Chapter 8
The Relevance of Irrelevance: Explanation in Systems Biology

Fridolin Gross

Abstract In this chapter I investigate explanations in systems biology that rely on dynamical models of biological systems. I argue that accounts of mechanistic explanation cannot easily make sense of certain features of dynamical patterns if they restrict themselves to change-relating relationships. When investigating the use of such models, one has to distinguish between the concepts of causal or constitutive relevance on the one hand, and explanatorily relevant information on the other. I show that an important explanatory function of mathematical models consists in elucidating relationships of non-dependence. Notably, the fundamental concept of robustness can often be accounted for in this way, and not by invoking separate mechanistic features. Drawing on examples from the scientific literature, I suggest that an important aspect of explaining the behavior of a biological mechanism consists in elucidating how in the systemic context components are not, or only weakly, dependent on each other.

Keywords Systems biology • Mechanism • Dynamical system • Explanatory relevance • Robustness • Invariance

1 Introduction

The molecules inside a living cell do not behave like the molecules in a gas. In a gas the single particles freely move around and interact randomly (if at all), while showing no apparent organization. A gas seems to be the perfect example of a simple aggregate (Wimsatt, 1997, 2007) whose macro-level properties are invariant under various kinds of changes at the micro-level. Obviously, the cell is not such a simple aggregate. However, a cell does not appear to behave like a mechanical clock either. The mechanism of a clock almost certainly breaks down if we remove one part or try to exchange two different components, whereas living

F. Gross (✉)
Institute for Theoretical Biology, Charité Universitätsmedizin, Philippstraße 13, 10115 Berlin, Germany
e-mail: fridolin.gross@hu-berlin.de

P.-A. Braillard, C. Malaterre (eds.), *Explanation in Biology*, History, Philosophy and Theory of the Life Sciences 11, DOI 10.1007/978-94-017-9822-8_8

systems are often surprisingly stable under a wide range of perturbations.[1] From this one may infer that on the spectrum of organizational complexity cells assume a position somewhere between gases and clocks. But arguably many people would strongly object to this classification. Aren't biological systems much *more* complex and organized than the artifacts of mechanical engineering? And aren't there very specific kinds of perturbations, such as mutations, to which living systems react in very sensitive ways? Isn't it more plausible to assume that living systems are very complex mechanisms, but differently from clocks they have additional features that account for their particular ways of resisting perturbations? The coexistence of extreme complexity and robustness is undoubtedly one of the most fascinating features of life. As the physiologist Walter Cannon remarked:

> When we consider the extreme instability of our bodily structure, its readiness for disturbance by the slightest external forces and the rapid onset of its decomposition as soon as favoring circumstances are withdrawn, its persistence through many decades seems almost miraculous. (Cannon, 1932, 20)

Cannon and many others have shown that the stability of physiological processes can often be explained with reference to particular mechanisms. For example, the homeostasis of blood sugar levels can be explained with reference to a simple feedback mechanism involving the hormones insulin and glucagon.

Systems biology has recently started to investigate similar phenomena at a more fine-grained level, such as the robustness of genetic or metabolic networks. Is it obvious that the robustness of such fundamental structures of living systems can be understood in the same way? In what follows I want to show how some of the insights gained by systems biologists challenge widely held intuitions about mechanistic explanations in the life sciences. By doing so, I do not want to suggest that systems biology will eventually give rise to an alternative, non-mechanistic paradigm of explanation in biology. I merely want to draw attention to certain issues that mechanistic accounts will need to address in order to capture the explanatory ambitions of systems biologists. The relationship of systems biology to the framework of mechanistic explanation and the particular importance of the concept of robustness have already been discussed elsewhere (e.g. Braillard, 2010; Breidenmoser and Wolkenhauer, 2015; Brigandt, 2015). Here, I want to pursue some of the already developed ideas further and address some of the issues from a different angle by focusing on the idea of "non-dependence."

The currently most refined accounts of mechanistic explanation link the understanding of systemic behavior to manipulationist criteria of causal or constitutive relevance (e.g. Craver, 2007; Woodward, 2010). The claim I want to defend in this chapter is that these accounts rely on a clock-like picture of biological mechanisms by assigning importance solely to *change-relating* relationships. I argue instead for

[1] An impressive example of such robustness is revealed, for instance, in the experiments of rewiring the *E. coli* gene regulatory network by Isalan et al. (2008).

the explanatory value of relationships that are not change-relating,[2] especially when it comes to explanations of the behaviors of very complex mechanisms. Examples of explaining biological robustness by means of dynamical modeling will provide the right kind of illustrations for this point.

The chapter is organized as follows. In Sect. 2, I introduce the manipulationist account of mechanistic explanation and discuss how biological robustness might be accounted for from this perspective. Section 3 takes a closer look at Woodward's concepts of *invariance* and *stability*, which seem to be closely related to robustness, and investigates their role in manipulationist explanations. For illustrative purposes, a toy example of gene expression is presented in Sect. 4. Here my aim is to show that the significance of non change-relating relationships arises especially in the context of dynamic modeling. Dynamic (or steady-state) equilibrium is arguably the simplest case of dynamic stability, and it reveals some typical features of explanations that are given in systems biology. Then, in Sect. 5, I will present a real case study from systems biology that illuminates how information about non-dependence plays an essential role in our understanding of biological mechanisms. After this more formal discussion, these issues are connected in Sect. 6 to a general perspective on robustness and the global architecture of living systems.

2 Mechanistic Explanation

It seems natural to think that real understanding of a system implies the ability to predict how it will respond to various kinds of interventions. To understand a phenomenon means to know how changes in it can be brought about, and this idea seems, at least implicitly, to underlie many of the recent conceptions of mechanistic explanation in the philosophy of science. The relationship between intervention and explanation has been made most explicit by James Woodward (2002, 2003) in his manipulationist account of causation and explanation which was subsequently adopted and further developed by some of the main proponents of mechanistic explanation (e.g. Craver, 2007; Glennan, 2002). On Woodward's account a causal relationship holds between two variables or events, roughly speaking, if it is possible (at least in principle) to systematically bring about changes in one by intervening on the other. The importance of these relationships for explanation, according to Woodward, lies in the fact that they allow us to answer a range of counterfactual *what-if-things-had-been-different* questions about the explanandum. We may, for instance, explain why a particular person has contracted lung cancer by referring to the fact that the person was a heavy smoker. The causal knowledge that the occurrence of cancer can be influenced by intervening on smoking behavior

[2] I will synonymously speak of "non change-relating relationships" and "relationships of non-dependence".

increases our understanding since it allows us to infer the counterfactual claim that the person (probably) wouldn't have gotten the cancer if she hadn't smoked.

However, mechanistic explanations are not necessarily explanations of effects in terms of their causes, but usually are understood as explanations of the activity of a whole in terms of the properties of its parts. This distinction can be further illuminated by considering the different types of questions that different explanations are supposed to answer. While explanations of effects in terms of their causes are directed towards *why-questions*, such as 'why did this person get lung cancer?,' the description of the mechanism underlying a behavior may be understood most intuitively as answering a *how-question*, such as 'how does the heart pump blood?'. In this context, Wesley Salmon referred to explanations in terms of underlying structure as *constitutive* and distinguished them from *etiological* explanations that cite the causal history of an event or phenomenon (Salmon, 1984, 275). More recently, Craver (2007) has argued that constitutive dependencies between mechanisms and their components are metaphysically distinct from causal dependencies holding between objects at the same level. He argues, however, that usually both causal and constitutive relationships are employed in mechanistic descriptions, and that both can be understood within Woodward's general manipulationist perspective.

Mechanistic accounts are often presented as an alternative to nomological models of explanation. According to the latter, explanation consists in logically deriving the explanandum from premises that include law-like generalizations. Mechanistic accounts, by contrast, hold that the explanation does not lie in a derivation, but is directly given by objective relationships, causal or constitutive, that account for the explanandum phenomenon. The core idea of applying the manipulationist framework to mechanistic explanations can be illustrated with the following quote by Carl Craver:

> One need not be able to derive the phenomenon from a description of the mechanism. Rather, one needs to know how the phenomenon is situated within the causal structure of the world. That is, one needs to know how the phenomenon changes under a variety of interventions into the parts and how the parts change when one intervenes to change the phenomenon. When one possesses explanations of this sort, one is in a position to make predictions about how the system will behave under a variety of conditions. Furthermore when one possesses explanations of this sort, one knows how to intervene into the mechanism in order to produce regular changes in the phenomenon. (Craver, 2007, 160)

Craver thus thinks of *explanatory relevance* (which components and relationships should figure in a mechanistic explanation) in terms of *manipulationist relevance* (which factors can be manipulated to change the phenomenon). The proponents of the manipulationist view argue that, restricting oneself to change-relating relationships one obtains a criterion for explanatory relevance that avoids many of the problems afflicting both nomological accounts and traditional versions of causal-mechanical explanation (e.g. Railton, 1981; Salmon, 1984).

If one conceives of biological systems as clock-like, it is plausible to equate manipulationist relevance with explanatory relevance. In a clock it seems that exactly those interventions that bring about changes in the overall behavior are

the ones that reveal the relationships one needs to know in order to grasp the underlying mechanism. For instance, if the balance spring in a clock is replaced by an otherwise similar spring with greater stiffness, the balance wheel will oscillate with increased frequency, and, as a consequence, the hands of the clock will move faster. Generalizing from this example, one may say that in a certain sense machines like clocks are extremely fragile because changes in the components are rigidly connected to changes in systemic behavior. The observation that clocks do not easily fall apart, in spite of this fragility, is explained by the fact that the properties of the parts are not easily changed in the first place. Consider the effect of temperature on a clock. A clock made of metal can work reliably both in most climates because the temperatures that could significantly deform its components lie outside the typically encountered range. Similarly, most mechanical devices owe their robustness to the fact that properties of their parts are insensitive to a wide range of external perturbations or changes in external conditions. 'Mechanical robustness' therefore can be taken to mean that the components of a system are insensitive to interventions on external conditions.

It seems that biological systems are not fragile in the same sense. Biological robustness means roughly "that some property of the system remains the same under perturbation" (Gunawardena, 2010, 35). However, in this context 'perturbation' is understood not as a change in external conditions, but as a change in the components or the organization of the system itself. Thus biological robustness implies that certain interventions on the components do *not* bring about changes in a phenomenon, which in turn means that there are relationships between properties of the system and its components that are *not* change-relating. What role do such relationships play in our attempts to understand biological systems? And how could they be interpreted within a manipulationist account of mechanistic explanation?

There seem to be two strategies of dealing with such relationships from within the manipulationist framework. On the one hand, one may argue that they simply fail to meet the criterion for explanatory relevance. For example, a clock's behavior will not be altered by changing the color of the balance spring. Consequently, the color of the spring is considered irrelevant when it comes to explaining how the clock works. On the other hand, there might be occasions where the observation that something *doesn't change* itself is of explanatory interest. The manipulationist will then set out to look for an explanation of this behavior in terms of underlying relationships that actually *are* change-relating, such as in the case of blood sugar homeostasis (see Issad and Malaterre, 2015). Insofar as robustness is a somehow "surprising" or "almost miraculous" (Cannon, 1932) property of living systems, she will attempt to explain it by looking for specific mechanisms that are responsible for the resistance to change. In any case, change-relating relationships are the fundamental building blocks of mechanistic explanations for the manipulationist. Relationships that are not change-relating either are irrelevant or themselves have to be explained in terms of change-relating relationships.

By investigating examples of dynamical modeling in systems biology, I will show that relationships that are not change-relating (relationships of non-dependence) point to something deeper and draw our attention to complementary aspects of sci-

entific understanding that have been neglected in recent discussions on mechanistic explanation. Before turning to these examples, however, I will have to say a bit more about the connection between change-relating relationships and explanation according to the manipulationist picture and the problems it raises for explaining biological robustness.

3 Explanation and Invariance

To illustrate his counterfactual account of causal explanation, James Woodward (2003, 187) makes use of a simple example from physics that probably can be found in any textbook on electrostatics (Fig. 8.1). A very long straight wire carries a uniformly distributed electric charge with density λ. The explanandum in this example is the force of the electric field on a test charge at position P at a perpendicular distance r from the wire. Woodward describes a derivation that is based on Coulomb's law and determines the strength of the field at P by summing up the contributions dq from all the infinitesimal sections dx of the wire. As a result he obtains the following expression:

$$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r}\,. \tag{8.1}$$

Woodward argues that this relationship, together with its derivation, explains the field intensity at P because it allows us to predict how the value of E changes if we intervene on the system in various ways. For instance, if we increase the distance between the wire and the test charge, the formula tells us that the field intensity decreases proportionally to the reciprocal of the distance. Changing the relative charge λ, on the other hand, results in a proportional change in intensity.

Generalizing from this example, Woodward proposes that explanation amounts to exhibiting the systematic patterns of counterfactual dependence that can be expressed as functional relationships between variables. In doing so, however, he restricts himself to relationships that are change-relating, that is, to relationships in which an intervention on one variable brings about a change in the other. If we consider the derivation of (8.1), however, we notice that it also elucidates relationships that are not change-relating. For instance, we learn that moving the test

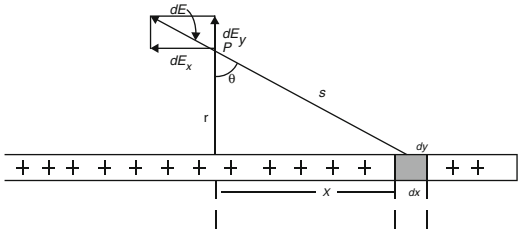

Fig. 8.1 Woodward's example of the charged wire. Reprinted by kind permission of Oxford University Press, Oxford, from Woodward (2003, 188)

charge to a new position P' at the same distance from the wire will not change the value of the field intensity because, as can easily be shown, such a transformation would not affect the result of the calculation. Likewise, we can infer that the x-component E_x of the field vector will never change to a value different from zero, no matter how we intervene, provided that we do not destroy the symmetry of the geometrical setup.

In Woodward's picture information about such non change-relating relations plays no direct role in explanation. However, the notion of *invariance*, that figures prominently in his account, seems closely related. He argues that causal claims are always associated with claims about invariant relationships:

> Invariance under at least one testing intervention (on variables figuring in the generalization) is necessary and sufficient for a generalization to represent a causal relationship or to figure in explanations. (Woodward, 2003, 250)

The idea is roughly the following. A generalization on which a causal claim is based can be described as a functional relationship between two variables of the type $Y = G(X)$. It is not necessary that such a generalization holds under all circumstances, instead it is required only that there are *some* possible changes of X under which it continues to hold. Invariance obviously comes in degrees, but as long as there is a minimum of invariance, a relationship is causal and, therefore, potentially explanatory. Highly invariant generalizations, such as the fundamental laws of physics, do not necessarily give rise to better explanations, even though they might have other desirable features.

In a more recent article Woodward uses the terms of *invariance* and *stability* interchangeably, and gives a slightly different characterization in terms of background circumstances. He argues that in order to qualify as causal, it is sufficient that a relationship of counterfactual dependence holds in some set of circumstances B_i. He then states:

> The *stability* of this relationship of counterfactual dependence has to do with whether it would continue to hold in a range of other background circumstances B_k different from the circumstances B_i. (Woodward, 2010, 291–292, emphasis in original)

According to this characterization, a claim of invariance or stability can be formally expressed as:

$$Y = G(X, B_i) = G(X, B_k) \text{for all } k \text{ in some set } K\,, \tag{8.2}$$

which obviously implies the existence of a relationship $F(B) = G(X, B)$ of non-dependence, that is, $F(B_k) = F(B_l)$ even if $B_k \neq B_l$. Thus Woodward's account actually relies on *both* change-relating and non change-relating relationships. However, the two kinds seem to play very different roles. Change-relating relationships on Woodward's view are the crucial elements; they provide the *content* of the explanation, so to speak, and elucidate the features of the explanandum phenomenon by providing patterns of counterfactual dependence. Relationships of invariance, on the other hand, largely keep in the background; they are necessary for specifying the range of application, or generality, of an explanatory claim, but strictly speaking do not provide any explanatory information.

There is thus a clear conceptual separation between the two types of functional relationships reflecting Woodward's distinction between causal explanatory claims, on the one hand, and claims about invariance, on the other hand. However, if we think again of the derivation of the field strength in the wire example, we recall that it also provides information about interesting relationships of non-dependence. If we keep track of both the x- and y-components of the field strength, we notice that all the infinitesimal contributions to E_x exactly cancel each other out, independently of the position P at which the field is evaluated. The relationship

$$E_x(P) = 0 \text{ for all } P, \tag{8.3}$$

however, does not play the role of an invariance condition, and it does not seem to be irrelevant in the same way as, for instance, the color of the wire. At least for mathematical structures it seems that both change-relating and non change-relating relationships are potentially important for our understanding. It is not obvious why one type of relationship should be more interesting or more informative than the other. The reason why non-change-relating relationships are often neglected might be due to the following feature: they can be represented in a very compressed way, or simply be left implicit. As Herbert Simon put it: "Mother Hubbard did not have to check off the list of possible contents to say that her cupboard was bare" (1962, 478). But the fact that non-change-relating relationships lend themselves to descriptive economy does not imply that they are explanatorily irrelevant.

Woodward's account suggests that change-relating relationships exhaust all that is needed for explaining the behavior of a system. But if relationships of non-dependence can contribute to our understanding of mathematical models, why shouldn't they be taken as contributing to our understanding of phenomena that are explained by means of such models?[3] As will be further illustrated later in this chapter, the functional relationships that play a role in the models of systems biology often are change-relating in some particular range of values while being non-change relating in a different range. I will show that usually both types of information are crucial for an understanding of complex behavior, without one necessarily being reducible to the other.

By that I do not want to deny the important role that change-relating relationships play in determining the causal or constitutive links within a mechanism. There is no doubt that these relationships provide explanations by allowing us to answer to why-questions of a particular type. But this alone does not entail the equivalence of information about manipulationist relevance and explanatory information when it comes to more complex mechanistic explanations.

A related issue, that Woodward's account leaves unclear, is how invariance or stability itself is explained. As Robert Batterman notices:

[3]For a more sophisticated analysis of the role of mathematics in scientific explanation see Baker (2015).

> Woodward stresses the importance for explanation of a kind of invariance and robustness that may be present in a given regularity to some degree or other. Thus, he discusses how "nonlaw-like" regularities may, because of their robustness, play crucial explanatory roles. Woodward is not concerned to answer why-questions about the universality or degree of universality of the regularities that he discusses. That is, he does not, as far as I can tell, ask the question why the regularity has the robustness that it has or has it to the degree that it has. (Batterman, 2002, 59)

Batterman argues that in the explanation of a phenomenon one has to distinguish between two different kinds of why-questions:

> A type (i) why-question asks for an explanation of why a given instance of a pattern obtained. A type (ii) why-question asks why, in general, patterns of a given type can be expected to obtain. Thus, a request to explain an instance of universality is a request to provide an answer to a type (ii) why-question. (Batterman, 2002, 23)

Batterman's ambition to explain universality and Woodward's efforts to elucidate explanation in terms of contingent causal generalizations point to different but possibly complementary aspects of scientific curiosity. These may be seen as loosely related to the different types of questions that are typically asked in the physical and the biological sciences, respectively. It is a philosophically interesting question how the new field of systems biology locates itself on this spectrum since, with regard to its methodological and explanatory resources, it has often been perceived as pushing biology more towards a physics attitude (see e.g. Poon, 2011). A closer look at some examples may help to shed light on this issue.

4 Explaining Equilibria

Let us start with a very simple case and consider the following minimal model of gene expression. The system consists of a protein with concentration X that is synthesized at a constant rate $S = \sigma$, while its degradation rate, $D = \delta \cdot X$, is proportional to the concentration. Figure 8.2 graphically represents the qualitative features of this model. The dynamics of X is captured by the following differential equation:

$$\frac{dX}{dt} = S - D(X) = \sigma - \delta \cdot X \,. \tag{8.4}$$

Solving this equation allows us to obtain the temporal behavior of X depending on a given initial concentration X_0 at time $t = 0$. As can be checked, the explicit solution is given by:

$$X(t, X_0) = \left(X_0 - \frac{\sigma}{\delta}\right) \exp(-\delta t) + \frac{\sigma}{\delta} \,. \tag{8.5}$$

After sufficient time, the value of the exponential will become very small and the first part of the right hand side of (8.5) can be neglected. Formally,

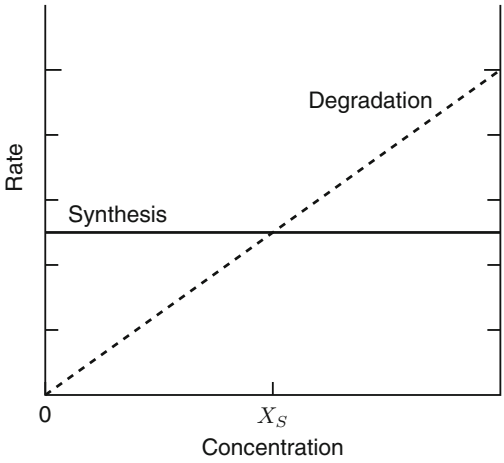

Fig. 8.2 Rate balance plot for the simple gene expression model. Synthesis is constant, while degradation depends linearly on protein concentration. Both can be represented as *straight lines*. The intersection of the two lines corresponds to the equilibrium state. The stability of the equilibrium can be inferred from the sign of the resulting rate when subtracting degradation from synthesis

$$X(t, X_0) \to \frac{\sigma}{\delta} \quad \text{for} \quad t \to \infty\,. \tag{8.6}$$

Note that the expression to which X converges does not contain X_0. This means that the protein concentration in the long run does not depend on its initial value, but assumes an equilibrium (or steady state) value $X_S = \sigma/\delta$ that depends only on the rates of synthesis and degradation. A further consequence is that, whenever the system is perturbed by changing the concentration to some value $X \neq X_S$, it will always return to X_S eventually. At first sight this derivation seems to provide a perfectly satisfactory explanation of equilibrium.

The model just described is very similar to an example that Elliott Sober (1983) has used to raise some questions about causal-mechanical approaches to explanation. He refers to an explanation given by R. A. Fisher for the 1:1 sex ratio observed in many sexually reproducing species.[4] Instead of providing a particular causal history for the occurrence of the ratio, Fisher points out why the long run ratio in many sexually reproducing populations does *not* depend on particular causal details. As Sober reports:

> Fisher's account shows why the actual initial conditions and the actual selective forces don't matter; whatever the actual initial sex ratio had been, the selection pressures that would have resulted would have moved the population to its equilibrium state. Where causal explanation shows how the event to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired. (Sober, 1983, 202)

Sober concludes that equilibrium explanations are not causal explanations of the etiological type:

[4]Baker (2015) briefly mentions the same example and identifies equilibrium explanation as one of three general classes of mathematical explanation in science.

> The causal explanation focuses exclusively on the actual trajectory of the population; the equilibrium explanation situates that actual trajectory (whatever it may have been) in a more encompassing structure. It is in this way that equilibrium explanations can be more explanatory than causal explanations even though they provide less information about what the actual cause was. This difference arises from the fact that explanations provide understanding, and understanding can be enhanced without providing more details about what the cause was. Equilibrium explanations are made possible by theories that describe the dynamics of systems in certain ways. (Sober, 1983, 207)

Sober thus hints at a discrepancy between information about particular causal events and information that is relevant for explanation. He suggests that explaining equilibrium means showing why particular causal facts do not make a difference to the outcome. The question thus arises how this idea relates to Woodward's account according to which such facts are simply explanatorily irrelevant. Is it straightforward to capture equilibrium explanations within the manipulationist framework?

Before trying to determine what kinds of explanations they are—or aren't, we should clarify what it is that equilibrium explanations are supposed to explain. Regarding the sex ratio, the general question is 'Why is there an equilibrium at a sex ratio of 1:1 in so many species?.' However, this question can be interpreted as actually including three different calls for explanation, depending on where we put the stress in the sentence.[5] First, it may be read as the question of why it is one and the same ratio that is observed across a wide range of sexually reproducing species. In other words, why does the rate not assume different values for different species? Second, as the question of why that ratio has the particular numerical value of $r = \#\text{males}/\#\text{females} \approx 1$, and not some other number in the interval $(0, \infty)$. And third, as the question of why the observed ratio represents an equilibrium point, that is, why it is stable and adjusts itself after perturbations.

Each way of interpreting the initial question calls for an account that makes use of different explanatory resources. The first interpretation, even though interesting in its own right, is not relevant for the current discussion since it seems to mainly depend on empirical facts that are specific to evolutionary biology. For this reason, my focus will be on the differences between the second and the third interpretation that more directly pertain to the phenomenon of equilibrium in general, and roughly correspond to Batterman's type (i) and type (ii) why-questions. I will discuss these differences in more detail using the particularly clear example of the gene expression model.

Let us therefore look at the explanation-seeking question and the two relevant interpretations when transferred to this example. The question is, 'Why is there an equilibrium at a concentration $X = X_S$?', and it can be interpreted as expressing an interest either in the particular numerical value or in the fact that there is an equilibrium. Responding to the first, the derivation of (8.6) can be taken to show why the protein concentration at steady state is given by the particular ratio σ/δ.

[5]See Morange (2015) for a general discussion of how the ambiguity of questions can lead to explanatory diversity.

This seems to represent a paradigmatic case of a Woodwardian explanation since the steady state concentration is explained in terms of the dependency relations characterizing the system. It clearly allows us to answer a range of counterfactual *what-if-things-had-been-different* questions. For instance, we can predict how the steady state value would change if we were to intervene in ways that change the rates of synthesis or degradation. Differently from the type of causal explanation that are the target of Sober's argument, however, this explanation refers to structural features of the model rather than to causal history. In the terminology introduced earlier, this explanation might, therefore, best be understood as constitutive. This is the way in which Kuorikoski (2007) interprets equilibrium explanations within a manipulationist framework:

> If explanations indeed track dependencies instead of persistence, the interesting explanatory relationship cannot be the one between the initial conditions and the equilibrium state, as might first be surmised, and indeed as seems to have been Sober's view. Instead, what the equilibrium state does depend on are the structural features of the system. *Equilibrium explanations are not causal explanations of events but structural or constitutive explanations of system-level properties*. (Kuorikoski, 2007, 154, emphasis in original)

However, stating the dependency relations between parameters and steady state value alone does not give an answer to Batterman's type (ii) question of why the pattern, in this case equilibrium, obtains in the first place. Instead, as we have seen, equilibrium seems to be explained precisely by deriving a relation of *non-dependence* between the initial conditions and the long-run concentration. Is there another way in which we can understand this aspect of equilibrium within a manipulationist framework of causation while avoiding Sober's puzzle about the irrelevance of particular causal facts?

To maintain a contrastive focus, one may try to interpret the existence of a single stable equilibrium as a property that systems either do or do not possess, and determine exactly what this property depends on. It turns out that in the present example this property depends only on the structure of the model.[6] This dependency may be expressed in terms of a binary variable $P \in \{0, 1\}$ in the following way:

$$P\left(\{S, D\}, \{X\}, \{\sigma, \delta\}\right) = 1, \tag{8.7}$$

where $S = \sigma$ and $D = \delta \cdot X$ represent the particular types of functions used to express the dynamic relationships, while $\{X\}$ and $\{\sigma, \delta\}$ stand for the sets of variables and parameters that appear in the model. By modifying this structure in particular ways, one may obtain a different model for which $P = 0$, that is, a model without an equilibrium state, or perhaps with more than one. An example of such a modification is the complete disruption of degradation, i.e. setting $D = 0$, or the addition of a more complex dependency $S(X)$ of synthesis on the concentration.

[6]Since the equilibrium is global it does, for instance, not depend on the initial concentration being within a particular range. However, similar arguments can be made for cases of non-global equilibrium.

This reasoning suggests that in principle it might be possible to find a representation of the (potentially very complicated) dependency relation between P and the structural properties of the model. Subsequently, one could make use of this relation to explain why a particular instance of the model does or does not possess the equilibrium property P. Furthermore, one may argue that P's structural dependency explains equilibrium by showing how it appears when the structural parameters are changed in particular ways. But have we thereby really explained equilibrium? It seems that by using the complex dependency relation, we have at best been able to give a more sophisticated answer to a type (i) why-question. We have explained that a particular system shows equilibrium because it belongs to a particular structural class. If we intervene on the structure of the system in such a way that it no longer belongs to this class, it will exhibit qualitatively different behavior.[7] We have not explained the behavior itself. It seems that by using a manipulationist strategy, we do not reach beyond the explanation of instances of equilibrium.

To summarize, a satisfactory explanation of equilibrium in causal-etiological terms fails for the reasons discussed in Sober's paper. In order to explain equilibrium constitutively, the manipulationist may invoke relationships that relate quantitative or qualitative changes in behavior to changes in structural features of the system, but thereby fails to give an account of how the behavior is produced in the first place. As I argue, and as Sober suggests, equilibrium is best explained by referring to a relationship of non-dependence. The case of equilibrium thus shows how knowledge about such relationships can be explanatorily relevant.

5 Dissecting a Dynamical Switch

After these initial considerations about relationships of non-dependence, one may ask whether they are of any importance for the description of actual scientific explanations. For this reason I will now turn to a real example taken from the scientific literature. The biological phenomenon I will discuss is an instance of so-called *bistable switching* which plays a role in many important biological processes, for instance in the control of gene expression, in cellular differentiation, cell-cycle progression, and in neural signaling. It is thus representative for a class of phenomena that are biologically relevant and widely discussed among theoretically minded molecular biologists (see e.g. Bhalla and Iyengar, 1999; Ferrell and Xiong, 2001; Novak et al., 2007; Savageau, 2001). My aim is to show how in the explanatory practice of systems biology manipulationist reasoning about causal mechanisms is integrated with dynamical modeling. Notably, it will become clear

[7] In the theory of dynamical systems the investigation of equilibrium states when varying the parameter values is known as bifurcation analysis. This type of analysis is carried out to investigate the circumstances under which a system shifts between qualitatively different behaviors, not to explain the behaviors themselves.

that relationships of non-dependence are crucial for an understanding of systemic behavior, and not only used to establish the invariance of the causal or constitutive relationships. Conveniently, the philosophically interesting features of this example can be elucidated without going into the mathematical details.

At a particular stage during the process of egg formation in the frog *Xenopus laevis*, oocytes are arrested in an immature state. When exposed to the hormone progesterone, they undergo maturation and complete the first meiotic division. This maturation has been observed to occur in a switch-like manner: cells are either in the immature or in the mature state, but apparently cannot be in intermediate states for extended periods of time (Ferrell and Machleder, 1998). A crucial step in triggering maturation, and a convenient read-out, is the phosphorylation of the protein kinase p42 MAPK. When treating individual oocytes with intermediate doses of progesterone, Ferrell and Machleder observed either very high (>90 %) or very low levels (<10 %) of phosphorylated p42 MAPK. In the following they were interested in understanding how "a continuously variable stimulus—the progesterone concentration—is converted into an all-or-none biological response" (Ferrell and Machleder, 1998, 895). They hypothesized that the all-or-none character of the phenomenon is due to bistability, that is, the system can shift between two alternative stable equilibrium states. In what follows I will present the way in which these and other scientists have explained the switching behavior in oocyte maturation.

Bistability can arise in certain types of dynamical systems that involve nonlinear relationships.[8] It is easy to see how the existence of multiple equilibria is possible when we consider a rate balance plot which, in contrast with the case of simple equilibrium discussed in Sect. 4, is not restricted to straight lines. Figure 8.3 shows the balance of phosphorylation (forward reaction) and dephosphorylation (back reaction) of the kinase p42 MAPK in the oocyte system. The two curves indicate how the rates of these reactions depend on the fraction of phosphorylated kinase (denoted by A^*/A_{tot}). While the back reaction curve is simply a straight line, the forward reaction curve is essentially flat in the left portion of the plot and has a pronounced maximum when about half of the total amount of kinase is phosphorylated. The particular nonlinear behavior of the forward reaction curve is due to an underlying positive feedback: the phosophorylation reaction is slow unless there is a considerable fraction of kinase that is already phosphorylated.[9] Since the slope of this curve is initially less steep than the slope of the back reaction curve, there are three intersections of the two curves and hence three equilibrium points.

[8]Nonlinear relationships are not necessarily change-relating. It is perhaps as a result of the hype around chaos theory that 'nonlinearity' is usually associated with the idea that small changes can have large effects. However, in nonlinear dynamical systems the converse is also possible: large changes with negligible or small effects.

[9]Ferrell and Xiong (2001) suggest that several mechanisms are jointly responsible for this effect. Notably, p42 MAPK is involved in a positive feedback loop by contributing to the accumulation of Mos, its upstream activating kinase.

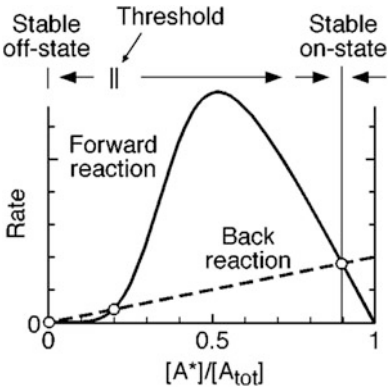

Fig. 8.3 Rate balance plot for the oocyte maturation model. Due to non-linearities the forward reaction is not a straight line. The three intersections correspond to three equilibrium points. The one in the middle is unstable. Note that, instead of balance of degradation and synthesis, equilibrium in this case requires equal rates of forward and back reaction. Reprinted by kind permission of AIP Publishing LCC, from Ferrell and Xiong (2001, 232)

The middle one is unstable, however, since in its vicinity the system will always be driven away from it, towards one of the two outer equilibria, which are both stable.

In this way we have established that there are two stable equilibria at low and high concentrations of phosphorylated kinase, respectively. These can be interpreted as *off* and *on*-states of a switch; but how can this maturation switch actually be effectuated? It turns out that such a shift from *off* to *on* occurs at a critical level of progesterone concentration because the *basal rate* of the forward reaction is proportional to the level of the activating progesterone stimulus. The basal rate is the rate at which the reaction would proceed in the absence of the positive feedback mechanism, and its dependence on the stimulus affects the shape of the total forward reaction curve as shown in Fig. 8.4. With increasing stimulus the *off*-state and the unstable equilibrium point move closer together until the two points finally coalesce. The curves corresponding to even higher levels of stimulus each have only one intersection with the back reaction curve. Therefore, if the system is initially in the *off*-state, it will at some critical level of stimulus jump to the *on*-state which then is the only remaining equilibrium. This explains the observed all-or-none behavior in the maturation of oocytes.

The behavior of the switch can be further illustrated by representing the position of the stable equilibria as a function of the stimulus (Fig. 8.5). This plot elucidates another important property of the switch: After the system has been driven from the *off*-state to the *on*-state by continuously increasing the stimulus, it will remain in the *on*-state even if the stimulus is subsequently withdrawn. Thus, once an oocyte has received a hormonal stimulus of sufficient size, it is irreversibly committed to maturation and does not shift back and forth between the two states. Obviously, this irreversibility is crucial for the reliability of developmental pathways.

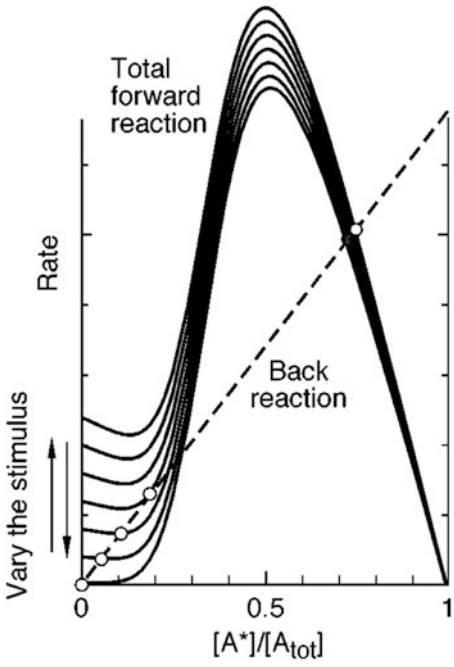

Fig. 8.4 Rate balance plot for varying progesterone concentration. Above a critical level the *off*-state disappears. Reprinted by kind permission of AIP Publishing LCC, from Ferrell and Xiong (2001, 233)

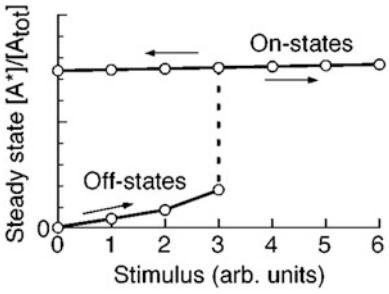

Fig. 8.5 Stimulus response curve. Once the stimulus reaches the threshold level, the system is locked in the *on*-state in which the concentration of the phosphorylated kinase A^* is always high. Reprinted by kind permission of AIP Publishing LCC, from Ferrell and Xiong (2001, 233)

Let us now try to understand in more detail how the given account explains the initiation of maturation. If we first consider only the "switching on" part of the story, we can represent the mechanism in terms of a simple causal relationship between two binary variables: A stimulus variable that can take on the values 'below threshold' or 'above threshold', and a kinase activity variable that accordingly assumes either of the values 'on' or 'off'. Obviously, this is exactly what we expect from a simple switch. Note, however, that this behavior is exhibited by a system with a high number of degrees of freedom. The simplicity of the behavior, as will be shown, arises from the fact that possible dependencies among the variables are removed or attenuated.

Let us go back to Fig. 8.4 from which we can infer how the total forward reaction rate curve changes as the stimulus is varied. The first thing to notice is that the

important changes concern only the lower left portion of the plot. Which are the relevant features for the behavior of the switch? First of all, it is necessary, as we have seen, that there exists a threshold level for the stimulus above which the curves do not intersect in this region of the graph. The value of this threshold is biologically important since it determines the sensitivity of the switch. A very low threshold would cause the system to shift already at small levels of hormone, which might lead to premature differentiation. Second, it is crucial that the highest value of phosphorylated kinase in the *off*-state, is not so high as to already activate the maturation process. As long as these conditions are met, however, the details of the relationship between stimulus and *off*-state concentration do not matter. The organization of the mechanism, notably the positive feedback, ensures that there is a range within which the level of phosphorylated kinase depends only weakly on the stimulus (corresponding to the branch of *off*-states in Fig. 8.5) and ensures that the system remains in the *off*-state even if the progesterone level is varied significantly.[10] Only around the threshold level there is sensitive dependence on the stimulus. But as soon as the system has switched, the level of phosphorylated kinase becomes virtually independent from the level of hormone (the branch of *on*-states in Fig. 8.5). The reason is that the stimulus does not significantly affect the reaction rates at high levels of phosphorylated kinase and cannot destabilize the equilibrium.

Thus in order to explain the phenomenon, we have to invoke both the non-dependence of the *off*- and *on*-states on the stimulus as well as the very sensitive dependence around the threshold. What the example shows is that the explanation of complex dynamical behaviors requires information both about relations of dependence and of non-dependence. In order to understand features of persistence, such as robustness or memory, we have to illuminate how some variables in certain ranges do not or only weakly depend on others. Moreover, this kind of knowledge allows us to explain how systems built of many parts may show behaviors that can be described in comparatively simple terms. The simplicity of the behavior at the level of the whole mechanism is due to the fact that many changes at the level of the components are not constitutively relevant, in the sense of not being change-relating. We cannot fully comprehend how this behavior is brought about if we restrict ourselves to information about manipulationist relevance. This suggests that the manipulationist conception of mechanistic explanation is insufficient to account for many aspects of phenomena that involve dynamical patterns.

As already noted, the description of the switching behavior itself represents a change-relating generalization. Therefore, it can be used as a basis for further explanation. For example, one may explain why one particular oocyte did not initiate maturation by referring to this generalization plus the fact that the given hormonal stimulus was not sufficient. Relationships of non-dependence partly account for the invariance of this generalization. They illuminate, for instance, why different

[10]Note however, that the role of feedback is not merely to confer robustness to the system. Instead, it is an integral part of the switching mechanism since in its absence the system would not show bistability in the first place.

oocytes initiate maturation even when given slightly different doses of stimulus. It might be argued, therefore, that information about change-relating generalizations is sufficient to explain the phenomenon of interest, and that information about non-dependence comes into play only if we want to generalize for further purposes of explanation. But as I hope to have shown, both kinds of relationships are in fact already used in the explanation of the basic features of the switch. The mechanistic explanation that shows *how* the system brings about the behavior contains answers to both types of Batterman's *why*-questions. Systems biologists want to understand the factors on which changes in observed dynamical patterns depend, but they also want to explain why these patterns obtain in the first place.

The discussion of this particular mechanism has touched upon the concept of robustness on several occasions. In the following section I will return to the idea of robustness as a fundamental property of living systems and show that relationships of non-dependence play an important explanatory role here as well.

6 Robustness and the Architecture of Living Systems

Investigating robustness is often invoked as one of the key motivations for research in systems biology. Hiroaki Kitano, for instance, holds that "[it] is one of the fundamental and ubiquitously observed systems-level phenomena that cannot be understood by looking at the individual components" (Kitano, 2004, 826). How does this idea of robustness as a fundamental property of living systems connect to the discussion about relations of non-dependence in the preceding sections? We have seen in the example of the switch that the particular dynamical organization of a mechanism can lead to weak relationships between variables or components, which in turn confers reliability and robustness to the system as a whole. However, in the just cited article Kitano notes:

> Robustness is often misunderstood to mean staying unchanged regardless of stimuli or mutations, so that the structure and components of the system, and therefore the mode of operation, is unaffected. In fact, robustness is the maintenance of specific functionalities of the system against perturbations, and it often requires the system to change its mode of operation in a flexible way. (Kitano, 2004, 827)

This seems to imply that it would be overly simplistic to explain robustness by referring to the causal or the constitutive irrelevance of particular factors under certain conditions. Instead, the quote suggests that the reliable performance of a system requires sophisticated underlying structures. A more refined view of the mechanistic structure of living systems would, therefore, consist in holding that robustness can be explained by invoking particular 'robustness mechanisms.' Indeed, Kitano mentions four different features that could play the role of such mechanisms: system control, redundancy, modularity, and decoupling (Kitano, 2004, 827). Even though it may not have been his intention, the fact that Kitano is speaking of "mechanisms that insure the robustness of a system" (Kitano, 2004,

827) suggests a particular picture: A living system may at its core be clock-like, but its reliable functioning in environments that are characterized by uncertainty and noise is guaranteed by an intricate machinery of additional features that has evolved around this core. If this picture were accurate, the general strategy of understanding mechanisms in terms of change-relating generalizations alone would be justified after all. Robustness would not be a fundamental property of the mechanisms themselves, but rather a separate phenomenon that could be explained by referring to independent mechanistic features. Yet, we have seen in the previous section that there are at least some cases where it is not possible to separate the explanation of a behavior from an explanation of its robustness. Moreover, a closer look at Kitano's alleged robustness mechanisms suggests that this conceptual separation might in general not be obvious. What he means, for example, by 'systems control' is the use of certain control strategies in the building of biological circuits, something that also fits the example discussed in the previous section. So his robustness 'mechanisms' are probably better understood as 'design features' of biological mechanisms. Just as in the case of oocyte maturation, robustness is often *in-built* and not in any obvious way *added* to the mechanism. In the remainder of this section, I will illustrate how mathematical modeling has recently been applied to elucidate the features underlying biological robustness of this kind.

In the attempt to simulate the interactions among the genes responsible for segmentation in *Drosophila*, von Dassow et al. (2000) developed a dynamical model and systematically investigated its behavior under changes in parameters. The segment polarity network described by this model generates a periodic expression pattern across cells early in development. Initially, von Dassow et al. had hoped that the requirement to reproduce the behavior of the target system would impose sufficient constraints on the model to obtain reasonable estimates for the nearly 50 parameters of the model. Consequently, they expected that only a relatively small subset among all the states in the high-dimensional parameter space would lead to biologically meaningful versions of their model. Strikingly, however, they found that solutions in this space were not rare at all:

> Among 240,000 randomly-chosen parameter sets we found 1,192 solutions ($\sim$ 1 in 200). This is very frequent; as this search involved 48 parameters, on average a random choice of parameter value has roughly a 90% chance of being compatible with the desired behaviour. (von Dassow et al., 2000, 189)

Apart from their abundance, solutions were apparently not isolated in parameter space. For many of them the model was found to be tolerant to variation of individual parameters over several orders of magnitude. Thus the scientists concluded that the model's ability to reproduce the target behavior was "intrinsic to its topology rather than to a specific quantitative tuning" (von Dassow et al., 2000, 189).

The case of the segment polarity network, therefore, supports the idea that robust behavior is not always achieved by adding structural components to an otherwise fragile mechanism. Robustness, therefore, cannot necessarily be analyzed as a separate feature, but instead appears to be entangled with a system's overall functionality. We have also seen this clearly in the example of the bistable switch

in Sect. 5, where the positive feedback is necessary both for the robustness and for the basic behavior of the mechanism. In general it seems that the molecular organization of biological mechanisms can lead to a weak dependence of system behavior on the behavior of the components. The following quote nicely illustrates how the scientists' initial assumptions about the robustness of the segment polarity network were overturned by their detailed investigation of the mathematical model:

> We originally expected the core topology to be frail and easily perturbed, and expected to achieve robustness only by adding additional complexity; we expected the reconstitution approach to tell us which architectural features confer robustness. Confounding that expectation, the simplest model that works at all emerged complete with unexpected robustness to variation in parameters and initial conditions. (von Dassow et al., 2000, 191)

Robustness of this kind does not seem to be restricted to the generation of developmental patterns in *Drosophila*. Gutenkunst et al. (2007) investigated 17 different systems biology models and systematically examined the sensitivity of their behavior to parameter changes. Their set of models covers a wide range of different biological mechanisms and, aside from von Dassow et al.'s network, includes models of circadian rhythm, metabolism, and signaling. In all of them they found what they call 'sloppy parameter spectra:' the behavior of the model is sensitive to variation along a few 'stiff' directions in parameter space, but insensitive along a large number of 'sloppy' directions. It is important to emphasize that these directions (or 'eigenvectors,' mathematically speaking) do not correspond to individual model parameters but rather to combinations thereof:

> Naively, one might expect the stiff eigenvectors to embody the most important parameters and the sloppy directions to embody parameter correlations that might suggest removable degrees of freedom, simplifying the model. Empirically, we have found that the eigenvectors often tend to involve significant components of many different parameters. (Gutenkunst et al., 2007, 1873)

This essentially means that the investigated systems do not react in a clock-like fashion to most perturbations on individual components. In order to bring about significant changes in systemic behavior, it is necessary to intervene on multiple components simultaneously.

The observed 'sloppiness' of biological systems provides resilience towards many disturbances at the molecular level and therefore appears to be beneficial for the survival of the organism, but it is not necessarily an evolved feature of living systems. Daniels et al. (2008), for example, conjecture that sloppiness might be a universal property of a particular class of dynamical models which naturally accounts for many types of robust behavior. With regard to von Dassow et al.'s model of the segment polarity network they note:

> The model is robust in these [sloppy] directions not because of evolution and fitness, but because of the mathematical behavior of chemical reaction networks, which are naturally weakly dependent on all but a few combinations of reaction parameters. (Daniels et al., 2008, 393)

In general, however, it is clear that the investigation of biological robustness must pay attention both to evolved robustness mechanisms and to generic features of biological organization.

The results of Gutenkunst et al. suggest that the relation between components and system behavior is often not as straightforward as analogies to machine-like mechanisms would make us believe. Information about non-dependence is highly relevant for the explanation of the behavior of many biological mechanisms, and their robustness is not just an interesting feature that requires separate explanation. Thinking about robustness may have a profound influence on the way in which we conceive of mechanistic explanations in the life sciences.

7 Conclusion

In this chapter I have tried to assess particular accounts of mechanistic explanation, according to which explanatory relevance relies on change-relating generalizations, by looking at dynamical modeling in systems biology. The motivating question was whether this manipulationist framework can adequately account for what we know about the robustness of living systems, a property that is extensively studied by systems biologists. I have argued that certain aspects of the explanation of dynamical patterns, first and foremost simple dynamical equilibrium, are not captured by approaches that solely focus on change-relating relationships. Instead, the explanation of such features relies on information about relationships of non-dependence, that is, on information about factors or relationships that are irrelevant from a manipulationist standpoint. By presenting the example of a bistable switch in the maturation of *Xenopus* oocytes, I have shown that this kind of reasoning is actually applied in mechanistic explanations as they are found in the scientific literature. I have tried to illuminate how in the case of this mechanism the discussion of robustness cannot be separated from its functional behavior. I have then turned to a more general discussion of the concept of robustness. There is evidence that at least some of the robustness we find in biological systems cannot be accounted for by invoking separate 'robustness mechanisms.' Instead, it can often be explained by the fact that interactions among components of a system are reducible to only a few significantly sensitive dependencies. Therefore, if we want to mechanistically explain how these systems work, we have to understand how biological organization results in weak dependencies and leads to coherent behavior and robustness at the systemic level.

Robustness is often presented as one of the paradigmatic examples of an emergent property; at least systems biologists frequently describe it as such. The ideas discussed in this chapter may shed some light on the reasons for this usage. Philosophical accounts of emergence have mostly focused on system properties that somehow 'exceed' the capacities of the components, or are unpredictable based on information about individual parts (e.g. Bedau, 1997; Boogerd et al., 2005; Kim, 1999). However, as the French philosopher Edgar Morin has noticed, a system is

not only more than the sum of its parts, it is also *less* than the sum of its parts in certain respects (Morin, 2008). The behavior of the components is constrained in various ways by the structure and organization of the system, which keeps them from exhibiting many of the properties that they might show in isolation or in different contexts. Robustness is thus a striking example of a restriction of component potential. What scientists mean by 'emergence' might often simply be the idea that the system is *different* from the sum of its parts. Restricting ourselves to change-relating relationships may prevent us from understanding how such kinds of emergent behavior are brought about.

I briefly discussed in Sect. 3 that the manipulationist picture might be defended by maintaining that a relationship of non-dependence simply expresses the fact that some element is explanatorily irrelevant. For instance, even though physics tells us that the current positions of remote stars exert a non-vanishing gravitational force on objects on the earth, we do not mention them in biological explanations because they are not considered to make significant differences to biological phenomena. However, many phenomena sensitively depend on factors that we would not want to include in their explanations either. The croaking of a frog, for instance, depends on whether or not the frog has just been run over by a car, yet we do not cite facts about cars when we explain how frogs croak. Acknowledging this problem of extravagant causes, Carl Craver resorts to a pragmatic solution by restricting himself to changes that *typically* occur in a system. But extravagant causes are a threat only if we insist on equating the notions of manipulationist and explanatory relevance. If it is true, as I have tried to argue, that information about causal or constitutive irrelevance can be explanatory, then the line between what is relevant for explanation and what is not must be drawn elsewhere anyway. I propose that there are interesting 'non-dependencies' just as there are uninteresting dependencies. The question of what makes a relationship interesting, however, may not be answered so easily.

Acknowledgements Apart from the editors, I would like to thank Pierre-Luc Germain, Mark Bedau, Jan Baedke, Sara Green, Veli-Pekka Parkkinen, and Robert Meunier for their helpful suggestions.

References

Baker, A. (2015). Mathematical explanation in biology. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 229–247). Dordrecht: Springer.

Batterman, R. W. (2002). *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence*. Oxford: Oxford University Press.

Bedau, M. A. (1997). Weak emergence. *Noûs, 31*(Suppl s11), 375–399.

Bhalla, U. S., & Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science, 283*(January), 381–387.

Boogerd, F. C., Bruggeman, F. J., Richardson, R. C., Stephan, A., & Westerhoff, H. V. (2005). Emergence and its place in nature: a case study of biochemical networks. *Synthese, 145*(1), 131–164.

Braillard, P.-A. (2010). Systems biology and the mechanistic framework. *History and Philosophy of the Life Sciences, 32*, 43–62.

Breidenmoser, T., & Wolkenhauer, O. (2015). Explanation and organizing principles in systems biology. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 249–264). Dordrecht: Springer.

Brigandt, I. (2015). Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 135–173). Dordrecht: Springer.

Cannon, W. B. (1932). *The wisdom of the body*. New York: W. W. Norton.

Craver, C. F. (2007). *Explaining the brain*. New York: Oxford University Press.

Daniels, B. C., Chen, Y. J., Sethna, J. P., Gutenkunst, R. N., & Myers, C. R. (2008). Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology, 19*(4), 389–95.

von Dassow, G., Meir, E., Munro, E. M., & Odell, G. M. (2000). The segment polarity network is a robust developmental module. *Nature, 406*(6792), 188–92.

Ferrell, J. E., & Machleder, E. M. (1998). The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science, 280*, 895–898.

Ferrell, J. E., & Xiong, W. (2001). Bistability in cell signaling: how to make continuous processes discontinuous, and reversible processes irreversible. *Chaos, 11*(1), 227–236.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(3), S342–S353.

Gunawardena, J. (2010). Models in systems biology: The parameter problem and the meanings of robustness, chap 2. In H. M. Lodhi & S. H. Muggleton (Eds.), *Elements of computational systems biology* (pp. 21–47). Hoboken: Wiley.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R. & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology, 3*(10), 1871–1878.

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., & Serrano, L. (2008). Evolvability and hierarchy in rewired bacterial gene networks. *Nature, 452*(April), 840–846. doi:10.1038/nature06847

Issad, T., & Malaterre, C. (2015). Are dynamic mechanistic explanations still mechanistic? In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 265–292). Dordrecht: Springer.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies, 95*, 3–36.

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics, 5*(11), 826–37.

Kuorikoski, J. (2007). Explaining with equilibria. In J. Persson & P. Ylikoski (Eds.), *Rethinking explanation* (pp. 149–162). Dordrecht: Springer.

Morange, M. (2015). Is there an explanation for ... the diversity of explanations in biological studies? In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 31–46). Dordrecht: Springer.

Morin, E. (2008). *On complexity*. New York: Hampton Press.

Novak, B., Tyson, J. J., Gyorffy, B., Csikasz-Nagy, A. (2007). Irreversible cell-cycle transitions are due to systems-level feedback. *Nature Cell Biology, 9*(7), 724–728.

Poon, W. C. K. (2011). Interdisciplinary reflections: The case of physics and biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 42*(2), 115–118.

Railton, P. (1981). Probability, explanation, and information. *Synthese, 48*(2), 233–256.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press

Savageau, M. A. (2001). Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos, 11*(1), 142–159.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Sober, E. (1983). Equilibrium explanation. *Philosophical Studies, 43*, 201–210.

Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. In *Proceedings of the 1996 biennial meetings of the philosophy of science association*, Cleveland (Vol. 64, pp. S372–S384).

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

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, S366–S378.

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

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy, 25*(3), 287–318.