behavior of the system following perturbations [say, through gene knock-outs or altering expression levels], that is, gene network models can be used to predict the response of a network to an external perturbation and to identify the genes directly "hit" by the perturbation' (Bansal et al. 2007: 1). A greater degree of invariance to intervention is therefore likely to be found in the relations between functional modules, permitting more robust predictions of perturbation effects. In this way, there is a close connection between talk of functional modules among gene network scientists and a system's possessing modularity in Woodward's sense.

Just as with the CMC, Cartwright has argued against interventionist claims that the concept of causal relations requires modularity (see her 2007: ch. 7). For our purposes, the matter is once again not whether modularity applies in all cases of causal inquiry, but whether its assumption yields reliable information about the relevant target systems. In practice, researchers appear comfortable with the risks of assuming a substantial degree of modularity. For some, such as Peter and Davidson (2015), invocations of modules are part of the basic description of gene regulatory network structure. According to them, the expression of individual genes is controlled by sequences residing on the same DNA molecule. These '*cis*-regulatory modules' interact with transcription factors to define the conditions under which a given gene is expressed. They do so by acting, for example, as cofactors that determine where in the developmental plan the transcription of a gene is initiated, or by isolating it from other regulatory domains and preventing it from being transcribed when certain other genes are active. The specificity of transcriptional control due to *cis*-regulatory modules serves to insulate genes from influence by factors other than those that commonly affect them, lending gene regulatory networks a higher degree of modularity than would be expected otherwise. The authors then specify a further level of modularity in these networks due to the presence of 'subcircuits'. These, like the 'network motifs' and 'functional modules' referenced earlier, are highly recurrent patterns of connections between regulatory genes, such as feed-back loops, which serve to coordinate the joint expression of several genes in a way that carries out a distinct function. Peter and Davidson elaborate: 'A given developmental GRN will include several separate subcircuits joined by encoded regulatory linkages. Thus, considered from the perspective of the structural elements that perform its overall control functions, the developmental GRN has a modular character' (Peter and Davidson 2015: 44). If we assume that the large numbers of interactions among sub-circuit elements and sparser interactions between separate sub-circuits protects sub-circuit interactions from being affected by perturbations of individual sub-circuit components, then this 'modular character' instantiates causal modularity between the network sub-circuits.

5 Multi-stage model specification in systems biology and beyond

I hope to have shown that the practice of modeling cellular networks using Bayesian inference techniques is guided by a notion of underlying causal structure that corresponds to a Woodward-style interventionist conception of causality. How, then, should we characterize the role causal concepts have played in this process of model specification? Recall that model specification is a procedure by which researchers arrive at a satisfactory amount of detail in their representation of some phenomenon of interest or its underlying structure. In the case of cellular networks, we have seen how certain levels of detail may be obscured on the basis of model underdetermination; for example, if a given strategy is unable to specify whether arrows between nodes are directed one way or another, then the most accurate representation available will have to leave this feature undetermined. Figure 5 shows a schematic depiction of the role of causal concepts in model specification, which involves overcoming the stages of underdetermination discussed above. Here model specification is achieved through the incorporation of compounding background theoretical assumptions which aid in the elaboration of a particular notion of underlying causal structure.

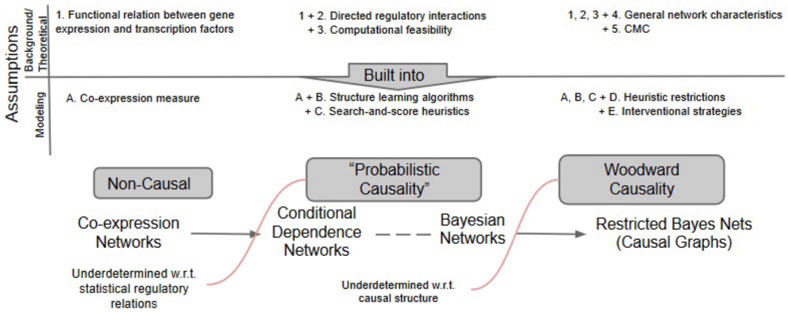
Assumptions
Background/Theoretical
Modeling
1. Functional relation between gene expression and transcription factors
1 + 2. Directed regulatory interactions + 3. Computational feasibility
1, 2, 3 + 4. General network characteristics + 5. CMC
Built into
A. Co-expression measure
A + B. Structure learning algorithms + C. Search-and-score heuristics
A, B, C + D. Heuristic restrictions + E. Interventional strategies
Non-Causal
"Probabilistic Causality"
Woodward Causality
Co-expression Networks
Conditional Dependence Networks
Bayesian Networks
Restricted Bayes Nets (Causal Graphs)
Underdetermined w.r.t. statistical regulatory relations
Underdetermined w.r.t. causal structure

Figure 5: Causal concepts in network model specification.

This graphic helps show how it is the notion of underlying causal structure that helps guide the process. Otherwise it becomes hard to make sense of the idea that a model is underdetermined. That is, we can always ask 'underdetermined with respect to what?' Co-expression networks account for the data as well as other networks, in the sense that they are constructed through the direct incorporation of the available measurements. It is only because researchers, motivated by their understanding of the relevant biological systems, posit a more fine-grained structure, one resembling a network of Woodward-style causal relations, that such correlation networks are viewed as insufficient. A supposition of causal regulatory structure thus guides the development of strategies to represent 'deeper' structures in the same data, supplemented with further assumptions and, eventually, interventional results.

In this way the modeling practice of researchers of cellular networks in systems biology can be seen to involve the multi-stage specification of an adequate representation of its subject matter. Model specification is achieved through a compounding series of theoretical and modeling assumptions, which serve to elaborate and refine an informational structure designed to produce reliable predictions with respect to some phenomenon of interest. In reviewing the features of this practice, I seek to describe them with enough generality that they may be recognized in other areas of science, as I believe they can be.

Researchers begin with the most widely adopted modeling assumption to aid in simplifying impending computational tasks (e.g., measurements of transcription factors have a functional relationship to gene expression; co-expression is a reliable indicator of possible regulatory interactions between genes). This is followed by the choice of specific mathematical objects (e.g., graphs of conditionally dependent data points), which pick out, from the available ways of inquiring into the system under study, a specific informational structure that is relevant to researchers' aims (e.g., such graphs better predict downstream results of gene knock-outs). More detailed features of these mathematical objects and their sub-components are then specified, again with reference to the aims of research, with trade-offs being made by researchers based on the particular problem-solving context in which the modeling effort takes place (e.g., a binomial form for probability distributions is more computationally efficient than multinomial distributions, but may fail to reliably represent regulatory feedback loops; a search heuristic that penalizes high-scoring graphs for complexity or for over-fitting of data risks rejecting accurate models of highly involved networks).[13]

These stages are accompanied by a form of computational opportunism, in which concerns about finding 'one true representation' may be overshadowed by an interest in selecting from a menagerie of tweaks and variations on model sub-components suited to serve different purposes. Such strategies are also seen to be rooted in a local domain of inquiry embedded within a collection of ongoing peripheral research programs. This embedding in a local domain provides crucial contextual features that orient scientific problem-solving in the form of established empirical facts that justify basic modeling assumptions. It also provides peripheral research programs and sources of background theory for modelers to draw on and refine their results (e.g., incorporating gene location data into search heuristics).

Finally, the use of concepts relevant to the formulation and execution of strategies is seen to be highly purpose-driven, and is in many

[13] For another account of multi-stage model construction, instead from the perspective of 'bottom-up' systems biologists, see MacLeod and Nersessian 2013.

ways a function of the modeling context. The interventionist notion of causation proves to be a highly useful means to further specifying regulatory network structures, as it gives researchers a way to differentiate statistically indistinguishable graphs. This notion of causation also plays something of a guiding role for modelers: without it, it is hard to make sense of the idea that a model is underdetermined. That is, it is because researchers posit a more fine-grained structure—one that may be approximated by a system of Woodward-style causal relations—and because they seek the kind of inferential reliability that such a structure brings, that representations such as co-expression networks are viewed as insufficient depictions of the sources of experimental data.

By paying attention to these features of the model specification process—the stages of decision-making, the research context, the driving concepts—we gain a better understanding of scientific modeling practices. In addition, it allows us to see how the inductive risks that accompany various assumptions can be localized and at times individually examined, rather than attributed wholesale, say, to the finished product or to the very act of inductive inference.[14]

Dana Matthiessen
University of Pittsburgh
Dietrich School of Arts and Sciences
Department of History and Philosophy of Science
1017 Cathedral of Learning
4200 Fifth Avenue
Pittsburgh, PA 15260
dam228@pitt.edu

References

Albert, Réka. 2007. Network inference, analysis, and modeling in systems biology. *The Plant Cell* 19: 3327–38.

Bansal, Mukesh; Belcastro, Vincenzo; Ambesi-Impiombato, Alberto; and di Bernardo, Diego. 2007. How to infer gene networks from expression

[14] Thank you to the organizers and attendees of CaMitS 2015 and to two anonymous reviewers for stimulating comments and helpful suggestions on earlier versions of this paper.

profiles. *Molecular Systems Biology* 3: 1–10.

Braillard, Pierre-Alain. 2010. Systems biology and the mechanistic framework. *History and Philosophy of Life Sciences* 32: 43–62.

Cartwright, Nancy. 1993. Mark and probabilities: two ways to find causal structure. In *Scientific Philosophy: Origins and Development*, ed. by F. Stadler. Dordrecht: Kluwer.

Cartwright, Nancy. 2002. Against modularity, the causal markov condition and any link between the two: comments on hausman and woodward. *British Journal for the Philosophy of Science* 53: 411–53.

Cartwright, Nancy. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. New York: Cambridge University Press.

Cartwright, Nancy; Shomar, T.; and Suárez, M. 1995. The tool box of science. In *Theories and models in scientific processes*. Amsterdam: Rodopi.

Chickering, David Maxwell. 1996. Learning bayesian networks is NP-complete. In *Learning from Data: Artificial Intelligence and Statistics V*. New York: Springer-Verlag.

Chickering, David Maxwell; Heckerman, David; and Meek, Christopher. 2004. Large-sample learning of bayesian networks is NP-Hard. *Journal of Machine Learning* 5: 1287–330.

Craver, Carl. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

De Backer, Phillipe; De Waele, Danny; and Van Speybroeck, Linda. 2010. Ins and outs of systems biology vis-à-vis molecular biology: continuation or clear cut? *Acta Biotheoretic* 58: 15–49.

Douglas, Heather. 2000. Inductive risk and values in science. *Philosophy of Science* 67: 559–79.

Friedman, Nir; Linial, Michal; Nachman, Iftach; and Pe'er, De'er. 2000. Using bayesian networks to analyze expression data. *Journal of Computational Biology* 7: 601–20.

Hartemink, Alexander J.; Gifford, David K.; Jaakola, Tommi S.; and Young, Richard A. 2002. Combining location and expression data for principled discovery of genetic regulatory network models. In *Pacific Symposium on Biocomputing 2002: Kauai, Hawaii, 3-7 January 2002*: 437–49.

Hausman, Daniel W.; and Woodward, James. 1999. Independence, invariance and the causal markov condition. *British Journal for the Philosophy of Science* 50: 521–83.

Hausman, Daniel W.; and Woodward, James. 2004. Modularity and the causal markov condition: a restatement. *British Journal for the Philosophy of Science* 55: 147–61.

He, Feng; Balling, Rudi; Zeng, An-Ping. 2009. Reverse engineering and verification of gene networks: principles, assumptions and limitations of present methods and future perspectives. *Journal of Biotechnology* 144: 190–203.

Le Novère, Nicolas. 2015. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews: Genetics* 16: 146–58.

Levins, Richard. 1966. The strategy of model building in population biology.

American Scientist 54: 421–31.

Levy, Arnon; and Bechtel, William. 2013. Abstraction and the organization of mechanisms. *Philosophy of Science* 80: 241–61.

MacKay, David J. C. 1992. Bayesian interpolation. *Neural Computation* 4: 415–47.

MacLeod, Miles; and Nersessian, Nancy. J. 2013. Building simulations from the ground up: modeling and theory in systems biology. *Philosophy of Science* 80: 533–56.

MacLeod, Miles; and Nersessian, Nancy J. 2015. Modeling systems-level dynamics: understanding without mechanistic explanation in integrative systems biology. *Studies in History and Philosophy of Science Part C—Biological and Biomedical Science* 49: 1–11.

Markowetz, Florian; and Spang, Rainer. 2007. Inferring cellular networks—a review. *BMC Bioinformatics* 8 (Suppl 6) S5. <www.biomedcentral.com/1471-2105/8/S6/S5>

Matthiessen, Dana. 2015. Mechanistic explanation in systems biology: cellular networks. *British Journal for the Philosophy of Science*. <http://bjps.oxfordjournals.org/content/early/2015/04/09/bjps.axv011.full.pdf+html>

McMullin, Ernan. 1985. Galilean idealization. *Studies in the History and Philosophy of Science* 16: 247–73.

Mitchell, Sandra. 2008. Exporting causal knowledge in evolutionary and developmental biology. *Philosophy of Science* 75: 697–706.

Morgan, Mary; and Morrison, Margaret. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Opgen-Rhein, Rainer; and Strimmer, Korbinian. 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology* 1. <www.biomedcentral.com/1752-0509/1/37>.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Peter, Isabelle S.; and Davidson, Eric H. 2015. *Genomic Control Process: Development and Evolution*. New York: Elsevier.

Sachs, Karen; Perez, Omar; Pe'er, Dana; Lauffenburger, Douglas A.; Nolan, Garry P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308: 523–9.

Spirtes, Peter; Glymour, Clark; and Scheines, Richard. 1993. *Causation, Prediction, and Search*. New York: Springer-Verlag.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, James. 2010. Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology and Philosophy* 25: 287–318.

Woodward, James. 2013. Mechanistic explanation: its scopes and limits. *Proceedings of the Aristotelian Society* S87: 39–65.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.

Westerhoff, Hans V.; and Kell, Douglas B. 2007. The methodologies of systems biology. In *Systems Biology: Philosophical Foundations*. New York: Elsevier.

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

Wouters, Arno G. 2007. Design explanation: determining the constraints on what can be alive. *Erkenntnis* 67: 65–80.

Yu, Jing; Smith, V. Anne; Wang, Paul P.; Hartemink, Alexander J.; and Jarvis, Erich D. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20: 3594–603.