DRAFT—PLEASE DO NOT CITE. COMMENTS ARE WELCOME

Paper to be presented at the PSA 2014, Chicago, 6-8 November 2014

Symposium: "How Adequate Are Causal Graphs and Bayesian Networks?"

***

On the Incompatibility of Dynamical Biological Mechanisms and Causal Graph Theory

Marcel Weber
Department of Philosophy
University of Geneva
marcel.weber@unige.ch

Abstract

I examine the adequacy of the causal graph-structural equations approach to causation for modeling biological mechanisms. I focus in particular on mechanisms with complex dynamics such as the PER biological clock mechanism in *Drosophila*. I show that a quantitative model of this mechanism that uses coupled differential equations – the well-known Goldbeter model – cannot be adequately represented in the standard (interventionist) causal graph framework, even though this framework does permit causal cycles. The reason is that the model contains dynamical information about the mechanism that concerns causal properties but that does not correspond to variables that could be subject to independent interventions. Thus, a representation of the mechanisms as a causal structural model necessarily suppresses causally relevant information.

1. Introduction

Recent decades have seen the advent of elaborate formal techniques for causal modeling (Spirtes, Glymour, and Scheines 2000; Pearl 2000). These techniques, which essentially link causality to manipulability, have been instrumental in taking philosophical debate about causation as well as about scientific explanation to a new level (e.g., Woodward 2003, 2011; Woodward and Hitchcock 2003; Hitchcock and Woodward 2003; McKay Illari, Russo and Williamson 2011). Furthermore, this formal approach to causality has been productively applied in order to analyze causation in specific scientific disciplines. Originally developed mainly in the context of econometrics, it was recently also applied to various other sciences, e.g., neuroscience (Craver 2007, Weber 2008), genetics (Waters 2007; Woodward 2010), evolutionary theory (Otsuka forthcoming), psychiatry (Woodward 2008), or public health policy (Russo 2012), to name just a few.

The basic tools of this approach are the formally definable concepts of directed acyclic graph (DAG), Bayesian network, and structural equation. In the standard approach, the causal interpretation of these formal concepts is provided by means of the concept of idealized intervention. The result are causal models that contain information about counterfactual dependencies between a set of variables as well as, in some cases, probability distributions defined over these variables.

While the fruitfulness of this approach to causal modeling in scientific practice as well as for philosophical analysis is beyond doubt, there have not been many attempts to explore its limits in adequately representing causal systems. There has, of course, been

quite some debate concerning the question of whether a certain conception of *mechanism* is adequate for explaining biological phenomena (e.g., Bechtel 2005, 2013; Bechtel and Abrahamsen 2010; Braillard 2010; Kuhlmann 2011; Waskan, 2011; Weber 2012; Dupré 2013; Woodward 2013). This debate focused on the issue of whether the standard conceptions of mechanism can account for biological processes that feature complex dynamical behavior and/or spatial structures. However, none of this work has directly challenged the underlying interventionist theory of causation itself. In fact, there is a whole range of more recent studies that attempt to show that Bayesian networks are actually adequate for modeling complex biological mechanisms (Casini et al. 2011; Clarke, Leuridan and Williamson 2014; Gebharter 2014; Gebharter and Kaiser 2014; Gebharter and Schurz, this symposium; Casini and Williamson, this symposium).

In part, this problem turns on the question of how narrowly the term "mechanism" should be understood (Woodward 2013). In this paper, I will not be concerned with this issue. Rather, I want to examine to what extent the contemporary interventionist approach to causality is apt for representing the causal properties of a certain kind of mechanism in the first place.

A critical issue will be the extent in which causal models that basically contain causal difference-making information can account for the dynamics and for spatial features of mechanisms, as such features are absolutely crucial for the explanatory force of many mechanisms, in biology and elsewhere. Woodward (2013) has argued that spatio-temporal information can always be integrated with the causal difference-making information

contained in causal models. While this may be true in some sense, I will show that it glosses over a basic problem pertaining to the dynamics of certain kinds of causal system.

I will closely examine an example from biology that involves a mechanistic model consisting of a system of coupled differential equations with complex dynamics. This model describes the operation of a biological clock. I will assume without further argument that this model captures the essential causal properties of the biological clock mechanism, at least with respect to certain explanatory goals.[1] Then, I will show that formal causal models fail to correctly represent these causal properties. Specifically, I will argue that such a model will not be able to treat time derivatives as causally relevant variables.

I shall proceed as follows. In Section 2, I shall briefly review the core notions used in the causal modeling literature, in particular the notions of causal graph, structural equations, and ideal intervention. In Section 3, I analyze a dynamical model of a biological clock mechanism and show that it has no adequate causal graph representation. In Section 4, I consider some attempts from the current causal modeling literature to represent differential equations in structural causal models. I show that the results coming from these

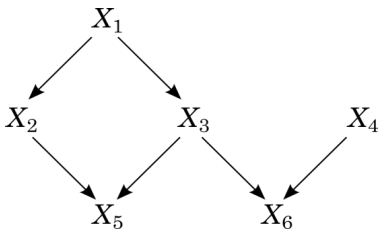
[1] I am not assuming that there is just one correct way of representing a causal system. Thus, I accept the pluralist thesis according to which there is always a variety of different perspectives on the world none of which succeeds in providing a complete picture (Kellert, Longino and Waters 2006; Dupré 2013). In fact, I suggest that my arguments presented here could be used for actually defending such a strong form of scientific pluralism, but this would go beyond the scope of this paper.

attempts actually support my thesis. Section 5 summarizes and integrates my conclusions with regard to the limitations of causal modeling.

2. Causal Modeling

The formal concepts used in causal modeling include directed acyclic graphs (DAGs), structural equations and Bayesian Networks. In this paper, I shall focus on DAGs and structural equations and leave Bayesian networks aside, but it should be noted that any problem concerning DAGs will also affect causally interpreted structural equations as well as Bayesian networks because the latter two kinds of causal models contain DAGs.[2]

A DAG is an ordered pair $\langle V, E \rangle$, where V is the set of variables and E is a set of directed edges.



A DAG becomes a causal graph as soon as its edges are interpreted causally, about which I will say a little more below.

Causal dependencies can also be represented by using so-called *structural models* (Pearl 2000). Such a model consists of an ordered triple $\langle U, V, Q \rangle$ where U is a set of exogenous variables, V a set of endogenous variables, and Q a set of *structural equations*. The structural equations give the value of each endogenous variable as a function of the

[2] I wish to thank Lorenzo Casini for pointing this out to me.

values of other variables in *U* and *V*. The variables may also be interpreted as nodes that are connected by causal arrows. But in contrast to pure causal graphs, the structural equations also provide quantitative information as to how much some dependent variables change per unit change of the independent variable.

Pearl (2000, p. 160) gives the following "operational" definition of a structural equation:

> *An equation* $y = \beta x + \varepsilon$ *is said to be structural if it is to be interpreted as follows: In an ideal experiment where we control* X *to* x *and any other set* Z *of variables (not containing* X *or* Y*) to z, the value* y *of* Y *is given by* $\beta x + \varepsilon$*, where ε is not a function of the settings* x *and* z*.*

According to this definition, it is obvious that structural equations *sensu* Pearl are *linear* equations in the sense of not containing derivatives of the variables. As we shall see, this feature constitutes a major limitation when it comes to modeling systems with complex dynamics.

Pearl's definition of a structural equation contains the idea of an "ideal experiment". This notion has been elaborated in great detail by Woodward (2003, 94-99), who defines it in terms of the notion of ideal intervention. On this account, an (ideal) intervention on some variable *X* with respect to some variable *Y* changes *Y* by changing *X* without changing any other variable that is a cause of *Y*.

In a nutshell, these are the basic concepts of causal modeling. Thus, when I speak about a "causal model" in what follows, I mean a model that is expressed by using either causal graphs or structural equations and that uses an interventionist criterion for

interpreting the graphs and equations causally. The goal of this paper (as well as the paper by Kaiser, this symposium) is to show that these concepts fail to fully account for certain *causal* explanations in biology.

In the following section, I show what problems are created for causal models by complex dynamical information. Kaiser (this symposium) does the same for spatially complex mechanisms. Thus, while Kaiser's paper is about space, this one is about time.

3. It's About Time: Modeling Dynamic Processes

3.1. Classic Examples of Dynamic Models in Biology

There is an important class of biological models that try to account for complex series of events in dynamical terms. A classic example is the Hodgkin-Huxley model of the action potential (see, e.g., Weber 2005, 2008). This model (henceforth HH model) shows how changes in membrane conductance generate a temporary membrane depolarization that can spread along an axonal membrane and thus form the basis of information processing by neurons. A more recent example is Goldbeter's (1995) model of the circadian oscillations of the PER protein in *Drosophila*, which is the heart of a biological clock mechanism. There are many more such models, but for the purposes of this paper we shall concentrate on these two.

3.2. Bechtel and Abrahamsen on Dynamic Mechanistic Explanation

In a recent series of papers, Bill Bechtel and Adele Abrahamsen have provided a very illuminating account of models and mechanisms in circadian clock research, including

Goldbeter's model and the PER mechanism (Bechtel and Abrahamsen 2010, Bechtel 2013). Their account will prove to be useful for our analysis, which is why it will be briefly reviewed here. We take the gist of their account to be that circadian clock models provide what they call *dynamic mechanistic explanations*. According to Bechtel and Abrahamsen, such explanations differ from other kinds of mechanistic explanations in providing *quantitative* information about the behavior of the systems in question. Dynamic mechanistic explanations (may) contain *sequential mechanistic models* that describe a series of events in purely *qualitative* terms. Figure 1 shows such a sequential mechanistic model.

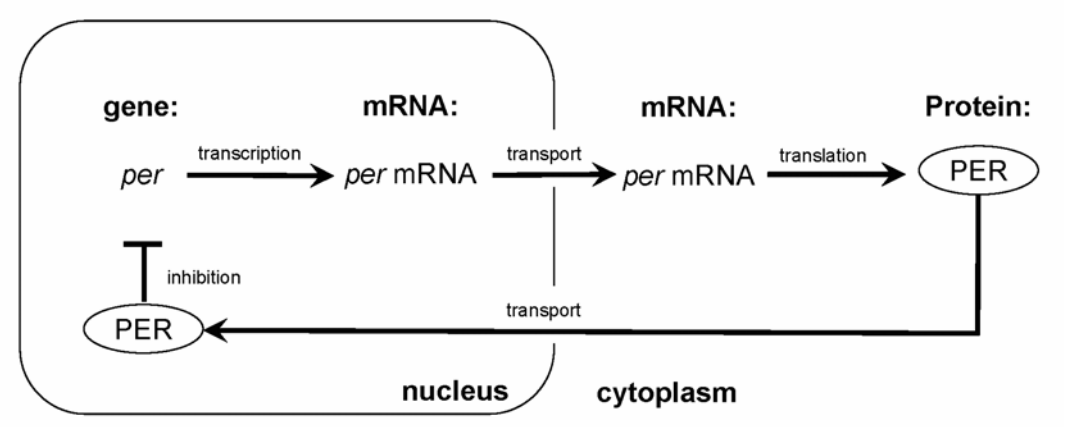


Figure 1. The sequential mechanism of the *Drosophila* circadian clock gene *period*. After Hardin et al. (1990).

An interesting feature of this sequential model according to Bechtel and Abrahamsen is the fact that it is possible to *mentally rehearse* the individual steps as well as their temporal arrangement.

But the most important claim made by Bechtel and Abrahamsen for our purposes is the following: The qualitative sequential model as shown in Figure 1 is *incomplete*. For what the model must show is that the circadian system is capable of generating *stable* oscillations. This is where the dynamical, quantitative model constructed by (Goldbeter 1995) comes in. The model describes the change in cytoplasmic concentrations of PER mRNA (M) as well as the different phosphorylation states of cytoplasmic (P_0, P_1, P_2) as well as nuclear (P_N) PER protein with the help of differential equations. The model uses standard Michaelis-Menten enzyme kinetics where the V_i are maximal reaction rates and the K_i the so-called Michaelis constants for the different biochemical reactions involved (the Michaelis constant gives the substrate concentration at which the reaction rate is half the maximal rate).

The structure of the dynamical model can be extracted from Figure 5.

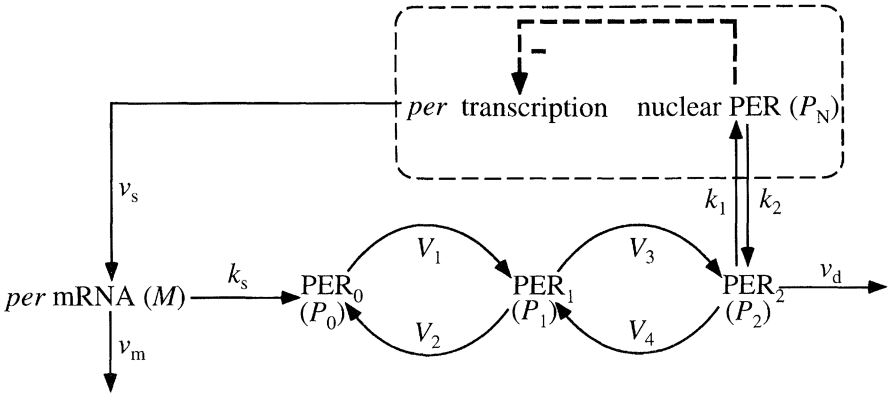


Figure 2. The structure of Goldbeter's dynamical model (after Goldbeter 1995). The concentration of *per* mRNA is represented by M, that of different forms of the PER protein by P_i. P_0 is the unphosphorylated, P_1 the monophosphorylated and P_2 the biphosphorylated form. P_N is for the nuclear PER protein, all the other concentrations are cytosolic. v_s is the

maximal rate of mRNA synthesis, v_m and K_m are the maximum rate and Michaelis constant for the enzymatic degradation reaction of the mRNA. The V_i and K_i give the maximum rates and Michaelis constants for the kinases and phosphatases catalyzing reversible phosphorylation reactions. v_d and K_d are the enzymatic parameters for the degradation reaction of fully phosphorylated PER. k_1 is a rate constant for the transport of PER protein into the cell nucleus, k_2 for the reverse transport. Feedback inhibition of *per* mRNA by nuclear PER is modeled by a Hill equation with a cooperativity of n and a repression threshold constant K_I.

Goldbeter wrote down the reaction rates for the different molecular species as follows:

$$\frac{dM}{dt} = v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M}{K_m + M} \tag{1a}$$

$$\frac{dP_0}{dt} = k_s M - V_1 \frac{P_0}{K_1 + P_0} + V_2 \frac{P_1}{K_2 + P_1} \tag{1b}$$

$$\frac{dP_1}{dt} = V_1 \frac{P_0}{K_1 + P_0} - V_2 \frac{P_1}{K_2 + P_1} - V_3 \frac{P_1}{K_3 + P_1} + V_4 \frac{P_2}{K_4 + P_2} \tag{1c}$$

$$\frac{dP_2}{dt} = V_3 \frac{P_1}{K_3 + P_1} - V_4 \frac{P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{K_d + P_2} \tag{1d}$$

$$\frac{dP_N}{dt} = k_1 P_2 - k_2 P_N \tag{1e}$$

The total (nonconserved) quantity of PER protein, P_t, is given by:

$$P_t = P_0 + P_1 + P_2 + P_N \tag{2}$$

Using numerical integration techniques, Goldbeter was able to show that for some parameter values there is indeed a limit cycle, in other words, a stable oscillation of the concentrations of mRNA and PER protein.

Bechtel and Abrahamsen stress that without this quantitative model, the sequential model provides no explanation for the *stability* of the circadian behavior. Without introducing quantitative parameters, the sequential model could produce all kinds of behavior, only some of which generate a limit cycle. Thus, the dynamical model must complement the sequential model to obtain the full explanation.

I will argue now that *at best* the sequential model *sensu* Bechtel and Abrahamsen can be represented as a causal model. The dynamical model cannot be so represented, even though it clearly represents a *causal process* (in an idealized and simplified way). Thus, I shall argue that the Goldbeter model is a case of a biological explanation that cannot be accounted for by causal graph models.

3.3 The Sequential Model as a Structural Causal Model

I shall first attempt to represent what Bechtel and Abrahamsen call the sequential model within this causal framework. There is an apparent difficulty in that the sequential model is *cyclical* whereas causal graphs are *acyclical*. However, this problem is not new and solutions have been proposed by several authors (Kistler 2013; Gebharter and Kaiser 2014; Clarke, Leuridan and Williamson 2014). Briefly, one way of doing this is by introducing a time index on some of the nodes of the causal graph structures. When a system comes to the end of a cycle, time has passed. This new state of the system should thus be represented

by a different node, a variable that represents the state of the system at a later time. This way, the cyclical path is broken up and "rolled out" in time and presents no problems for the causal modeler.

However, it should be clear that such a causal graph fails to fully explain the explanandum phenomenon, because essential dynamical information is missing. The graph would merely represent what Bechtel and Abrahamsen refer to as the sequential model. In the next section, I shall examine how the dynamical model could be represented.

3.4 The Dynamical Model as A Structural Causal Model

Could the same strategy that works for the sequential model also be used for representing Goldbeter's dynamical model by using causal graphs? It could be suggested that the causal structure of the model is captured by the following time-indexed causal graph:

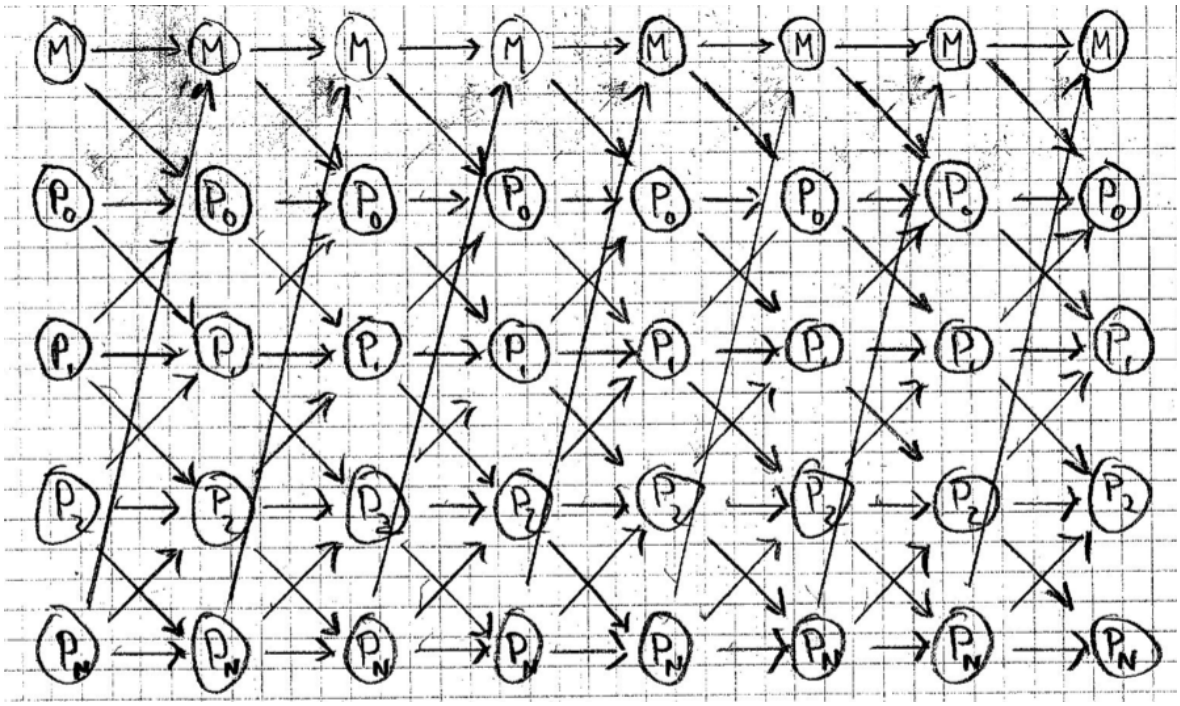


Figure 3. Proposed time-indexed DAG representing the causal dependencies in the Goldbeter model.

It could be argued, perhaps, that this DAG contains all the causal relations posited by the Goldbeter model. A quantitative structural model could also be constructed, for example by writing down rules for updating the values of the salient variables from each discrete time point to the next.

However, it should be clear that such a causal structural model would not be the same as the Goldbeter model. Differential equations with continuous time are

mathematically clearly different from a model with discrete time points.[3] Perhaps there is a discrete-time model that makes approximately the same prediction as Goldbeter's model. In fact, numerical simulations of the equation system use pretty much this strategy. However, the following difficulty arises: In order to really explain the explanandum phenomenon, a model must incorporate temporal information, namely information about how *rates of change* affect the behavior of the system. Goldbeter's differential equations contain precisely such information, and this information is crucial for the model's explanatory force. In fact, I wish to maintain that rates of change are *causally* relevant, because they are important determinants for the behavior of the whole causal process. Thus, I will show now that the Goldbeter model contains causally relevant variables that cannot be represented in the causal graph framework. Furthermore, to the extent that the model is substituted by a discrete-time model that is (approximately) predictively equivalent, the same difficulty arises.

My main argument is that Goldbeter equations do not have the right manipulability properties that are required by structural causal models. I will show, first, that not all causal variables can be subject to *ideal interventions* as required by the causal graph theory. Second, I want to show that the equations do not satisfy the *modularity* requirement that is widely thought to be important in causal models.

First, to see the problem with ideal interventions, consider for example equation (1a) of the Goldbeter model. Suppose we wished to intervene on *M*, the mRNA

[3] It is known that difference equations can have quite surprising properties, see May (1974).

concentration. This obviously cannot be done in a way that leaves the time derivative dM/dt unchanged (if I want to go faster on my bike, thus changing the value of v, I have to accelerate and thus change the value of dv/dt). The same problem occurs for all the other causally relevant variables in the model. Note also that a discrete model faces the exact same difficulty; the only difference is that the rates of change are defined over a time interval instead of a time point. Thus, these equations cannot be subject to the idealized interventions that define causal relations according to causal modelers.

Second, to see the failure of *modularity*, consider for example equations (1b) and (1c). Let us examine what happens when we replace (1b) by the following equation (1b*): dP_0/dt = p_0, where p_0 is some real number. This would not only wipe out the r.h.s. of (1b), it would also affect the equations that determine the value of P_0. The reason is, once again, that dP_0/dt and P_0 cannot be manipulated independently of each other. The same problem occurs for the other M- and P-variables. Therefore, the system of equations fails to satisfy the condition of modularity *sensu* Woodward (2003, 48-49, 327-39), which can also be viewed as a kind of manipulability.[4]

What features of the Goldbeter model are responsible for this lack of manipulability, including modularity? It seems to us that the main such feature is the fact

[4] The purpose of the modularity condition is normally to ensure that different equations represent different causal pathways or mechanisms. Perhaps it could be argued that, indeed, some causal mechanisms in the Goldbeter model overlap. For instance, there is a causal cycle between M and P_0 as well as between P_0 and P_1 and these causal cycles share P_0 as a common constituent (cf. Casini and Williamson unpublished).

that some causal variables that occur in the model *affect their own rate of change*, and that for these variables their rate of change is of crucial relevance – indeed *causal* relevance – for the behavior of the whole system. For example, the rate of change of mRNA (*M*) depends on its own concentration. This is due to a *causal* process that is mediated by RNA-degrading enzymes. Furthermore, the concentration of monophosphorylated protein P_1 depends causally on the concentration of unphosphorylated protein P_0, which in turn depends on P_1. Both causal dependencies are mediated by kinases, thus they are *causal* processes.[5]

In Goldbeter's representation of these processes, not only the *values* of the variables at a given time point but also their *rates of change* are *causally relevant*. In other words, it matters not only that a variable X change its value from x_1 to x_2, which is the kind of information that can be encoded in causal graphs. It matters also *how long it takes* for a variable to change by some amount, including an infinitesimally small amount. This rate of change is a *causally relevant property*, but this causal relevance cannot be represented as a causal dependence in the causal framework because the rate of change cannot be

[5] An anonymous referee suggested that these dependencies are not causal but constitutive or due to part-whole relations. While there might be some part-whole relations involved in the model (e.g., in the way in which different processes contribute to the overall rate of change of a variable), the dependencies we are talking about here, e.g., the dependence of the rate of change of *M* on the concentration *M* (equation 1a) are not of this kind. This dependence is due to an enzyme-directed biochemical reaction, which is clearly a causal processes.

manipulated independently of all the other variables and equation as the standard causal theory requires it (see above; lack of independent manipulability and modularity). Rather, in these causal processes, concentrations and their rates of change are so intimately intertwined and integrated (cf. Mitchell 2009) that it is not possible to disentangle causal difference-making and dynamical information.

Why can the differential equations in the Goldbeter model not be replaced by something more akin to the causal modeler's structural equations, e.g., difference equations with a discrete time variable? As I have argued, it seems the same difficulty would arise: as soon as the concentration variables and the time intervals are fixed, the rates of change are determined and therefore no longer independent.[6] Furthermore, replacing the differential equations by standard structural equations would be like trying to do Newtonian mechanics without using calculus; what would be the point?

A possible response by the causal modeler might be to deny that the differential equations are even contenders for representing causal dependencies. Differential equations contain functions of time and their derivatives and need to be *integrated* in order to predict or explain physical events. Surely, when we want to discuss the causal content of models such as Goldbeter's we have to consider suitably *integrated* forms of equations.

The problem with this reply is that systems of differential equations such as Goldbeter's or HH can only be integrated numerically. The solutions of these equations that are available, showing the concentrations of various molecular species, have been obtained with the help of computer simulations. In these solutions, whatever causal

[6] This was pointed out to me by an anonymous referee.

difference-making information was represented in the differential equations (if any) is irretrievably lost. However, these simulations provide a different kind of information: They show under which parameter values certain kinds of behavior are stable. In the case of the Goldbeter model, the behavior that is of particular importance, for obvious reasons, is stable oscillatory behavior. It can be represented by a limit cycle in a plane defined by mRNA and total PER protein concentrations. The limit cycle gives the initial conditions for M and P_t (= total PER protein concentration) that generate a stable oscillation, which functions as the basic *Zeitgeber* for Drosophila's biological clock. I would not refer to this kind of information as causal difference-making information but as *stability information*.

Perhaps it could be argued that the integrated model provides some kind of causal difference-making information as well. In his original 1995 paper, Goldbeter showed that the rate of PER protein degradation has a strong effect on the period of the oscillations. The more rapidly the protein is degraded in the cell, the longer the period of the oscillations become. The reason is intuitively clear: The more rapidly the protein disappears, the longer it takes for protein synthesis to rise the concentration above the threshold where the repression of transcription by nuclear PER protein significantly slows down gene expression such that the concentration of PER starts to drop after a period of increase. However, as intuitively obvious as this may be, the exact effect of the rate of decay on the period of the oscillations can only be predicted by such a dynamical model, which, as I have shown, contains causal information that is highly *integrated* with temporal, dynamical information and thus not representable by standard causal models, because the independent manipulability and modularity requirements are not satisfied.

In the following section, I will consider some results from the causal modeling literature as to how systems of differential equations can be represented by structural causal models. As I will show, these results, while it is highly illuminating for the problem at hand, actually support my thesis about the limitations of causal modeling.

4. Differential Equations and Causal Structural Models

Attempts to describe at least the equilibrium states of systems of differential equations with structural causal models can be found in the causation literature, for example, Mooij, Janzing and Schölkopf (2013); henceforth abbreviated as "MJS".[7] MJS treat systems of ordinary first-order differential equations such as they feature in many scientific models, e.g., the Lotka-Volterra model of predator-prey dynamics or the coupled harmonic oscillator in mechanics. The systems described by such equations may be considered to contain causal cycles. For instance, in a predator-prey system the density of predators affects the density of prey, which causally feeds back to the predator density. This is the kind of causal cycle that we also find in our biological clock case examined in the previous section. Even though causal graphs (DAGs) are typically acyclic, this is not a constraint that would somehow be necessitated by the formalism. I have already mentioned possible approaches to modeling causal cycles in Section. MJS take a somewhat different approach: They show that the equilibrium solutions of systems of coupled differential equations that describe systems with some causal feed-back correspond to a structural causal model.

[7] I wish to thank an anonymous referee for calling this work to my attention.

It is not possible here to reproduce the full treatment given by MJS. Basically, they consider dynamical systems represented by systems of coupled differential equations of the following form:

$$\dot{X}_i(t) = f_i\left(X_{pa_{\mathcal{D}}(i)}\right), X_i(0) = (\mathbf{X}_0)_i \;\; \forall i \in \mathcal{I}$$

where the indices $pa_{\mathcal{D}}(i)$ range over the set of parents of the variable X_i, each f_i is a smooth function of X, and each $(X_0)_i$ is an initial condition. Then, they provide an account of what it means to *intervene* on such a system, as intervention is part of the standard semantics of causal models. In a nutshell, an idealized intervention can be described as:

$$\dot{X}_i(t) = \begin{cases} 0, & i \in I \\ f_i\left(X_{pa_{\mathcal{D}}(i)}\right), & i \in \mathcal{I} \setminus I \end{cases}$$

$$X_i(0) = \begin{cases} \xi_i, & i \in I \\ X_{0_i}, & i \in \mathcal{I} \setminus I \end{cases}$$

In such an intervention, some set of components I of the system are forced to take some target value, such that the first time derivative of the variable X_i takes the value zero (i.e., X remains constant), while the variable takes some fixed target value ξ_i.[8] Thus, whatever

[8] Note how this intervention must fix the values for both the variables and their rate of change at the same time (cf. Section 3.4). This is exactly how the structural causal model obliterates information that is explanatorily relevant.

mechanism previously determined the value of the X_i, the intervention exogenously breaks this mechanism and sets the variables to a fixed value. This corresponds to the well-known breaking of directed edges by intervention variables in ordinary causal graphs.

What is interesting to note in the present context is that, according to the definition of an idealized intervention given by MJS, such an intervention changes not just one but two equations. This shows, once again, that such a system of equations is not modular in the sense discussed in the previous section. (It might be modular in the sense that it doesn't change any further equations, though).

The idealized interventions obviously change the equilibrium states of the system. For example, a Lotka-Volterra system has a steady state in which the predator and prey populations show an undamped oscillation. If intervened upon in the manner shown above, such a system changes its equilibrium state. If, for example, the intervention sets the predator density in a Lotka-Volterra system to ξ_2, the system's new unique stable equilibrium state is $(X^{eq}_1, X^{eq}_2)=(0, \xi_2)$. In general, equilibrium states of systems of intervened differential equations can always be obtained by setting the rates of change of the variables to zero by an intervention. The resulting equilibrium is then described by some equilibrium equations.

Just as in ordinary causal graph representations, nodes and directed edges can be used to represent the outcome of possible interventions on the variables figuring in systems of differential equations. In the cases such as the ones considered here, there will be a set of equilibrium equations for each possible intervention of the kind introduced above. MJS show how such equilibrium equations can be derived in general, and that they form causal

structural models in accordance with the causal framework assumed. Thus, it seems that the causal graph framework with its standard interventionist semantics is able to deal with systems of differential equations. MJS suggest that this approach "sheds more light on the concept of causality as expressed within the framework of Structural Causal Models, especially for cyclic models."

I wish to draw a different conclusion from MJS's highly illuminating treatment. In my view, their approach to dynamical systems described by differential equations reveals precisely the limitations to the causal graph framework that I wish to expose in this paper. For it is clear that such an approach can only deal with stable equilibrium states of a system, i.e., such equilibrium states where there is no more change. This is a simple consequence from the kind of interventions introduced, where the first derivatives with respect to time of the variables considered are set to zero. Thus, the structural causal models represent static situations rather than dynamic processes. For some intents and purposes, this may be fine. But if it is accepted that the dynamical models examined here are representations of the causal properties of a system and that the rates by which variables change is such a property, this kind of causal property does not seem to be captured by ordinary causal structural models.

I wish to end this argument by disenabling a potential misinterpretation. My thesis of this paper should not be understood as a claim about causal *discovery*. None of the considerations presented here support the conclusion that a causal search procedure of the kind developed by Spirtes, Glymour and Scheines (2000) would be unable to identify all

the variables that values of which affect the behavior of the system.[9] I am only claiming that the entities referred to by these variables have causal properties – in particular the rate of change – that cannot be given a causal interpretation by using the standard formalisms.

6. Conclusions

My intention in this paper has not been to argue that there exist forms of explanations in biology that are not causal. There clearly is a sense in which DNA sequence recognition by proteins (Kaiser, this symposium) as well as the biological clock mechanisms discussed here are causal processes. What I as well as Kaiser (this symposium) want to show is that these biological explanations contain *causal* information that is not reducible to causal difference-making information of the kind that can be expressed in the formal causal models available today. Biological explanations often contain causal information that is inextricably intertwined with, first, *spatial* information and, second, *dynamical* information. The spatio-temporal aspects represented in these explanations are not such that they could simply be integrated with the causal difference-making information to give the full picture. At least in the case of the dynamical information contained in systems of differential equations, there appears to be a deep incompatibility between the axioms of causation and the dynamical model. Just as the circadian clock mechanism cannot be

[9] For an illuminating discussion of this important issue in the context of systems of differential equations see Dash (2005). It should be noted that, just like in the Mooij, Janzing and Schölkopf (2013), time derivatives of variables are never treated as independent causes.

understood by looking at the level of individual molecules, complex spatially organized cohesive interactions in DNA-protein complexes (see Kaiser, this symposium) cannot in practice be expressed by causal graphs in a way that brings out the explanatory power and utility of these models.

My conclusion with respect to dynamical mechanistic models differs thus somewhat from Bechtel's and Abrahamsen's illuminating analysis: What they call the sequential and dynamical mechanisms, respectively, represent not two models that complement each other. Rather, in my view they represent incompatible perspectives on the same phenomenon of the kind that scientific pluralists have postulated (Kellert, Longino and Waters 2006).

Thus, rather than just the need of supplementing causal graphs with spatio-temporal labels such as to fine-tune them, a close examination of biological explanations rather reveals some intrinsic limitations of a certain type of causal model. Perhaps a new theory of causation is needed in order to do (more) justice to such explanations, in biology as well as in other sciences that deal with complex dynamical processes.

References

Bechtel, William (2005), *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge: Cambridge University Press.

Bechtel, William (2013), "From Molecules to Networks: Adoption of Systems Approaches in Circadian Rhythm Research", in Hanne Andersen, Dennis Dieks, Wenceslao J.

Gonzalez, Thomas Uebel and Gregory Wheeler (eds.), *New Challenges to Philosophy of Science*, Berlin: Springer.

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* 41:321-333.

Braillard, Pierre-Alain (2010), "Systems Biology and the Mechanistic Framework", *History and Philosophy of the Life Sciences* 32:43-62.

Casini, Lorenzo, Phyllis McKay Illary, Federica Russo, and Jon Williamson (2011), "Models for Prediction, Explanation and Control: Recursive Bayesian Networks", *Theoria. An International Journal for Theory, History and Foundations of Science* 70 (1):5-33.

Clarke, Brendan, Bert Leuridan and Jon Williamson (2014), "Modelling Mechanisms With Causal Cycles", *Synthese* 191 (8):1651-1681.

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

Dupré, John (2013), "Mechanism and Causation in Biology. I—Living Causes", *Aristotelian Society Supplementary Volume* 87:19–37.

Gebharter, Alexander, and Marie I. Kaiser (2014), "Causal Graphs and Biological Mechanisms", in Marie I. Kaiser, O. Scholz, D. Plenge and A. Hüttemann (eds.), *Explanation in the Special Sciences. The Case of Biology and History*, Berlin: Springer.

Gebharter, Alexander (2014), "A Formal Framework for Representing Mechanisms?", *Philosophy of Science* 81 (1):138-153.

Goldbeter, Albert (1995), "A Model for Circadian Oscillations In the Drosophila Period Protein (PER)", *Proceedings of the Royal Society of London. B: Biological Sciences* 261 (1362):319-324.

Hardin, P. E., Hall, J. C., and M. Rosbash (1990), "Feedback of the *Drosophila Period* Gene Product On Circadian Cycling of Its Messenger RNA Levels", *Nature* 343 (6258): 536-540.

Hitchcock, Christopher, and James Woodward (2003), "Explanatory Generalizations, Part II: Plumbing Explanatory Depth", *Noûs* 37 (2):181-199.

Kellert, Stephen H., Helen E. Longino, and C. Kenneth Waters, eds. (2006), *Scientific Pluralism*, *Minnesota Studies in Philosophy of Science, Vol. XIX*. Minneapolis: University of Minnesota Press.

Kistler, Max (2013), "The Interventionist Account of Causation and Non-Causal Association Laws", *Erkenntnis* 78:1-20.

Kuhlmann, Meinard (2011). Mechanisms in Dynamically Complex Systems. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the Sciences*. Oxford: Oxford University Press.

May, R. M. (1974), "Biological Populations with Nonoverlapping Generations: Stable Points, Stable Cycles and Chaos", *Science* 186:645-647.

McKay Illari, P., Russo, F., & Williamson, J. (Eds.) (2011). *Causality in the Sciences*. Oxford: Oxford University Press.

Mooij, Joris M., Dominik Janzing, and Bernhard Schölkopf (2013), "From Ordinary Differential Equations to Structural Causal Models: The Deterministic Case", in Ann Nicholson and Padhraic Smyth (eds.), *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, Corvallis: AUAI Press, 440-448.

Otsuka, J. (forthcoming), "Causal Foundations of Evolutionary Genetics", *The British Journal for the Philosophy of Science*

Pearl, Judea (2000), *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Russo, Federica (2012), "Public Health Policy, Evidence and Causation: Lessons From the Studies On Obesity", *Medicine, Health Care and Philosophy* 15 (2):141-151.

Spirtes, Peter, Clark Glymour and Richard Scheines (2000), *Causation, Prediction, and Search*. Cambridge, Mass.: MIT Press.

Waskan, Jonathan (2011), "Mechanistic Explanation at the Limit", *Synthese* 183 (3):389-408.

Waters, C. Kenneth (2007), "Causes That Make a Difference", *The Journal of Philosophy* CIV (11):551-579.

Weber, Marcel (2005), *Philosophy of Experimental Biology*. *Cambridge Studies in Biology and Philosophy*. Cambridge: Cambridge University Press.

Weber, Marcel (2008), "Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation", *Philosophy of Science* 75:995-1007.

Weber, Marcel (2012), "Experiment in Biology", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/biology-experiment/

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

Woodward, J. (2008). Cause and Explanation in Psychiatry: An Interventionist Perspective. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*. Baltimore: Johns Hopkins University Press.

Woodward, J. (2010). Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation. *Biology and Philosophy, 25,* 287-318.

Woodward, James (2011), "Mechanisms Revisited", *Synthese* 183 (3):409-427.

Woodward, James (2013), "Mechanistic Explanation: Its Scope and Limits", *Proceedings of the Aristotelian Society Supplementary Volume* 87 (1):39-65.

Woodward, James, and Christopher Hitchcock (2003), "Explanatory Generalizations, Part I: A Counterfactual Account", *Noûs* 37 (1):1-24.