Author:
Manas, Prakhar

Title:
Explanations in Biology

Perspectives on a Model-based science

EXPLANATIONS IN BIOLOGY

PERSPECTIVES ON A MODEL-BASED SCIENCE

Prakhar Manas

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Arts.

November 2020

Word count: Seventy-seven Thousand Four Hundred and Eighty-six

ABSTRACT

Biology is a model-based science. Biologists routinely offer explanations by building models of the systems that they wish to study. In this thesis, I explore the different ways in which biologists use models as explanatory tools. To do this, I explore how models explain within three biological domains – molecular biology, systems biology, and systems theory, and situate them within three explanatory frameworks – mechanistic, minimal model, and comparative-functional explanations. I argue that the explanations offered by models within these frameworks are distinct from each other, but that the explanatory imports of one type of model for explanatory frameworks other than the one they are constructed within, is significant. But that to understand how models explain in different frameworks, attention must be paid to their construction which utilises heuristic tools, abstraction, idealizations, and optimising assumptions. I do so by showing how mathematical models of systems biology can be used as heuristics for discovery and in tracking organisational features of mechanisms, but that reducing the role of mathematical models merely to this heuristic role is a mistake. By taking the example of bacterial chemotaxis, I argue that all three types of explanatory models are required to explain the features of the chemotactic network and, by extension, any complex biological system. Lastly, I argue for an explanatory pluralism in biology based on scientific practices as an epistemic tool to deal with the overwhelming complexity of biological systems.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: PRAKHAR MANAS DATE: 11th November 2020

Dedicated to

my wife Divya,
for helping me through the most challenging phase of my life,

&

my parents,
for all their support.

ACKNOWLEDGEMENTS

The past six years have been really challenging for me, and a lot of people have contributed in making them a bit less of a challenge.

First and foremost, I would like to express my gratitude to my supervisor – Samir Okasha, for all his help and guidance even while residing in different continents. Thanks also to Jonathan Gross and Tudor Baetu who acted as my supervisors at different points during my PhD. A special thanks to Anthony Everett, who was always there to talk to about more personal issues.

I would also like to thank the staff and student body at the Philosophy Department of the University of Bristol for all the lively discussions on every topic under the sun, for the illuminating seminars, and the wonderful feeling of community while I was there.

A special thanks to Kit Patrick – a champion of the Socratic method – for discussions that lead to a more in-depth exploration of many issues.

Finally, thanks to – my wife, my parents, my brother, and my three adorable cats for their support, encouragement, and love.

CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION

Waters (2014) has noted that the traditional approaches in philosophy of biology have had a theoretical bent where the major questions have been about theoretical concepts, theory reduction, theory relations, and adequacy of explanation from a philosophical perspective. Questions about reduction of Mendel's laws to molecular genetics; providing naturalistic definitions for genes, species, and function; and conceptualising a theoretical framework for explanations in molecular biology (amongst others) have formed the bedrock of philosophy of biology in the last century. But, more recently, there has been a significant shift in the attention of philosophers towards practices in biology[1]. Instead of theorising about the nature of explanations, how they can be justified, and how can they be extended to different domains of biology; philosophers have started paying attention to how biologists use theories and explanations in their practice. As a result, the question has shifted from how to construct a good explanation to how explanations offered by biologists achieve their goals. There has been a lot more attention on practices in biology including the study of how biologists control, predict, and understand the system that they are working on [Leonelli (2009)]. The perspectives on how biologists explain has introduced a pluralistic picture of scientific practices and the concept of explanation and understanding itself.

The aim of this thesis is to analyse how biologists explain through models. The general framework, thus, is placed under the perspective of modelling that has come to

[1] A recent example would be Green (ed.) (2017) which takes perspectives on systems biology from both scientists and philosophers.

be known as 'models as mediators' perspective [Morgan and Morrison (1999a,b)] which breaks from the traditional view of models as guided by theory to models being the focus of study in sciences. Models, in this view, are autonomous entities that are not restricted to theory application but have a wide range of functions. Therefore, the challenge in front of philosophers is not only to understand how models explain, but to understand how models are used by biologists in their practices. In the thesis, I present a study of diverse way in which biologists explain through models focusing on three aspects of modelling a given system and its behaviour – mechanistic, mathematical, and functional. And to show how all of them are important in the study of a given system. Furthermore, I show that these models do not work in isolation but have consequences for frameworks other than the ones they are constructed under.

A clarification must be made here. The thesis draws from the work that the biologists have done as published. The examples that are used and the models that are analysed derive explicitly from readings of scientific literature and not from observing scientific work in any laboratory. In this respect, scientific practices, as used in this thesis is the study and analysis of the final work presented by the biologists and not an examination of 'science in practice' – which would take an anthropological dimension into consideration. The aim, thus, is to look at the final models/explanations presented by scientists, in an attempt to understand what their models achieve and how they explain a particular phenomenon, without taking the pragmatic dimensions decision making, problem solving, and routine scientific work that takes place in a laboratory.

Another recent development in biology has been the advent of big data in form of omics (genomics[2], proteomics, metabolomics). Omics is the systematic study and

[2] See Güttinger and Dupré (2016) for definition and analysis.

quantification of data about different biological molecules that define the structure and function of the organism. The huge repository of data that has been collected since the turn of the century through projects like genome projects that have sequenced the genome of different species. This has led to a new way of constructing models and explanations in biology. The nascent field of systems biology has seen huge progress both due to availability of the data and increase in computational power in the recent years. Technological advancements in science are generally associated with a leap towards progress. For example, the improvements in microscope in the 17th century by Robert Hooke radically changed how organisms were studied. In other cases, the improvements in technology have aided scientific revolutions. Galileo's improvement of telescope and the subsequent study of phases of Venus contributed greatly to the study of heliocentrism.

Although technology plays a great role in scientific progress, it should not be taken as an advent of new paradigm on face value. Systems biologists generally see their study of the system as more 'holistic' than traditional approaches of molecular biology. The claims are generally derived from the observation that systems biology studies more complex systems than molecular biology and accounts for interaction between different parts which is important to understand the system's functioning[3]. While the validity of such claims is very much debatable, the main difference between systems biology, and traditional approaches in molecular biology is in their methods of modelling the system in question. Therefore, instead of focusing on the two disciplines and their differences, I will be focusing on the different ways of modelling in mechanistic explanations and

[3] See Powell and Dupré (2009) for a critical analysis.

the explanations provided by mathematical/computational models and show how both differ from each other with respect to their methods, aims, and focus of study and therefore, on the kind of explanations they provide. This is not to say that the different frameworks cannot have explanatory utility for each other – they do, and I will highlight how – but that to properly understand how they can be utilised by the explanatory strategies used in each other, we must first understand what they precisely explain and therefore, how they can be complementary and informative about each other.

Lastly, I explore the capabilities of systems theory to provide avenues for explanations in biology. Systems theory in biology has a long history that has worked in tandem with, and many a times independently of, the more experimental and research-oriented streams of biology. One of the aims of systems theory is to explain aspects of biological systems and how they differ from the realm of non-living. In this respect, I explore a more recent idea of how to characterise biological organisations. The idea focuses specifically on how biological systems achieve autonomy through taking control of their own constraints. I explore the idea of constraints and show that the general theoretical framework – dubbed the "Closure of Constraints" – has practical explanatory import based on the elucidation of regulation, robustness, and adaptive behaviour under the explanatory framework of comparative functional explanations.

The structure of the thesis is as follows –

In *Chapter 1*, I introduce some basic terminological and conceptual clarifications that will be instrumental in constructing the arguments in the coming chapters. I give a sketch of what models in biology are, and some of their characteristics and aims. I further argue for explanatory pluralism based on the plurality of models, plurality of

questions asked by biologists, and complexity of the biological domain. Finally, I analyse the role of heuristics, idealization, and abstraction in modelling exercises.

In *Chapter 2*, I introduce the concept of mechanistic explanations and explore a few problems for the framework. I argue that in the light of these problems, the ontic conception of mechanistic explanations in its stronger form is untenable with the scientific practices. But that a milder, limited view of the ontic conception in form of normative constraints on models can be accommodated. I, therefore, argue for a multi-model approach to mechanistic explanations (and explanations in general).

In *Chapter 3*, I provide an introduction to the field of systems biology and discuss the various modelling exercises that fall within its ambit. I further analyse how these models can aid the construction of mechanistic models but note that this is not how the mathematical/computational models of systems biology derive their explanatory powers. Instead, I argue that they explain by the virtue of constructing minimal models that show how some of the features are irrelevant for the derivation of the system's behaviour.

In *Chapter 4*, I introduce the idea of closure of constraints as a theoretical framework for biological systems and argue that a weak ontological conception of emergence is needed to understand the concept of constraints. I further argue that the constraints thus construed are important to understand closure, control, and regulation. Finally, I show that the explanatory imports of the closure of constraints idea can be presented within the explanatory framework of functional explanations through the concept of design explanations.

In *Chapter 5*, I take the example of chemotaxis in *E. coli* to show how different explanatory frameworks, as explored in the rest of the thesis play important part in explaining the chemotactic system. This will be done by showing the limitations of each model in capturing all the aspects of the chemotactic network.

The hope is that, through this thesis, I can show that models –

a. Play diverse roles in explanations in biology,
b. Are constructed within their explanatory framework and therefore must be judged according to the norms of the framework and the explanations they offer, and
c. Are not completely explanatory for the system in question, rather only model/represent the system in limited ways.

All these considerations should pragmatically force us to embrace explanatory pluralism in model-based sciences like biology.

Lastly, I explore the avenues for these different models to be integrated, thus taking a step towards 'Integrative Pluralism.

CHAPTER ONE

Models, Explanations, and Complexity in Biology

The aim of this chapter is to provide a general framework within which the arguments of the rest of the thesis will be made. Since the thesis consists of comparing different explanatory practices in biology – encompassing different domains of molecular biology, systems biology, and systems theory – a general framework needs to be presented for some relevant ideas – that of explanation, modelling, representation, and the methods and epistemic tools used in these endeavours. The structure of the chapter is as follows. First, I consider modelling in biology to present a few general characteristics of what a model is, what it must do, and how it relates to other models and explanations. Although, it will not be an exhaustive account of modelling, it will be sufficient for the analysis provided in the thesis. Following this, I argue for explanatory pluralism in biology based on scientific practices and diversity of models. A further argument will be provided for pluralism based on the idea of complexity and how it is captured in our representations. This will also enable me to assess another aspect of how representations of complex system force us to adopt various heuristic strategies. I end the chapter with a discussion on abstraction and idealization.

Before I start, there is some basic terminology that I must clarify. I use the term 'system' in context of biological organisation throughout the thesis. What I mean by systems is synonymous with organisations. A system is a collection of parts and processes that construct an organised whole such that it shows some behaviour. While a biological system can vary from a cell up to an ecosystem and even the whole biosphere, and the explanatory models can be equally applied to these levels, I tend to

focus mainly on what can be called ‘organismal system’ – i.e. systems that can be construed primarily at the level (or sublevels) of organisms.

Another consideration must be paid to the examples used throughout the thesis and the motivation behind the selection of these particular examples. Most of these examples are motivated by the literature that has been studied for this thesis. The motivations for the particular examples are also rooted in how they explicate some of the modelling methods that have been used, but at the same time, can be generalized to (if not all) some particular classes of systems. In chapter 2, for example, I have used the example of Axis formation in *Xenopus*, as a case to show how mechanistic modelling might be unsuitable for developmental systems in general. The criticisms raised by this example are applicable to other developmental systems like morphogen gradient during morphogenesis (a process ubiquitous to development in many species). Similarly, in chapter 3, many examples are used as vehicles to discuss some features of particular models. But the use of these examples does not mean that they are specific to the system in question, but they can essentially be generalized. For example – the persistence detector in bacterial arabinose system shows an architecture that has been found in various systems and acts in similar ways despite the difference in constituents. Similarly, topological models show some underlying dependencies that can be generalized for various systems that show similar topology. The examples are, thus, illustrative of modelling techniques that can be applied to a wide variety of systems. The examples used in chapter 4 are even more generalizable. The closure of constraints theory, as discussed in this chapter, aims to provide very generalized necessary conditions for the existence of biological systems. The examples, like glycolytic pathway, lac operon, glucose homeostasis, are used to portray these very generalized

necessary conditions that show the ubiquity of design explanations, regulation, and robustness as applied to specific cases. Finally, the example used in chapter 5 – that of bacterial chemotaxis – is motivated by the abundance of literature present about this system. This chapter aims to show how multiple model systems might be required to explain biological phenomena. While the focus, throughout the chapter, is on bacterial chemotaxis, I believe that what is analysed and proposed in this chapter applies to biological systems in general due to the ubiquitous presence of regulatory systems that confer robustness to the system and the complexity of the system which opens up avenues for construction of different models of these systems. Thus, the example is again a paradigmatic case used to make general claims about biological systems.

1.1: On Models

'Modelling' is a ubiquitous practice in sciences. It would be uncontroversial to note that scientists, including biologists, engage in the exercise of model construction in an attempt to understand the target system. Models in science can have various functions – it can represent a system, it can offer explanations about the target system, and it can help scientists intervene on the system. But modelling as an exercise and as a theoretical unit in sciences has encountered various problems in describing –

1. What a model exactly is?
2. How do they represent a system?
3. How are they related to the theory? And
4. How do they offer genuine explanations and understanding about the target system?

All these questions have been major themes in philosophy of science and have produced a rich and diverse literature. Providing an exhaustive account of all these positions, arguments, and views is beyond the scope of this dissertation. Instead, I will focus on themes that are pertinent to the concepts of models and explanations explored further in the thesis. My aim, thus, is to provide a general idea of what a model is, and of what models do in terms of offering explanations and fostering understanding of the target system. The exploration of these questions will inevitably touch upon all the four questions (1-4) mentioned above, although, I will not attempt to give a definitive answer to all of them.

To start with, different conceptions of theory in science assign different roles to models. In this respect, I focus on the role of modelling as presented in the semantic view of theory and the pragmatic view. The former construes theory as a family (sometimes hierarchical) of mathematical models. There are various interpretations of theory within the ambit of the semantic conception. Suppes (1960) presents a set-theoretic approach to mathematical modelling. On this approach models are "a set-theoretical entity which is a certain kind of ordered tuple consisting of a set of objects and relations and operations on these objects"[4]. A more pertinent idea to our discussion though, would be the state-space approach by Van Fraassen (1980). As per this approach, models are state spaces associated with dynamical equations [Weisberg (2013)]. According to this approach, real systems that are the focus of our studies, can be said to be in particular states (that provide the description of properties of the system) at a particular time. As time progresses, the system might transition from one state to another. Therefore, the system will transition through what can be called a state-space

[4] Suppes (1960), p.290

– or the collection of all the possible states that a system can be in. These states are defined by variables that the system has. For example, we can talk about the thermodynamic state of a system which will be defined by certain variables like temperature, pressure, volume, energy etc. Over time, the system will evolve from one state to the next – for example, in the absence of outside interference, the system will move towards an equilibrium where its total energy will be minimized. Therefore, a system is seen as transitioning from one state to another as time progresses according to some laws (in this case, the laws of thermodynamics). These laws then limit the possible trajectories that a system can take through the state space. A model, in this view, is the representation of one of the possible trajectories that the system might occupy in the state space. Laws in the semantic view are therefore constraints on the possible trajectory in the state space with models representing the possible trajectories. The theory is accurate (empirically adequate) if it represents the trajectory according to various laws (like the law of succession, laws of coexistence, and laws of interaction) that can be embedded in some model of the theory, where the relationship between the model and the real system is that of isomorphism.

The semantic view to theory thus, places models at the centre of defining the application of theory. In such a view, the laws are constraints on the state space, and models are the representations of the actual trajectory of the system in this restricted state space.

While I think that the semantic view of theory and (by extension) models captures some aspects of the role models play in representing the system in question, there are some problems with the view. Many of these problems have been highlighted by what has been come to known as the “Model and mediators” perspective of modelling under

the general purview of the pragmatic view of theories. As per the 'models as mediators' perspective, models are not just 'models of theories' but autonomous entities that are used by scientific practices in various ways. Therefore, in general, the view enables a more practice-based account of models (and explanations) which I believe is a strength of the view.

Morgan and Morrison (1999a, b) express a view of models that they believe captures the way scientists use and reason about them. In this approach to models and modelling are dictated more by the practical consideration rather than the theory. 'Mediators' here specifically refer to the idea that models can lie between theory and the real system (or data) but must be seen as independent entities in themselves. As Morgan and Morrison (1999b) notes –

> "It is common to think that models can be derived entirely from theory or from data. However, if we look closely at the way models are constructed we can begin to see the sources of their independence. It is because they are neither one thing nor the other, neither just theory nor data, but typically involve some of both (and often additional 'outside' elements), that they can mediate between theory and the world"[5]

The outside elements alluded to can include metaphors, analogies, policy views, stylized facts, mathematical techniques, and mathematical concepts. What elements play a what role in a specific model in a specific field of research and how prominent is the role cannot be known a priori but must be explored for the specific field and specific models.

[5] Morrison and Morgan (1999b), p.10.

A significant way in which the mediator view differentiates itself from the semantic view is that models are not necessarily mathematical as per the mediator view. In the semantic view, many authors have stressed that models are ultimately mathematical. Van Fraassen (1970) writes – "this meaning structure has a representation in terms of a model (always a mathematical structure, and must usually some mathematical space)"[6]. Morrison and Morgan (1999b) on the other hand, note – "[m]odels may be physical objects, mathematical structures, diagrams, computer programmes, or whatever, but they all act as a form of instrument for investigating the world, our theories, or even other models"[7].

Three consequences stem from the mediator view –

1. Models are independent entities that are neither completely dictated by the theory or by the data. In this respect, models can be called 'autonomous'[8]. Autonomy here, is intended to capture the idea that models are independent from theory, and have their own life, representative capacities, and explanatory roles.
2. Models can have various functions that are dependent on how they are used. They can be representations of the world, representations of the theory, instruments for measurements and interventions, simulations, etc. Furthermore, there is a dual plurality of models. Models are internally pluralistic in terms of concepts used in them (metaphor, analogy, assumptions, etc.) and can be externally pluralistic in the sense of what they explain (mathematical, mechanistic, historical). This plurality of modelling enables different kinds of

[6] van Fraassen (1970), p.327.
[7] Morrison and Morgan (1999b), p.32.
[8] Morrison (1999).

ways in which modellers can learn about the world, theory, and the model itself (representation, comparison, simulation, etc).

3. Lastly, the mediator view represents disunity of science. Models are piecemeal investigation tools and can be used to understand and to intervene on the system in several different ways. Modellers can ask different questions about the world and aim to answer them with myriad representations that capture some aspects of the world. By shifting the focus to practices, unity of science and scientific theories become secondary perspectives which may be or may not be the expectations that the modeler has from a particular model.

Within the models as mediators perspective, three important consequences come forth. Firstly, models are not in isomorphic relation with the system they are models of. Instead, representation constitutes 'a kind of rendering – a partial representation that either abstracts from, or translates into another form, the real nature of the system or theory, or one that is capable of embodying only a portion of a system"[9]. This allows the account to be permissible of various kinds of models many a time of the same phenomenon. The upshot of this is that models show variability – both in their construction and aims. The variability can be seen, especially in biology, where models can take the form of model organisms, diagrammatic schema, material models, mathematical models, etc. I will be exploring the construction and aim of some of these types of models throughout the thesis. The permissibility of models as mediators account to incorporate these models – both as the study of the phenomenon and for biological theorising provides a tool for understanding their epistemic content as modelling exercises envisioned by biologists to explore the phenomenon.

[9] Morgan and Morrison, (1999a), p.27.

Secondly, the models as mediators account provides a way of thinking about models as tools for intervention. Since within the account, models are mediating instruments between the world and theories, models can provide a way to gain insights about the world and construct more intelligible theories. As Morgan and Morrison (1999a) write, 'models are not passive instruments, they must be put to work, used or manipulated'[10]. This can be widely seen in biology, where both material models – like model organisms, and conceptual models – like mathematical equations, are manipulated to understand the phenomenon in a better way. Biologists routinely intervene on their models to produce more intelligible explanations about the phenomenon. This can be in the form of knock-out experiments in model organisms, intervening on causes in mechanisms, or modifying mathematical equations to change the system's behavior. Manipulation of models, thus, opens up the avenue to learn something about what the model represents. Explanations therefore, are dependent on model manipulation to explain certain features of the represented system. Models can be manipulated in various ways depending on their construction and their ontological status. As Leonelli (2007) notes, there are inherent differences between material model manipulation and conceptual model manipulation. Nonetheless, both these kinds of models can be subjected to manipulation to either learn something about the represented system, or to construct theories. The implication of this is simple – models must be "put to work"[11] to build explanations.

This thesis is, therefore, situated within this larger context. The explanations explored here embody the two ideas from the models as mediators account. Firstly, models explored here are representations of a system or phenomenon with variability in

[10] Morgan and Morrison, (1999a), p.32
[11] Morgan and Morrison, (1999a), p.33.

their representation. Models explored throughout the thesis are mechanisms, mechanistic schemas, mathematical equations, simulations, and systems models built on ontological considerations of the system. All these models represent the systems in various ways incorporating abstraction and idealization to varying degrees. These models, thus, provide a way to represent and analyse the system in multiple ways and for various purposes. The relation between the represented and the representative model is therefore not isomorphic. The representational content depends on the utility and the construction of the model.

Secondly, models are useful as long as they are open to manipulation – both in their construction and their use. The construction of the model itself is an epistemic exercise that requires constant tinkering with the model to achieve the intended goals. The use of the model is to intervene on the system in specific ways to gain understanding and to construct explanations and theories. Throughout the thesis, I will show how biologists and philosophers think of models as instruments for intervention. The aim for models is thus, not only to represent but to construct explanations that capture something about the world out there. This is achieved through interventions – tinkering of models to understand how the modifications affect the studied system and the phenomenon. This can be achieved through intervening on mechanisms, mechanistic schemas, mathematical equations, proposed designs, etc. For example, the ideal intervention conditions explored by Woodward (2003) as a theory for causation plays a big role in understanding how mechanisms are productive. By intervening on proposed mechanisms, we can establish causal relations which are necessary to bring about the phenomenon. Similarly, by intervening on mathematical equations, we can understand

the modalities of the system and show why certain structural features are necessary or desirable for a certain behaviour.

Through both these conditions, we can see how models are instrumental in scientific practices, where their construction and use guides constructions of intelligible explanations and theories.

1.2: On Explanations

Giving an account of explanation has been one of the central topics in philosophy of science. This includes questions like – what is an explanation? When can we say that an explanation has been provided? And more lately, how do models explain? I will primarily focus on the last question as it is the most relevant to my thesis.

Historically, accounts of scientific explanations have tried to give an objectivist criterion for explanation in science. Therefore, the notion of scientific explanation consists of two questions in general –

1. What is a scientific explanation?
2. What distinguishes scientific explanation from other types of explanations – i.e. what is a good scientific explanation?

The history of what explanation entails, and how to differentiate between scientific and other forms of explanations is long and diverse. A pertinent entry point to this long history, though, can be provided by a discussion of the Deductive-Nomological (DN) model provided by Hempel and others [Hempel and Oppenheim (1948), Hempel (1965)].

According to the DN model of scientific explanation, scientific explanations consists of two components – the explanandum: sentences that describe the phenomenon to be explained, and explanans – sentences that are "adduced to account for the phenomenon"[12]. Explanans explain the explanandum if the explanandum can be shown to be the logical consequence of the explanans. The logical consequence has two requirements. First is 'deductive' – that the explanandum is a conclusion of the premises in the explanans. And second, 'nomological' – that the explanans must contain at least one "law of nature" in the premise. As Hempel (1965) writes –

> "[A] DN explanation answers the question "*Why* did the explanandum-phenomenon occur?" by showing that the phenomenon resulted from certain particular circumstances, specified in $C_1,C_2,\ldots,C_k$, in accordance with the laws $L_1,L_2,\ldots,L_r$. By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred."[13]

As many have noted, the DN account is problematic. I do not wish to go into the general problems with the DN account, but specific criticisms as applicable to explanations in biology should be addressed. Many biologists and philosophers have argued that there are no laws in biology (at least not in the paradigmatic sense) that apply to wide variety of cases. While biologists might cite 'laws' like Mandel's Laws of segregation, it has been established that such laws apply to limited cases. This lack of laws would mean than many explanations in biology will not be counted as genuine

[12] Hempel (1965), p.247.
[13] Hempel (1965), p.337. Italics in the original.

explanations. A way around this might be to point out that even if there are not laws, there can be generalizations that are invariant under interventions in biology [Woodward (2003)]. Therefore, explanations can be genuinely explanatory when they allude to such generalizations. But it will still leave many 'explanations' than scientists aim to give in biology outside the purview of genuine explanations. Moreover, philosophers, like Leuridan (2010) have cast doubts if this view can replace the laws of nature.

Other accounts of explanations have been provided based on other criteria of what an explanation should do. A unifications account by Friedman (1974), and subsequently, Kitcher (1989) see scientific explanations as providing a unification of a range of phenomena. As Kitcher notes –

> "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again and in demonstrating this, it teaches us how to reduce the number of facts we have to accept as ultimate"[14]

For example, Newton's theory increases the understanding of our world because it brings various phenomena – like the movement of celestial and terrestrial bodies – within the same set of principles. Therefore, explanations increase the understanding of the world by deriving description of different phenomena using as few arguments patterns as possible. Furthermore, this process must be repeated again and again to provide more stringent argument patterns that can describe different phenomena. In this

[14] Kitcher (1989), p.432.

way, we can provide a more unified explanation every time we are able to explain phenomena with more stringent argument patterns.

On the other hand, Salmon (1984, 1997) describes what he called the 'Causal-Mechanism' (CM) model of explanation. According to the CM model, explanations increase our understanding by showing how the underlying causal-mechanisms bring about the phenomenon. According to the CM model, an explanation of an event E must trace the causal processes and interactions[15] that lead up to E and describe these processes and interactions that make up the event. The former (tracing of the processes and interactions) forms the etiological aspect of the explanation, while the latter (description of the processes and interactions) is the constitutive aspect of the explanation. By doing so, the CM explanation shows how the event fits into the "causal nexus"[16].

While a critical analysis of both these approaches is beyond the scope of this thesis, the common feature of these models is that they view explanations as satisfying an objective criterion. All three accounts provide features of what a good explanation is and claim that any explanation that abides by these features explains the phenomenon. In contrast, many have argued that such a stringent condition cannot be expected to be met in scientific practice – especially in biology – and therefore have argued for an explanatory pluralism.

[15] Causal processes are physical processes that are capable of leaving a 'mark' or local modification. Processes can transmit this mark through causal interactions. Causal interactions are spatio-temporal 'intersection' between two or more causal processes.
[16] Salmon (1984), p.9.

Explanatory pluralism is the thesis that there is no one way to provide or assess what explanations must be and what they must do. The plurality of explanations in biology is argued for, because of two factors –

1. *Scientific practices* – scientific practices in biology are diverse, and so are the aims and methods. Biology as a science studies various phenomena – ranging from cells to ecosystems, and can have various domains in mind – historical, causal, functional. Furthermore, biologists aim to provide explanations for different types of question – many times about the same system/phenomena (for example, proximate and ultimate explanations or Tinbergen's four questions[17]).
2. *Different models in biology prioritize different explanatory aims and use different heuristics and tools* – for example, mechanistic models generally aim to offer a causal explanation using Woodward's (2003) interventionalist approach, but many biologists would disagree that this is the only kind of explanation that is possible. Humenan (2010), for example, argues that topological explanations in ecology are not mechanistic (a similar application has been considered in Chapter 3 regarding network analysis) but structural. While mechanistic and topological explanations still are representational models – i.e. they represent the target system albeit at different level of abstractness and with different idealizations, other explanations may construct false models that do not have any target system in reality. An example of this would be that of

[17] Mayr's (1961) proximate and ultimate explanations introduces the distinction between two kinds of explanations we can offer about a trait – proximate explanations offer explanation about how the trait works, ultimate explanation about a trait is what fitness advantage it provides. Tinbergen's (1963) four questions about traits include - Causation: what is the mechanism by which the trait produces its effects?, Survival value: how does the trait contribute to the organism's fitness?, Ontogeny: how is the trait constructed in development? Evolution: 'the elucidation of the course evolution must be assumed to have taken, and the unravelling of its dynamics.' [Griffiths and Stotz (2013), p.173]

three (or multi) sex species [Diéguez (2013)]. While these models do not provide representative content, they can answer questions about phenomena that do exist. By studying a multi-sex species, we can answer questions about why two-sex species is observed in the world. A similar explanatory model will be considered in Chapter 4 called 'Design explanations' [Wouters (2007)]. Similar to the multi-sex species example, design explanations are contrastive explanations that answer functional questions - like why terrestrial organisms do not have gills? By asking such questions, we are imagining terrestrial organisms with gills and showing how constraints placed by the internal organization and environment will either not make it possible or decrease the life chances of the imagined organism (see Chapter 4 for more details). By constructing and studying such models, while we do not represent a system that exists, but through comparative analysis we can still understand the features of what exists.

These two points show that the explanations in biology do not follow an objective feature that can be applied to all kinds of explanations in biology, instead what we must consider is how different models provide explanations of the phenomena they deal with. Therefore, the primary consideration of this thesis is to show how different models in different practices in biology explain. A cautionary note though, is that the presence of explanatory pluralism does not mean that there are no objective standards through which we can examine explanatory capacities of different models. It rather means that an all-encompassing objective standard does not exist. Therefore, to understand how different models explain, we must consider their context-dependent application – including their assumptions, idealizations, abstractions, and heuristics.

1.3: Biology and Complexity

Biological systems are complex systems. But what do we mean by this? In general terms, we can point out that biological systems – cells, organs, organisms, ecosystems consist of various components interacting in various ways such that they appear to be organized towards performing certain functions or forming structures that are dynamic in nature – i.e. show behaviour that changes over time. But complexity itself is a hard concept to ground in theoretical terms. As Ladyman, et al. (2013) note, various authors have ascribed different characteristics to complexity which include but are not limited to non-linear interactions, emergence, hierarchical organization, robustness, multiple realisability, etc. Ladyman et al. (2013), provide their own minimal definition of complexity based on the measurement of complexity as statistical complexity and computability [see Ladyman et al. (2013)]. As per this definition –

> "A complex system is an ensemble of many elements which are interacting in a disordered way, resulting in robust organization and memory."[18]

This necessary but not sufficient condition for complexity of a system is derived from minimal conditions that a complex system must show. Firstly, the interactions between elements of the complex system must be of various kinds. Secondly, from the seemingly disordered interactions between the elements, a robust organization can be achieved such that patterns or structures are formed. For example, the Bénard convection cell forms a pattern of rotation between the liquid molecules such that a macroscopic pattern is observed shows stability on the relevant timescale. Finally, by achieving a robust

[18] Ladyman et al. (2013), p.57

organization, the complex system shows ‘memory’ such that the organization is able to return back to its stable state after certain perturbations.

But, even with a characterization like this, complexity remains a difficult property to define in general terms. As Hooker (2011b) notes, complexity studies as a branch of scientific theorizing still is in its infancy and struggles with various limitations including a lack of mathematical framework, and a general lack of necessary or sufficient conditions. The term remains vague to some extent, and as Hooker (2011b) notes, the problem is compounded in biology where we “lack a general or foundational understanding of biological organization — whether cellular, organismic, communal or ecological — and of autonomy in particular”[19]. This points to two problems of applying the understanding of complexity to biological systems. Firstly, the concept of a general complex system is too narrow to apply to biology. Biological systems show autonomy such that they are able to self-maintain, regulate, and control their own organization in a wide range of conditions. While the general concept of complexity is able to capture some aspects of this problem like self-organization – a phenomenon found in both biological and non-biological systems – it is unable to account for the characteristic properties of biological systems. As I will argue in Chapter 4, an account specifically for biological organizations exists which takes autonomy into account and has explanatory imports for understanding why autonomy in biological systems is realized in a particular way, the idea of complexity simpliciter is insufficient for explaining this feature of biological systems. Secondly, the measure of complexity itself is still not possible [Hooker (2011b, 2013)]. While concepts like statistical complexity,

[19] Hooker (2011b), p.844.

computation, or thermodynamic depth might provide us with some conceptions of complexity, they are not truly useful in biology.

Adami (2002), for example, considers what complexity means in biology and if there is a trend in evolution towards an increase in complexity. As Adami (2002) notes, biologists consider different measures of complexity – specifically structural and functional complexity. McShea (1996) for example, studies structural complexity in metazoans using a narrow view of how differentiated a system is – the more differentiated a system is in its parts and interactions, the more complex the system itself is – in an attempt to make sense of what biologists mean when they say that complexity has increased in metazoans. In this way, McShea (1996) defines complexity in terms of objects, processes, and hierarchy, such that different measures of complexities can be achieved according to the organization. But as McShea notes, there is nothing conclusive that can be derived from this study. Similarly, functional complexity has been elaborated by McShea (2000) as a calculation of functions using parts as proxy, such that more complex organisms perform more functions. But, again, as McShea (2000) points out, these measures still require assumptions about the correlation between the number of parts and functions –

> "The argument here is that a functional system needs to be both integrated internally and isolated from its surrounding, and therefore in organisms, functions are expected to be localized, to some extent, in parts. Further, most parts will participate in a few functions, and few parts will have many functions overlapping in them."[20]

[20] McShea (2000), p.661.

The analysis done by McShea (1996, 2000) points to how biologists try to make sense of complexity, but also highlight the assumptions that need to be made to come at such conclusions – hence different justifiable perspectives may lead to a different picture of complexity. As Winther (2006a) notes, this might very well be the case. Winther (2006a) shows how different areas of study in compositional biology use different theoretical perspectives to define and delineate parts. Even though compositional biology is driven by part-based explanations where parts can be structural, functional, or processual; the partitioning framework is defined by theoretical units [Wimsatt (1972), (2007)]. These theoretical units form the perspectives that are driven by the explanations that the biologists working in the particular fields aim to offer. For example, in comparative morphology, the parts are structural since the explanatory account aims to compare homologous structures in different species. In contrast, in functional morphology, the partitioning framework is activity-based where the considerations for defining parts are based on the questions that functional morphologists aim to answer – i.e. how a function is performed.

Therefore, as Winther (2006a), shows, different partitioning frameworks in part-based science like compositional biology can be derived from applying different theoretical perspectives and answer different questions. Therefore, we cannot 'glean' directly into the makeup of the system without first defining some theoretical perspectives and related modelling methods. This leads to a problem for studying complexity in biological systems – there are several ways to model a complex system such that they have their inherent theoretical perspectives that are in service of the kind of explanation they aim to offer and therefore will divide the system in different ways. Wimsatt (1972, 2007) offers two encompassing ideas of complexity that can be seen as

proxies for exploring the complexity of the system. The first is *descriptive complexity* which appeals to the idea that different theoretical perspectives, explanatory requirements, and practices might pick out different parts that may not spatiotemporally coincide. Wimsatt calls these ' K decompositions' where $K(T)_i$ are decompositions relative to a range of theoretical perspective T_i. In simpler systems, the $K(T)_i$ will be relatively overlapping. But in complex systems that might not be the case and $K(T)_i$ may differ with the range of theoretical perspectives T_i.

The second is *interactional complexity*, which as per Wimsatt (2007) considers the system in a state-space representation that represents the causal interactions of the variables within the system. We can describe the system in terms of parameters that depend on the location of the system in phase state and the inter and intra-systematic relations between the subsystems. As Wimsatt (2007) notes, each theoretical perspective will pick out different properties of the system and therefore pay attention to a different set of variables. This is dependent on how they establish intra and inter-system relations. Different perspectives will neglect some relations that fall below a certain threshold and pay attention to only the stronger relations.

The idea that different theoretical perspectives can provide us with different ways to understand a complex system leads to a further ramification for understanding complexity. Many authors have argued that complexity should not only be understood as an intrinsic property of the system, such that certain general criteria can define it but as how they are represented in our models. As per this view, the complexity of the system is closely linked to the number of different ways it can be represented. This is derived from the idea that complex systems are difficult to represent, and no one model can capture all the features of the system. Badii and Politi (1997), for example, write -

> "[T]he concept of complexity is closely related to that of understanding, in so far as the latter is based upon the accuracy of model descriptions of the system obtained using a condensed information about it. Hence, a "theory of complexity" could be viewed as a theory of modelling… a system is not complex by some abstract criterion but because it is intrinsically hard to model, no matter which [mathematical] means are used."[21]

As the quote suggests, the complexity of a system can be found in our representations and models of the system such that a complex system will provide avenues for many models from different theoretical perspectives. This sentiment has been echoed by other authors too. Kauffman (1970), while discussing the application of cybernetics in biology, writes –

> "Since there may be more than one set of sufficient conditions for the adequate description of the behavior, more than one cybernetic model to account for the behavior may be constructed. Such different cybernetic models will not be isomorphic…not only are multiple views about what a system is doing possible, but also any system may be decomposed into parts in indefinitely many ways, and for any such part, it too can be seen as doing indefinitely many things."[22]

It must be noted, though, that what these approaches are advocating for is not that the system is not complex, or complexity is a subjective notion depending solely on the description of the system in particular ways, but that the intrinsic complexity of the

[21] Badii and Politi (1997), p.6.
[22] Kauffman (1970), p.258.

system forces us to have many representations of the system that are many times non-isomorphic.

As I will show in Chapter 5, this is precisely the case with explaining bacterial chemotaxis such that it requires the study of the system from various perspectives to explain its functioning and organization. In the context of current discussion though, I want to make two points –

Firstly, the complexity of biological systems and our limited cognitive abilities provides another justification for explanatory pluralism. This is an argument which is different from the ones presented in the previous section. The motivation for this argument is not an appeal to scientific practices, but the ontological features of biological systems. Since biological systems are complex, we must expect that there ought to be various representations and models possible for it. Therefore, we can argue that explanatory pluralism is not just contingent on our observation that explanatory practices are diverse, but that the explanatory practices are diverse because of the very nature of biological systems – i.e. not only are there many models and representations of a particular biological system, but that given the complexity of biological system, we must expect there to be multiple models and representations.

Secondly, and relatedly, the intrinsic complexity of the system leads to epistemic complexity – i.e. the complexity of studying these systems themselves. Epistemic complexity is the complexity faced by the scientists in their aim to represent, predict, and control a phenomenon or a system that is the focus of study. As Rheinberger (1997) notes, experimental design by biologists is an attempt to reduce the epistemic complexity of the problem at hand.

My view in this regard is similar to that put forward by Mitchell (2003). Mitchell (2003) notes that the complexity of biological systems does force us to adopt multiple representations as suggested by our different models, but it does not force us to commit to an ontological pluralism. Instead, the representations (and models by extension) are partial. These models are constructed with pragmatic considerations in mind and are therefore affected by considerations beyond the want for a complete representation. These considerations can be the desired degree of accuracy and simplicity, the context of the investigation, levels of representations, and cognitive manageability. Therefore, Mitchell suggests two important lessons –

Firstly, the plurality of descriptions does not mean that the descriptions are in competition with each other. Instead, they can be partial descriptions capturing different features of the system. The problem with taking one description as more general or more important than the other might mean missing out on the features captured by the latter.

Secondly, Mitchell argues that these partial descriptions can be integrated to form a fuller picture. Thus, Mitchell rejects 'Isolationist Pluralism' (that different models or representations are complete at different levels) and instead puts forth an 'Integrative Pluralism' that is committed to explanatory integration of the plurality of these descriptions. To me, this integration seems a worthy goal. But this goal requires two different steps – firstly, to dispel the isolationist pluralist, one must show that different models/descriptions do provide different but important perspectives to explain the system's behaviour in question. And secondly, we must show how they can possibly be integrated. In chapter 5, I take the example of bacterial chemotaxis to show how the former is possible. And eventually, I try to tackle the problem of integration and explore

certain ways it is possible. Although, a concrete solution to the problem of integration is beyond the scope of this thesis.

1.4: Epistemic Complexity and Heuristics

Complex systems, as I have argued, provide a problem for representations. At one end, the process of model building is itself rich and diverse with regard to pragmatic considerations; on the other hand, complex systems resist any full representation due to their intrinsic complexity.

When faced with complexity that exceeds our cognitive abilities, problem-solving practices need to be employed to make sense of the overwhelming complexity. This is for two reasons – firstly, the processing and computing powers of humans or even the fastest computers is not enough to go through all the logical possibilities, and second, the search space and criteria for what counts as an explanation for the phenomenon are not well defined. Therefore, scientists need some strategies to limit their search space to make a decision about where to dedicate the limited processing and computational power. This is, of course, a fallible strategy such that the methods employed to limit the search space may not provide the results that are explanatory or increase our understanding of the phenomenon. Simon (1962) notes that the process of problem-solving requires a "great deal of trial and error"[23], but that these trials and errors are not "completely random or blind"[24] but are guided by some strategies. These strategies have been called heuristics [Simon (1962), Bechtel and Richardson (1993), Wimsatt (1972,

[23] Simon (1962), p.472.
[24] Simon (1962), p.472.

2007)]. The way that heuristics work is by reducing the space for possible solutions by systematically making some solutions more likely given some research goals, types of explanations, and available knowledge about the system. Heuristics in this respect are particularly useful when dealing with complex systems where the available parameters and variables might be too overwhelming to work with.

In this respect, Wimsatt (2007) describes the four properties of heuristics –

1. *Fallibility* – using a particular heuristic strategy does not guarantee the correct results or even any results.
2. *Efficiency* – using heuristics is 'cost-effective' in terms of computational power or other resources needed.
3. *Systematically biased* – Errors produced by heuristic strategies are not random but are systematically biased. What is meant by 'systematically biased' is that a particular heuristics strategy will break down in a predictable way such that we can have a fair idea of a class of systems it will break down for. Secondly, the heuristic will make errors in a particular 'direction'.
4. *Transformative* – Use of heuristics leads to a transformation of the problem into a related but non-equivalent problem that might be easier to solve. Moreover, the relation between the initial problem and the transformed problem must be explored and is not a given.

Heuristics are thus useful tools in the hands of scientists dealing with complex problems and are used to reduce the epistemic complexity. In the context of the thesis, there are two heuristic strategies (used in mechanistic models, and models of systems biology) that must be considered. I first consider heuristic strategies in mechanism as

explained by Bechtel and Richardson (1993), following which I consider the heuristic strategies of motif identification in systems biology.

1.4.1: Localization and Decomposition: Heuristics of Mechanism

Bechtel and Richardson (1993) note that one of the most successful heuristic strategies in biology has been that of localization and decomposition. In many fields like molecular biology, cell biology, genetics, biochemistry, etc.; scientists aim to explain a phenomenon by dividing the system into parts such that their interactions and organization can account for the system's behaviour. Bechtel and Richardson (1993) codify these heuristics as the central aspect of mechanisms in biology. Mechanisms, as per Bechtel and Abrahamsen (2005) "is a structure performing a function in virtue of its component parts, component operations and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena"[25]. While this conception of mechanism is but one of many (and I return to this in chapter 2), it will serve the purpose of understanding the heuristics of decomposition and localization for now.

Decomposition, according to Bechtel and Richardson (1993) is a strategy to divide the explanatory task "so that the task becomes manageable and the system intelligible"[26]. To do this, the tasks performed by the system is divided into subtasks, and it is assumed "that one activity of a whole system is the product of a set of subordinate functions performed in the system"[27]. The heuristics of decomposition,

[25] Bechtel and Abrahamsen (2005), p.423
[26] Bechtel and Richardson (1993), p.23.
[27] Bechtel and Richardson (1993), p.23.

thus, allows us to transform the problem of defining the activity/behaviour of the whole system into defining subtasks that together produce the said activity/behaviour.

Localization, on the other hand, "is the identification of the different activities proposed in a task decomposition with the behavior or capacities of specific components"[28]. To understand this, we must understand the strategies of approaching a system from bottom-up and top-down perspectives. The bottom-up approach aims at identifying the components of a system, analyzing how they behave through intervention to understand their contribution to the behaviour to the system. The top-down approach, on the other hand, starts with the system behaviour and tries to understand how it can be explained in terms of subtasks. Localization hypothetically links the subtasks to the components of the system – or it maps the subtask onto a component part. Together decomposition and localization enable a structural and functional mapping between subtasks and subsystems, hence providing a middling approach.

Furthermore, as Bechtel and Richardson (1993) note, localization can be of two varieties – direct and complex. Direct localization "assumes that there are a number of components in the system, that these components function independently, and that any complexity in the behavior of the system is the effect of isolable subsystems"[29]. Therefore, direct localization places the locus of study on the independent components and functions, but it fails to account for the whole activity of the system as it merely locates underlying systems within the complex systems – it does not show how the complex behaviour of the whole is produced –

[28] Bechtel and Richardson (1993), p.24.
[29] Bechtel and Richardson (1993), p.64.

> "Simple localization differentiates tasks performed by a system, localizing each in a structural or functional component. Complex localization requires a decomposition of systemic tasks into subtasks, localizing each of these in a distinct component."[30]

Therefore, complex localization does not merely differentiate tasks performed by the system but yields an understanding of interactions and organization as well. Therefore, the complexity of the system is reduced by considering the behaviour of the whole as a composite of the tasks performed by the components. In this way, the researcher can concentrate on one subsystem at a point and then explain how the interactions between them explain the behaviour of the whole.

As Bechtel and Richardson (1993) note, the crucial assumption of this decomposition approach is that of the idea that the system is 'near-decomposable'. The idea of near-decomposability, as described by Simon (1962) is that a system is near-decomposable if the interactions between parts (or subsystems) are less negligible (or weaker) when compared to interactions inside the parts (or subsystems) when determining the properties of said parts (or subsystems). Simon bases his near-decomposition strategy on naturalistic grounds – that modular organization can be observed in many biological systems. But this assumption might fail, especially in highly interconnected systems. In Chapter 2, I argue that this is might be the case in certain systems. Therefore, the heuristic of decomposition and localization may not apply to all classes of systems in biology. Bechtel and Richardson (1993) recognize this–

[30] Bechtel and Richardson (1993), p.125.

> "[A] wide variety of organizations may be revealed by beginning with an assumption of near decomposability. The resulting models may not retain the integrity of the components, but may describe what we have termed an integrated system. In such a system nature is at best minimally decomposable. If organization becomes even more dominant in explaining the behavior of the system and we appeal to less different and distinctive functions performed by the components, we reach a point where decomposition and localization in any recognizable form have to be surrendered."[31]

Bechtel and Richardson (1993) thus allude to the limitations of the heuristic strategies used in mechanistic research. But as they note, failure of localization still increases our understanding on the system. Failure of localization may indicate that the system is more complex than it was assumed to be, and the near-decomposability of the system might be brought into question. This can force the researcher to look for other ways to analyze the system.

In this way, the heuristics of decomposition and localization form an important strategy for research in molecular biology, and especially in mechanistic models. But, as with every heuristic strategy, it is fallible. In chapter 2, I argue against an ontic conception of mechanisms and draw upon Bechtel and Richardson (1993) conception of epistemic mechanism to argue for an epistemic conception of mechanistic models. The use of heuristics and their limitations will play a role in this argument.

[31] Bechtel and Richardson (1993), p.199.

1.4.2: Motif Identification: Heuristics of Systems Biology

With the explosion of data in biology post the omics project of compiling and cataloguing the genomic, proteomics, metabolomics, etc. sequence of various species, and subsequent data generated by such (ongoing) projects, a new avenue for data-centric biology has been seen in the recent past. While an analysis of the methods, aims, and hopes of the big data in biology is beyond the scope of this dissertation, the data explosion has led to various techniques of data analysis. A prominent approach in this regard has been that of identifying and studying network architecture. There are two approaches – one dealing with large scale network architecture that has given rise to analysis of the whole network for properties like robustness when compared to other architectures of similar size. The other, and more prominent approach has been of searching networks for smaller sub-units that can be studied as functional units. In Chapter 3, I show how both these approaches lead to explanations in biology in the virtue of the study of networks through explanatory strategies of 'topological explanations' and 'minimal models'. For now, I want to analyze the heuristics used in these practices – especially in network motifs approach.

Motif identification, as pioneered by Alon and his collaborators [Mangan and Alon (2003); Mangan, Zaslavar, and Alon (2003); Alon (2007a); Alon (2007b)] is a process of identifying smaller functional units within a larger network. The method employed in this approach is that of statistical significance. The given network is compared to a randomized network of the same size and statistically analysed for units that tend to appear statistically more significantly than random chance.

These motifs are then analysed for their functional features through defining parameters for which a certain function is realized. The utility of the approach is that

we can study biological systems even without full knowledge of the molecular details by showing that a certain motif is instantiated in a particular case. While I show how this can be explanatory in Chapter 3, the point I want to make here is that there are significant assumptions – both explanatory and methodological that form the heuristics of this approach. These assumptions are fallible, and while the research on their limitations has been much less than that of mechanisms, a few observations on these heuristics is warranted.

The first assumption is that the statistical significance of motifs in real biological networks is due to the process of evolution [Gross (2016)]. The idea is that as compared to random networks of the same size, the appearance of the motifs in biological networks is explainable due to the actions of evolutionary forces on biological systems that are absent in the random networks. This assumption has been doubted by Solé and Valverde (2006) who argue that the motifs might not be present because of selection but because of how networks grow. Networks, according to them, "grow and change through the duplication of elements already present"[32]. Mutations, for example, can insert or delete links amongst the elements. They cite a study by Kuo, et al (2006) in which artificial transcription regulatory networks show an abundance of network motifs when random duplication and mutations are introduced, even without specifying a true functionality. The idea that motifs are found abundantly in evolved systems is constructed out of the observation that (a) they exist as compared to randomized networks of similar architecture, and (b) there is a functional role that can be assigned to them. For (a), a randomized network does not necessarily account for how network growth takes place. It is not subjected to network growth rules that the real network in

[32] Solé and Valverde (2006), p.420.

question has been subjected to. It is, after all, randomization of the real network without the history of network growth that has taken place in the real network. Thus, it does not function well as a comparative tool. For (b), as Kuo, et al (2006) show, a network when subjected to events that result in network growth – gene duplication, mutation, etc., can show network motifs despite not being selected for and with no function assigned to them.

The second problem is that with the heuristics used for statistical screening of motifs [Gross (2016)]. As Artzy-Randrup, et al. (2004) show, the construction of random networks and the rules used for them can have significant results for the statistical analysis of the network and therefore the presence or absence of motifs found during the comparison. Different rules will give rise to different results. An example cited by them is that of 'preferential attachment' in which new nodes are preferentially connected to nodes that already have a higher density of connection. In such cases, an abundance of motifs has been found. Therefore, how significant a motif depends on how the randomized network is generated.

My aim in this section is to argue that the use of heuristics is a prominent tool in modelling in science but is fallible. It is, therefore, necessary to understand both – what their explanatory significance is, and their limitations. In chapter 5, I argue by taking the example of chemotaxis in *E. coli* that sometimes the explanatory virtues of different approaches can be overstated. A remedy for this is to understand the very specific way in which models in biology explain which is dependent on their heuristics and accompanying assumptions.

Lastly, heuristics have consequences for model construction and the explanatory virtues and scope of the model. In this regard, it is essential to make a few more points about heuristics explicit –

1. Heuristics reduce epistemic complexity but using them without paying attention to their assumptions – both ontological and methodological – can lead to false understanding of the intrinsic complexity of the system in question. As Wimsatt (2007) notes, all methods of investigation are created to be applied in a certain context. But when used out of context, they can "produce biased or worthless results"[33]. Therefore, an important aspect to the use of heuristics is to understand the context in which they can be used. Therefore, while using heuristics, it is not only essential to make them explicit, but to understand what the methods derived from them explain.
2. The lack of explicit recognition of these assumptions and biases can lead to a view in which the ontological complexity of the system is conflated with the reduced epistemic complexity of our models. Wimsatt (2007), in this regard, considers the heuristics of reduction in biology and how it can lead to a 'reductionist view of the world'. As he states, the transformative nature of heuristics can lead a researcher to believe that the solution of the transformed problem is indeed the solution of the original problem. Even more severely "Taken one step further, the new problem may be put forward as defining the proper formulation of the original [problem], and thus it has replaced the old in a manner rendered largely invisible since it is now regarded as a "clarification

[33] Wimsatt (2007), p.78.

> of the old problem""[34]. A similar caution has been heralded by Levins and Lewontin (1985), who in their seminal book - *The Dialectical Biologist* - warn that the ontological commitments of models (i.e. the entities and structures implied by the theory/model) must not be taken as reflecting the ontology of the world. Instead, the commitments of the theory/models must be made explicit and therefore should be a part of the dialectics about the theory/model [Winther (2006b)].

Therefore, while heuristics provide a strategy that is inevitable in scientific research (due to the need for the reduction of epistemic complexity) they need to be handled with care and analysed for the 'limited' epistemic account of the explanation they can provide. By understanding how specific heuristic strategies construct and explain the system, we can foster a better understanding of the system and its behaviour.

1.5: On Abstraction and Idealization

As I have mentioned already, philosophers have recently paid a lot of attention to the methods of model construction in science. In this respect, two concepts have shown a ubiquitous presence – that of abstraction and idealization. Philosophers have noted for long that models show abstraction and idealization when representing the world.

Abstraction is generally construed as 'leaving out details' in a representation. Godfrey-Smith (2009b) notes "[a]n abstract description of a system leaves a lot out… but it is not intended to say things that are literally false"[35]. Similarly, Jones (2005) notes

[34] Wimsatt (2007), p.87.
[35] Godfrey-Smith (2009b), p.48.

"abstractions involve omission without misrepresentation"[36]. In comparison, idealization is seen as a departure from reality towards misrepresentation. Godfrey-Smith (2009b) notes – "idealization involves a departure from reality in the direction of some kind of simplicity"[37]. Jones (2005) notes "'idealization' applies, first and foremost, to specific respects in which a given representation misrepresents"[38].

Therefore, the primary distinction between abstraction and idealization is that of representing truthfully, albeit with details left out, and misrepresentation, respectively. As per this view, abstractions leave out properties for various reasons. Cartwright (1983), for example, draws a relation between abstraction and generality. As per this view, abstraction leads to the model/theory being more generally applicable - a more abstract model/theory can be applied to various cases since the abstraction can lead to leaving out the differences in the cases. As per the specific case, these omissions can be added back to give a fuller picture of the system/behaviour in question. Recently, Levy (2018) has argued that this relationship between generality and abstraction is not a given, but is dependent on portraying that the underlying relations do not matter – i.e. a more general and abstract model can only be applied to varied cases where we can show that the underlying relations that are omitted during abstraction do not make the difference to the explanation. In Chapter 3, I show that this is the strategy that is used in mathematical models in systems biology while applying the models to various systems such that they can be placed in a particular class of systems.

Another aspect of abstraction as highlighted by Levy (2018) is that abstractness is a comparative idea. The only way we can say that a model/representation is abstract is by

[36] Jones (2005), p.175.
[37] Godfrey-Smith (2009b), p.49.
[38] Jones (2005), p.174.

comparison to other models/representations. Therefore, while describing a representation as abstract, we must presuppose that there is indeed at least another representation of the same system that is more or less 'abstract' in comparison. This idea seems intuitively correct to me. We cannot argue that a representation is abstract *tout court*, especially when talking about fictional or false models; neither can we compare models of different systems for their comparative abstractness. Therefore, while abstraction is a property of the model, abstractness understood as a comparison between models captures the aspect of abstraction that modellers generally allude to when comparing two or more models.

Idealizations, on the other hand, are misrepresentations that are deliberately introduced into the model/representation to make it simpler to understand, provide explanations with. In this respect, idealizations stand in relation between the representation/model and the world rather than between the models/representations [Levy (2018)]. Unlike abstractions, idealizations are deliberate misrepresentations about what the world is actually like and therefore have a condition of falsity or truth attached to their analysis. Most models idealize some aspects of the world in their representations. A model of evolution that assumes the population to be infinite is idealized with regards to the actual population of a species that it can be applied to. Models thus stand in idealizing relationship with the system that then aim to represent or explain. Therefore, idealization, too, cannot be context independent. A model of asexual reproduction, with all the other things equal, is not idealized in the same way if it is employed to represent a population of bacteria as it would be if it were used to represent a population of a sexually reproducing species [Levy (2018)]. Idealization as a falsehood that is deliberately introduced in the representation or model is thus, a result

of the modeller's intentions regarding the utility of the model as a vehicle for understanding and explanation.

This distinction is essential to understand because the way that abstractions and idealizations alter our representations of the world are significant and different from each other. A model can be both abstract and idealized[39]. As a result, we must talk about both these ideas when considering the explanations and representations that our models provide.

In the following chapters of this thesis, the idea of idealization and abstraction (along with the aforementioned heuristics) will be utilized to understand what various models explain, how they explain, and their limitations.

1.6: Conclusion

In this chapter I have introduced some basic concepts that will be used and discussed throughout the thesis. I have provided a basic view of models as mediators and characterised the properties of models that capture the way in which I will be characterising them throughout the thesis. I have also argued for an explanatory pluralism based on pragmatic concepts which I have noted to be derived from the

[39] Although it is debatable if they can be abstract and idealized about the same features of the model. Godfrey-Smith (2009b) and Jones (2005) think they cannot be because abstraction is about leaving out details of a true representation while idealization is introducing falsehood. In such a case, leaving out information about falsehoods introduced by idealization does not make it abstraction of the truth. Levy (2018) on the other hand, argues that representations can be abstract and idealized about the same features. I tend to agree with Levy (2018) on this, because of how I have characterised abstraction and idealization [again following Levy (2018)]. If abstractness is seen as a relation between models, and idealization is the relationship of the model to the actual system that the model aims to represent, then we can have a discourse about abstractness of two idealized models. But I will not further argue for this position because it is not relevant to my thesis. See Levy (2018) for further discussion and explication of this idea.

complexity of the biological system. The pragmatic aspect of explanatory pluralism is thus based on the epistemic problem of constructing whole models and explanations of complex system. Finally, I have characterised heuristics, abstraction, and idealizations as tools used in model construction. All these factors will be shown to have consequences for model construction and explanations offered in ‘organismal’ biology.

CHAPTER TWO

Mechanisms and Mechanistic Explanations

Consider the process of protein synthesis from DNA in a cell. It involves many steps[40] –

1. *Unzipping* – the DNA helix in the nucleus unwinds to reveal nucleotide sequence. This step requires various enzymes, including helicase.
2. *Transcription* – With the help of DNA polymerase, the base sequence of the DNA strand is read, and the complementary base sequence is transcribed upon the mRNA.
3. *Transportation* – the mRNA is transported to the ribosome where it will eventually be translated into protein sequences with the help of tRNA. The tRNA is attached to free amino acids. Each tRNA strand has a particular amino acid attached to it.
4. *Translation* – The process of translation starts when the promoter region of the mRNA is read. The tRNA molecule pairs with a specific base triplet and adds the corresponding amino acid to the protein chain.
5. *Elongation* – The protein chain grows as tRNAs keep adding amino acids to the chain.
6. *Termination* – the translation continues, and amino acids are added to the protein chain until the termination region of the mRNA is read. At this point, the process stops, and the protein chain is finally formed.

[40] This is not supposed to be exhaustive details of the process, just an outline as an example.

This process of protein synthesis has many parts, and these parts take place in various activities. The DNA and helicase take part in the unwinding of DNA helix; the polymerase reads the DNA to form the mRNA; the tRNA reads the mRNA and is used in the formation of the protein chain. These parts and their activities form the process of protein synthesis from the setup condition (the unzipping of DNA) till the termination condition (the reading of the termination region of the mRNA by the tRNA). These entities and their activities thus, make it possible for the protein to be synthesised inside a cell.

Explanatory accounts like these have led philosophers to look into mechanisms that biologists study to explain the phenomenon of interest in question. To explain the phenomenon, biologists look at the constituents that take part in the process and how they interact with each other. As a result of these constituents and interactions, the biologists explain how the phenomenon takes place, thus rendering the phenomenon explained. Philosophers who have been interested in how explanations in biology work, have investigated this process of scientific discovery and tried to give an account of the explanation offered. In recent years, these observations have motivated a new philosophical program of 'New Mechanism'. The proponents of this program hold the position that explanations in biology[41] are primarily engaged in explaining the mechanism underlying the phenomenon in question.

In this chapter, I first give an account of mechanistic philosophy, especially the New Mechanism and its evolution and how they explain the biological phenomenon using strategies of localisation and decomposition. I focus majorly on Craver's account of mechanistic explanations, especially on his ontic account of mechanistic explanations.

[41] And many other special sciences – like sociology, chemistry, etc.

I argue that his position is untenable because of his insistence on mechanistic explanations (and models) alone being ontically adequate. Finally, I argue that while ontic and epistemic constraints are both important for mechanistic explanations, the former need to be relaxed significantly to accommodate various models that may have utility for mechanistic explanations. The motivation for this argument rests on the idea that mechanistic explanations must be considered as a part of a broader context of modelling and explanatory exercise in biology.

2.1: Mechanisms in Science

Mechanistic Philosophy has gained widespread attention in the philosophy of biology literature over the past decade and a half. But this tradition has older philosophical roots. Scientists generally describe the workings of phenomena (both biological and otherwise) with an associated mechanism for it – from mechanism of natural selection, mechanism for protein synthesis, how the engine of a car works, to water cycle. In this regard, mechanisms have an overarching meaning with significantly varied connotations – covering activities of a single cell, a whole ecosystem, to the whole universe. Thus, while discussing mechanisms, it is imperative to differentiate and demarcate what exactly is meant by the term. Nicholson (2012) traces the diversity and history of the term mechanism as is used in biology and characterises it into three broad categories –

1. Mechanicism
2. Machine mechanism
3. Causal Mechanism

Mechanicism, as Nicholson (2012) notes, is a philosophical thesis about the nature of living beings committed to the idea of "ontological continuity between the living and the nonliving"[42]. It is essentially atomistic and reductive in its outlook and maintains that the biological systems (wholes) are directly determined by the properties and activities of their parts.

Machine mechanism, on the other hand, is a more neutral term generally used to describe the workings of a machine. Nicholson (2012) notes that the usage of this term in the biological domain has mainly semantic connotations, where biologists describe the features and workings of a particular system by analogies to machines. As a result, machine mechanisms can be seen as a useful heuristic tool in biological explanations.

Lastly, *causal mechanisms* are engaged in the study of causal relationships that bring about a phenomenon. Though often conflated with machine mechanisms, there is a difference between the two. Causal mechanisms are concerned with exploring the workings of a phenomenon by making the underlying causal structure intelligible, and thus, rendering the phenomenon explained. They are not purely heuristic like machine mechanisms as they play a role in explicating the make-up of the world by exploring cause and effects, thus having an ontological significance.

The New Mechanism[43] philosophy is concerned with the third kind of mechanisms, i.e. causal mechanisms. The aim of the program is to extrapolate from scientific practices as to what works as a good explanation in biology[44]. Thus, the proponents of New Mechanism generally advocate looking at the way scientists explain a specific

[42] Nicholson (2012), p.153
[43] Generally called New Mechanicism, but I want to avoid confusion with the Mechanicism as noted above.
[44] And other special sciences.

phenomenon as a token case of a good explanation, and they aim to elucidate, using such token cases, the type of explanations that work in biology. Another way to put it may be that instead of looking at what a mechanistic explanation (a particular causal mechanism) is explaining, the interest of the New Mechanism is in addressing – how and why is the causal mechanism explanatory. Thus, they aim to capture the meaning of explanation in biological (and other) sciences that generally employ a mechanistic explanation and to draw a causal framework for the mechanistic explanations in biology.

We can thus safely forgo the ambiguity of the term and talk about the New Mechanism philosophy in a causal mechanism term, avoiding any reference to the mechanicism or machine-mechanism terminology (unless explicitly mentioned).

The discussion in this chapter will be about two theses – first, about mechanistic explanations in biology, their scope, their limitations, and their structure; and second, about mechanisms as causal structures. The broad aim is to understand how the mechanistic philosophy aims to understand mechanistic explanations and how they might help understand the causal structures that are out there in the world. Thus, mechanistic explanations have two aims

a. Epistemic – that of how mechanistic models represent the target system, and the other
b. Ontic – that of how mechanistic explanations map onto the real causal structures in the world.

Keeping this in mind, I will be using the term mechanistic explanations/mechanistic models to denote the former – that of representing a system in a certain way as per the

methods of mechanistic models. And mechanisms to denote what ontically minded mechanicists claim are the real causal structures out there in the world.

2.1.2: New Mechanism

Many authors have defined mechanisms in an attempt to explain the necessary and sufficient conditions for a mechanistic account. The most prominent definitions come from the following –

> **Machamer, Darden, and Craver (MDC, henceforth):** "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions."[45]
>
> **Glennan:** "A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interaction between parts can be characterized by direct, invariant, change-relating generalizations."[46]
>
> **Bechtel and Abrahamsen:** "A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena."[47]

All three sets of authors define mechanisms in terms of their parts/entities and activities/interactions/operations organised in such a way that they are capable of bringing about the observed phenomenon. Furthermore, all share some commonalities

[45] Machamer, Darden, and Craver (2000), p.3
[46] Glennan (2002), S344
[47] Bechtel and Abrahamsen (2005), p.423

regarding what it means to be a mechanism – a mechanism displays productive continuity that is dependent on parts, their properties, their activities, and their organisation. This productive continuity is responsible for the function of the mechanism as a whole – which is the phenomenon in question. But they also differ in significant ways – especially the goals of what a mechanistic explanation is.

Glennan, for example, is concerned with explaining the causal relations. For him, mechanisms answer the question of how two events are causally related to each other. In his 1996 paper, Glennan espouses mechanism as the answer to Hume's problem of induction[48] by arguing for a theory of causation based on mechanisms. Mechanisms, thus, offer a way to expose the 'secret connexions' [Hume's term] between the cause and the effect. In his 1996 paper, Glennan maintains that the interactions of the parts are characterised by 'direct causal laws'[49]. The significance of this, according to Glennan (although he does not go into much discussion), is that the action of one part on the other must be immediate (direct) – for example if in a machine cog 1 moves cog 2, and cog 2 moves cog 3, then cog 1 is not directly causal to cog 3 but only to cog 2. This direct action is a way to flesh out the connections to exact details.

On the other hand, 'causal' is used to evoke the difference between actual causes and mere correlation – only the first of the two can qualify for mechanistic explanations. As for laws, Glennan indicates that laws, as he uses it, are change-related generalizations. In his 2002 paper though, he changes the 'direct causal law' language for 'direct, invariant, change-related generalization'[50]. In both the papers though, Glennan's

[48] Hume argues that our experience is insufficient to establish any inferences about causation. All we can observe is conjunction and not explicate the 'secret connexions' between the cause and effect. Hume rather adopts a sceptical view of causation and defines a cause as a constant conjecture.
[49] Glennan (1996) p52.
[50] Glennan (2002), S344.

mechanism is not merely explanatory of biological phenomena, but a metaphysical theory of causation in biology (and other special sciences).

MDC (2000), on the other hand, have different goals in mind. Their concept of mechanism is explanatory. Their accounts offer a way to understand how mechanisms explain a phenomenon and how we can establish what is relevant or irrelevant to the explanation. Bechtel and Abrahemsen (2005) also offer a similar view of mechanisms. MDC (2000) explain the virtues of mechanism in their ontic, descriptive and epistemic adequacy. Ontically, MDC (2000) substantiate mechanisms as appealing to both substantivists and process ontologists. Since they require both entities and activities, they believe that individually substantivism with the appeal to entities and their properties, and process ontologists with the appeal to processes and activities, are ontically inadequate to capture how mechanisms show productive continuity. Herein lies a significant difference between Glennan and MDC – for MDC interactions are not sufficient for mechanistic explanations. Interactions imply two or more entities engaged in a change. This is much clearer in Glennan's 2002 paper than in his original 1996 paper, where he drops his insistence on direct causal law and adopts the language of change-related generalization. For Glannan, mechanisms are causal because a change in the properties of one part brings about the change in the properties of another part[51].

Furthermore, this relationship for Glennan will be invariant and hence generalizable[52]. Activities, on the other hand, vis-à-vis MDC (2000) are not necessarily changes induced in one entity by the other, but self-sufficient in their existence. For example, a protein folding to a specific configuration is an activity. It cannot be called

[51] Glennan (2002), S344

[52] Where generalizable means that the relationship holds under a range of varying conditions.

an interaction as it is dependent on one entity – the protein, and its activity – that of folding. As such, MDC do not provide any coherent definition of activities, but it is easy to see why they think that interactions are not sufficient to describe mechanisms[53]. The problem that arises from the reliance on activity for MDC is that activities alone are challenging to establish productive continuity. If an entity A is engaged in an activity and we do not establish how it brings about a change in entity B, then we cannot explain the productive continuity of the mechanism. As Woodward (2002) points out – the only way to establish a productive relationship is by generalization that is invariant under interventions. This creates a problem for MDC because activities are local and necessarily do not support counterfactuals. Woodward's suggestion is thus to shift from MDC's activity-based account to a counterfactual account rooted in interactions that can be generalized.

> "(MECH) a necessary condition for a representation to be an acceptable model of a mechanism is that the representation (i) describe an organized or structured set of parts or components, where (ii) the behavior of each component is described by a generalization that is invariant under interventions, and where (iii) the generalizations governing each component are also independently changeable, and where (iv) the representation allows us to see how, in virtue of (i), (ii) and (iii), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves."[54]

[53] The motivation of MDC is both ontological and epistemic. Ontologically, they are committed to dualism – that of substance and processes. Epistemically, they argue that biologists often use activities as an epistemic tool to elucidate mechanisms. Machamer (2004) explicitly states that defining activities is difficult but that we can know of their existence through experience.

[54] Woodward (2002), S375

For Woodward, it seems that although mechanistic philosophy does provide a descriptive framework for causal features that bring about the phenomenon, it lacks explanatory rigour. Woodward, in his book *'Making things happen'* (2003) provides an interventionist approach to counterfactual reasoning with the aim of understanding causality in sciences. The main aim of Woodward's interventionist approach is to map out a range of invariant relationships among entities in questions as generalizations. For Woodward, mechanisms, as described by MDC, could not achieve generalizations because their focus was on local activities and ill-defined productive continuity which made a manipulationist account difficult as in no sense could ideal interventions be defined. For Woodward, ideal interventions were necessary to be able to establish the causal relation. On the other hand, Woodward finds Glennan's reliance on causal laws to be inadequate. Hence, he sees that only by introducing an interventionist approach to the mechanistic program might they be useful as an explanatory tool. To do this, Woodward introduces the idea of ideal interventions on variables in question.

2.1.3: Woodward's Interventionism

Woodward (2000, 2003) explicates the idea of interventions as invariant, change-related generalizations as a part of his counterfactual interventionist theory –

> "Suppose that *I* is an intervention on (or manipulation of) the variable X, where X is some property possessed by the unit *i*, the intent being to assess some postulated relationship (G) according to which changes in *X* cause or explain changes in some other variable *Y* by observing whether the intervention on *X* produces the change in *Y* predicted by (G). Call the value of *X* possessed by *i* prior

to the intervention x_0 and the value after the intervention x_1. Then *I* should have the following con-junction of features (M):

M1) *I* changes the value of *X* possessed by *i* from what it would have been in the absence of intervention (i.e. $x_1 - x_0$) and this change in *X* is entirely due to *I*.

M2) The change in *X* produced by *I* is claimed by (G) to change the value of *Y*. That is, according to (G), the value, y_0, that *Y* takes when $X = x_0$, is different from the value, y_1, that *Y* takes when $X = x_1$.

M3) *I* changes *Y*, if at all, only through *X* and not directly or through some other route. That is, *I* does not directly cause *Y* and does not change any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I*-*X*-*Y* connection itself; that is, except for (a) any causes of *Y* that are effects of *X* (i.e. variables that are causally between *X* and *Y*) and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*. In addition, *I* does not change the causal relationships between *Y* and its other causes beside *X*. Moreover, a similar point holds for any cause *Z* of *I* itself—i.e. *Z* must change *Y*, if at all, only through *X* and not through some other route.

M4) *I* is not correlated with other causes of *Y* besides *X* (either via a common cause of *I* and *Y* or for some other reason) except for these falling under (M3a) and (M3b) above."[55]

Woodward's account of invariance under intervention postulates that an ideal intervention (which follows M1 through M4) on a variable X which is a cause of variable Y can reveal generalizations about the relationship between X and Y. The ideal

[55] Woodward (2000), p201.

intervention on X must be able to bring about some change in Y through the change in X and X alone. As Craver (2007) puts it in simpler terms – the ideal intervention (I) must –

1. not change Y directly,
2. not change the value of some causal intermediate Z between X and Y except by changing the value of X,
3. not be correlated with some other variable M that is a cause of Y,
4. act as a 'switch' that controls the value of X irrespective of X's other causes.[56]

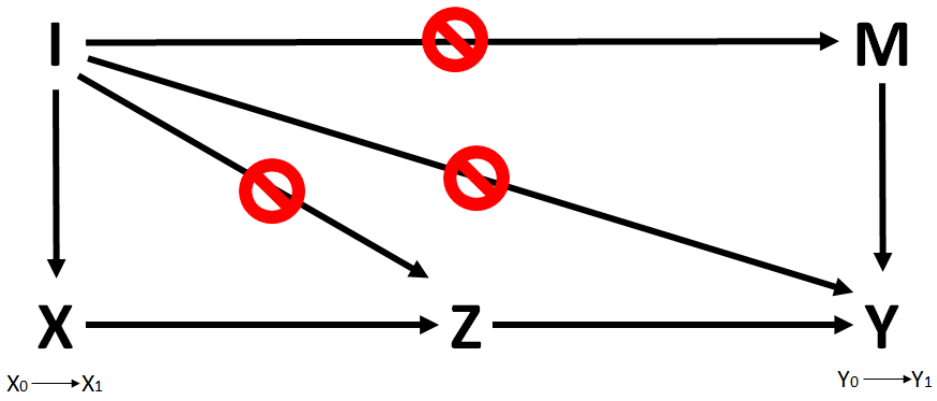

Figure 2.1 – A diagrammatic explanation of Woodward's ideal intervention. Where I is an ideal intervention on variable X such that it changes the value of X from X_0 to X_1 to change the value of variable Y from Y_0 to Y_1 correspondingly, with the conditions such that I does not change the value of Y directly, that it does not change the value of Y through an intervention on an intermediatory Z between X and Y, and that it does not change the value of M that is an independent cause of Y. In such a case, the relationship between X and Y can be said to be invariant under intervention.

If these conditions are met, then the relationship between X and Y is invariant under intervention and can be said to be generalizable. Woodward aims to provide a theory of

[56] Craver (2007), p.96.

causation for sciences that do not have laws (in the D-N sense). His theory is a part of broader manipulationist theories of causation that explain causation as the ability to manipulate the system to be able to study them. For Woodward, hence, the ability to manipulate and intervene on a system to study relationships between variables that can be generalized is a way of exploring causal relations in the absence of fundamental laws of causation.

The implications of Woodward's (2002) observations create a problem for MDC's and Glennan's accounts. For Glennan, the criticism centred on his reliance on laws. As noted, Glannan (2002) changes his earlier definition from his 1996 paper to accommodate Woodward's criticism to include change-related generalizations instead of direct causal laws as the causal force behind his mechanistic explanations. MDC, on the other hand, do agree with Woodward to some extent but maintain that activities are of ontological significance to mechanisms. Machamer (2004) argues that activities are what connects entities and are epistemically and ontologically significant to our understanding of mechanisms. A further, and in-depth, inquiry can be found in Tabery (2004). Tabery notes that both the terms – interactions and activities bring something to the table. While he acknowledged that MDC's idea of activities is a bit vague, he believes that their dualist stance about entities and activities is essential to understand the dynamic nature of mechanisms. As he puts it *"[f]or the dualists, the activity is the dynamic process of bringing about*"[57]. Activities according to Tabery (2004) are a way of opening the black box of how property changes are brought about – i.e. not just how interactions induce changes but what the process of this change exactly is. According to Tabery, both these concepts are not sufficient to capture the complexity of

[57] Tabery (2004), p10. (italics in original).

mechanisms alone. He thus, suggests a synthesis of the two where interactions and activities are closely related and need each other for a proper description of the mechanism. He calls it 'interactivity'. The synthesis that Tabery proposes is to explain both the productive continuity and dynamic nature of the mechanism.

Craver, on the other hand, incorporated Woodward's suggestions into his view of the mechanism in a different way.

2.1.4: Cravarian Mechanisms

One of the aims of the New Mechanism program (specifically MDC and Bechtel) has been to define what entities and activities form a component of the mechanism, i.e. how do we ascertain if a given component is relevant to the phenomenon that a given mechanism strives to explain. Mechanisms are complex systems and many a time they are situated within a larger system. Take protein synthesis as an example again. The whole phenomenon is situated within a cell which has many components and performs many functions. So, the question arises – what components of the cell can be said to be a part of the phenomenon of protein synthesis? This issue of constitutive relevance is a part of the larger question of mereological dependence – i.e. how wholes are realised by their parts. In biological organisations, which are generally accepted to be hierarchically organised[58], this question is important for any account of explanation to be able to answer. Since the phenomenon to be explained by a mechanism is that of a higher level

[58] By hierarchically organised it is meant that there are several levels of organisation composed of parts at a lower level of organisation. Biological organisations are generally thought to be nested, such that the higher levels of organisation are not just composed of lower level of organisation but are caused by the lower level.

realisation of the lower-level entities and activities, it is imperative that there be a clear demarcation process as to which entities/activities are responsible for bringing about the said phenomenon. Craver (2007) takes this issue head-on. For Craver, the interventionist framework provided by Woodward (2002, 2003) is a possible solution to address this question of constitutive relevance.

First, he distinguishes between causal and constitutive relationship. While the causal relationships are held between entities on the same level, constitutive relationship exists between the whole and its parts. For example, in a cell – the DNA, RNA, cell membrane, mitochondria, etc. exist on the same level, such that they causally interact with each other to perform functions. These components, therefore, are engaged in causal relations which can be studied by Woodward's interventionist approach. On the other hand, these components together form the cell, hence have a part-whole relationship with the cell. They are not causal for the cell as these components, and their causal relations constitute the cell itself. Hence, they have a constitutive relationship with the cell – that of mereological composition and dependence, i.e. they together form the cell and that their fate is intrinsically linked with the existence of the cell. Craver and Bechtel (2007) emphasise this while discussing top-down causation[59].

Craver and Bechtel (2007) argue that mechanistically mediated effects, which are composed of both causal and constitutive relationships, can solve the problem of top-down causation without becoming epiphenomenal or overdetermined. According to them, instances of interlevel causation can be substituted with mechanistically mediated effects that allow for constitutive relationships to affect the lower level rather than be

[59] Top-down causation is the idea that the whole, composed of the parts, can have causal effects on their parts. It is controversial since if the parts are causal for the whole, how can the whole, which exists because of the parts, have any causal influence on it?

causal, i.e. if any cause is seen as acting upon the higher level then it translates to the lower level being affected, not because of the causal influence of the higher level, but because of the intra-level causal relationships at the lower level being altered due to the inter-level constitutive relationship. This entails that explaining a phenomenon on different levels necessarily means explicating mechanisms and the causal and constitutive relationships that are at play. Craver (2007) builds further on this idea, both to explicate the nature of mechanistically mediated effects and to establish the constitutive relevance of what forms parts of the mechanism in question–

> "[A] component is relevant to the behaviour of a mechanism as a whole when one can wiggle the behavior of the whole by wiggling the behavior of the component *and* one can wiggle the behavior of the component by wiggling the behavior as a whole. The two are related as part to whole and they are *mutually manipulable*."[60]

Here, Craver is suggesting that top-down and bottom-up interventions can help explicate the relationship between parts and wholes. More formally –

> "(i) *X* is part of *S*; (ii) in the conditions relevant to the request for explanation there is some change to *X*'s φ-ing that changes *S*'s ψ-ing; and (iii) in the conditions relevant to the request for explanation there is some change to *S*'s ψ-ing that changes *X*'s φ-ing."[61]

Craver further explicates (ii) and (iii) as –

> "(CR1) When φ is set to the value φ_1 in an ideal intervention, then ψ takes the value f(φ_1)"[62]

[60] Craver (2007), p153, original emphasis.
[61] Craver (2007), p153.
[62] Craver (2007), p155.

"(CR2) If ψ is set to the value ψ_1 in an ideal intervention, then φ takes the value $f(\psi_1)$"[63]

Taken together, (CR1) and (CR2) explain the idea of mutual manipulation. Ideal interlevel interventions can be used to determine constitutive relevance if through the intervention when the value of S is changed, the value of the component changes to a function of this new value. This intervention on S to change the value of the components is a top-down intervention. Similarly, in a bottom-up intervention, if the value of a component is changed, the value of S changes as a function of this new value.

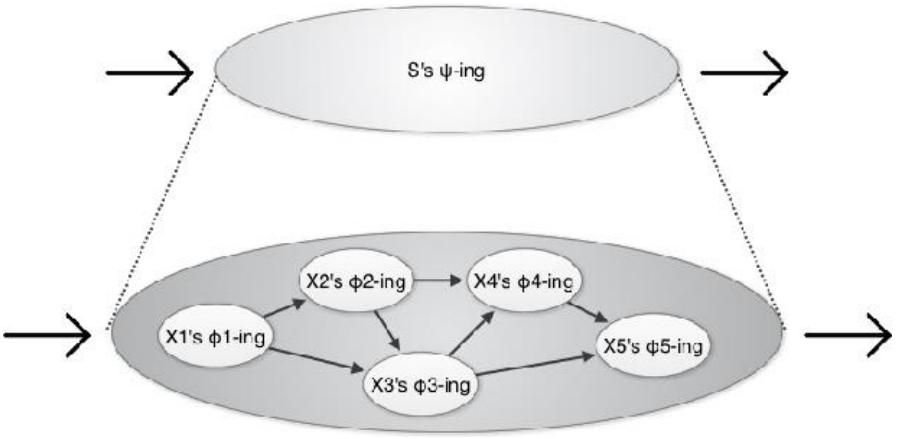

Figure 2.2 – Mutual Manipulation [Craver (2007), p7]. The X_n can be said to be components of S if they satisfy the mutual manipulation condition, where a change in X_n's φ_n-ing changes S's ψ-ing and a change in S's ψ-ing changes X_n's φn-ing.

In what follows, I elucidate some problems for Craver's account and show how Craver solves some of them. Then I analyse Craver's motivation for an ontic account for mechanistic explanations.

[63] Craver (2007), p159.

2.1.5: Craver's Mutual Manipulation and its Problems

Craver's mutual manipulation account is an attempt to demarcate the boundaries of a mechanism by defining constitutive relevance and to understand what entities and activities must feature in a mechanistic explanation. It is based on Woodward's manipulationist account of causation. The difference between the two is that Woodward's account is about causal dependencies, while Craver's account is about interlevel constitution. But the question remains – how well are both these accounts applicable to biological systems? We know that biological systems, unlike simple physical systems, show degeneracy and redundancy – i.e. a system can follow multiple paths, or have multiple copies of the same component that can compensate for the loss of a particular component. Similarly, they show properties like reorganisation – i.e. the system reorganises itself to compensate for perturbations. These properties have evolved to help a biological system to resist external changes and perturbations and continue their operations despite some deviation from the normal[64]. So, the question remains – unlike the physical systems, can the manipulationist account be of use in biology?

The answer is not as straightforward. Woodward (2002), for instance, starts with the example of a block sliding down an inclined plane, with relationships between components that can be represented with equations. Generalization in such a system is relatively easy. It is also easy to demarcate and define the parts. But as noted, biological systems are more complex than a simple physical system. Many times, interference with a component does not yield any change in the whole phenomenon. In such a case, not

[64] See Chapter 4.

only is it difficult to reach any generalization but also to demarcate the boundaries of the mechanism or to define the components that are relevant to the mechanism.

It will help us to take an example to explain this. In the early stages of the embryo, there are different concentrations of various chemicals. According to these concentrations, the scientists decide some regions which have different functions for what happens to the embryo. The Neuwikoop centre is identified by some combination of these chemicals.

The Neuwikoops centre appears in cells of early embryos and is found especially clearly in the embryos of toads [McManus (2012)]. It induces the primary organiser, which in turn determines whether the cell becomes an eye cell, or a skin cell, etc. So, the Neuwikoops centre determines the fate of the cell. However, the Neuwikoops centre is defined by the gradients of certain chemicals within the cell. It is near the extreme end of those gradients, which is typically located where the dorsal axis intersects with the back of the cell.

More precisely, the Nieuwkoop centre in Xenopus axis formation is defined as the dorsal most region of the blastula. It is a region formed as a result of an overlapping expression of VegT, Vg1, Nodal-related protein (Xnr), and β-catenin (βcat). It is a causal part in the induction of the Primary Organiser. As Schoenwolf, et al (2009) note in Larsen's Human Embryology, the Nieuwkoop centre is not structurally distinct. Its location is not determined by any structural features, but merely by the difference in expressions of the aforementioned proteins and therefore the location of the Nieuwkoop centre itself can also be identified by gene expression patterns.

The importance of the Nieuwkoop centre can be seen in grafting experiments performed by Gimlich and Gerhart (1984). They removed the Nieuwkoop centre from one embryo and graft it onto another embryo which already had its own Nieuwkoop centre intact. As a result, two, instead of one, embryonic axis were formed. Hence, they showed that Nieuwkoop centre was causal for the formation of the embryonic axis and that when more than one Nieuwkoop centre is present, a corresponding number of axis are formed.

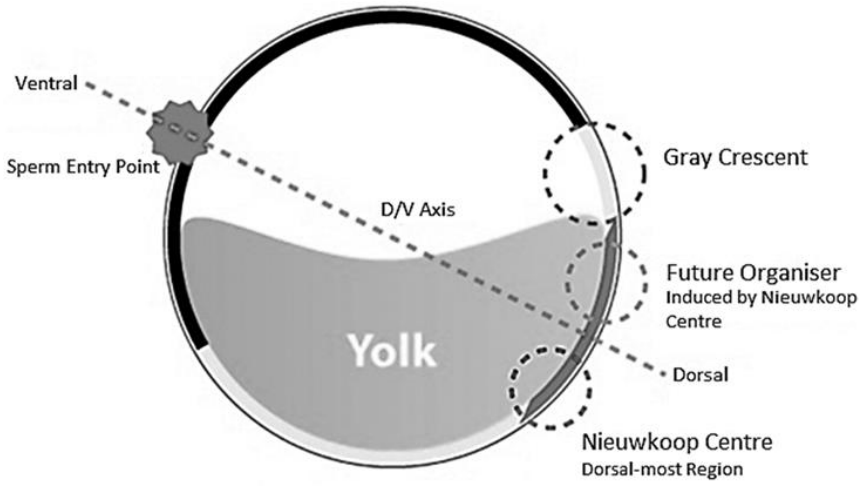

Figure 2.3 – The embryo of Toad showing Nieuwkoop centre and the Dorsal-Ventricle axis. (From https://studentreader.com/F31LK/nieuwkoop-center/)

On the other hand, if the Nieuwkoop centre is removed from an embryo, it has been found that the process of axis formation still takes place and the cell still specialises as before. This is because even after the removal of the Nieuwkoop centre, there is a maxima of the gradients. So there remains a Niewkoop centre – albeit constituted by different molecules and amounts of the relevant proteins. This enables the cell to compensate for the removal of the Nieuwkoop centre and induce the axis anyway.

However, the overall size of the embryo, which subsequently develops, may be affected by the removal of the mass. Thus, the embryo is a highly plastic entity, capable of overcoming deviations from the norm during its developmental cycle.

Biologists treat the Nieuwkoop centre as a part of the mechanism of axis formation because it induces of the phenomenon of axis formation as can be deduced from the grafting experiments. On the other hand, the Nieuwkoop removal experiment shows that the embryo still shows axis formation when the Nieuwkoop centre is not present. The axis still forms in the absence of Nieuwkeep centre because the Nieuwkoop centre is not rigidly defined. The removal still leaves us with a Nieuwkoop centre, albeit of a different concentration. Hence, the counterfactual condition 'if the Nieuwkoop centre was not present, the axis will not be formed', fails when we are referring to the molecules that form the Nieuwkoop centre but is true when we are referring to the Nieuwkoop centre as the peak of the gradient at the dorsal end. For comparison, it does not matter what the mountain top is, as long as there is a mountain top.

Thus, the Nieuwkoop centre is a part of a mechanism, but is unlike a cog in the system in various respects –

1. It is not spatially demarcated. A mechanism's part is spatially demarcated when there is a region of space which forms the boundary between the part and other parts. This region may be vague, but it cannot constitute the whole of the part since that would no longer be a boundary. Examples of things which are not spatially demarcated outside of biology include the red part of the rainbow. By contrast, a mountain top may have a vague boundary, but it nonetheless is bound and so is spatially demarcated. The Nieuwkoop centre is far from the only part of a biological mechanism which is not spatially demarcated in this

sense. Another example is the morphogenic gradient during embryonic development that determines the location of the cells in the developmental trajectory.

2. It is defined by its relative location in the embryo. Many things are defined by their relative location in the non-biological domain. A mountain top is relatively defined with regards to the mountain itself. A table, on the other hand, is not relatively defined to anything. But in biology, many things are relatively defined. The nucleoid in a prokaryotic cell is defined relative to the location of the DNA and ectoplasmic gradient.
3. It is materially robust as the matter it is composed of can be removed.

These features of Nieuwkoop centre create a problem for the mechanistic account as presented by Craver. The Nieuwkoop centre is causal for axis formation – the grafting experiment confirms that. But at the same time, the removal experiments show a different picture – i.e. even with the Nieuwkoop centre removed, the axis formation process still takes place. Thus, in one case, the Nieuwkoop centre is causally relevant for the axis formation process. Still, in the other, it is not – i.e. if one goes only by the grafting experiment, the Nieuwkoop centre is a part of the phenomenon of dorsal-ventral axis formation mechanism. But if one goes by the removal experiment, it is not. This shows that in at least some cases in biology, the mutual manipulation criteria fail to determine the components of a mechanism since the ideal intervention needed for the mutual manipulation account might not be present.

Craver is aware of this problem. Therefore, he introduces two sub-conditions to his (CR1) criterion –

> "[T]here should be an ideal intervention on X's φ-ing that changes the value of S's ψ-ing under the conditions (CR1a) that the intervention, I, leaves all of the other dependency relations in S's ψ-ing unchanged and (CR1b) that other interventions have removed the contributions of other redundant components."[65]

Craver notes that (CR1) might not be satisfied by the system because there might be recovery, reorganisation and redundancy. Recovery and reorganisation taken together, form the robustness of a system – the property of a system to resist and minimise the effects of perturbations – both internal and external. Thus, if an intervention results in the system reorganising itself and recovering from the intervention, then (CR1) will not be satisfied. This violation of (CR1) can happen through various means – like changing the relationship between the components of the mechanism. Thus, Craver adds (CR1a), which blocks these conditions. Redundancy is the property of a system through which it might be able to function even when a component is lost – a feature seen in many biological systems where many components can serve the same function. Different proteins can bind to the same location, for example. If the system has redundant components, then even on intervention on one component may not be sufficient to change the systems' behaviour as an identical component with similar functions can replace the component intervened upon. In both these cases, (CR1) fails. Craver (2007) is aware of this problem and thus he adds (CR1a) and (CR1b) to his conditions for the satisfaction of mutual manipulation criteria.

But this makes his conditions too stringent. It severely limits the types of systems that it can be applied to. In our example of the Nieuwkoop centre, the removal experiment showed that the system has the property of redundancy. So, the question

[65] Craver (2007), p156–7

arises – is the mutual manipulation criteria applicable to the system as an analysis of the mechanism of axis formation? If we abide by (CR1a) and (CR1b), that might not be true. Craver himself notes that (CR1), with its sub-conditions, and (CR2) might not be sufficient nor necessary to demarcate the components of a mechanism, but that they are a good epistemic tool used by scientists to probe into mechanisms. Craver's motivation is to explicate how mechanistic explanations work. He does not necessarily need them to abide by his conditions of mutual manipulation. This condition is, for Craver, at best an epistemic tool to discover mechanisms that exist out in the world.

Craver holds an ontic conception of mechanism rather than purely epistemic. The distinction between the ontic and the epistemic account of explanation, as proposed by Salmon (1984), holds that the ontic account is about explaining the ontic structure of the world rather than representing it. That is, to hold the ontic concept is to try to explain how events fit into the causal picture of the world. Salmon (1984) writes –

> "To understand the world and what goes on in it, we must expose its inner workings. To the extent that causal mechanisms operate, they explain how the world works… A detailed knowledge of the mechanisms may not be required for successful prediction; it is indispensable to the attainment of genuine scientific understanding."[66]

For the proponents of ontic conception, explanations explain causal facts that are out there in the world. As a contrast, the epistemic account of explanation is about representing the world in a certain way. The ontic account can still use representations,

[66] Salmon (1984), p133

but they are mere pragmatic tools not the aim of the explanation itself. As Craver (2007) notes –

> "explanations are not texts; they are full-bodied things."[67]

Craver (2014) postulates the relationship between mechanistic explanations and nature of explanation –

> "Good mechanistic explanatory models… are good in part because they correctly represent objective explanations."[68]

Craver (2007, 2014) is committed to an ontic realist account of explanation. But his framing of mechanistic explanations as good ontic explanations does not necessarily hold to scrutiny. If he wants to reconcile his mutual manipulation account with his ontic conception, he must find a way to make them compatible such that there is a clear way of moving from mechanistic models to mechanisms out in the world while explaining how examples like the Nieuwkoop centre might be incorporated.

In this section, I have analysed what mechanisms are, what mechanistic explanations entail, and how new mechanists provide a theoretical framework for their arguments. I have also argued that there are cases in biology where mechanistic explanations might fail. In what follows, I elucidate Craver's ontic conception of mechanisms, how it can be interpreted as either claim about mechanisms or mechanistic explanations, and some problems with these accounts.

[67] Craver (2007), p27.
[68] Craver (2014), p27.

2.2: The Ontic Account of Mechanism

The Craverian mechanistic account is committed to ontic realism. According to this account, mechanisms are explanatory because they explain the causal structure of the world as it is. It would be pertinent to clarify two points here –

1. Craver is not claiming that all cases of mechanistic explanations are ontically adequate, but that good mechanistic explanations are.
2. Craver is not claiming that other models of explanation do not represent or describe the phenomenon in question, but that they do not explain the objective world.

Craver is imposing ontic constraints on what a good explanation is. For him, a good explanation is one that shows how the mechanism fits in the causal structure of the world through the phenomenon they describe. Craver's argument is as follows –

(P1) Several types of models can be drawn for any given phenomenon – they can be phenomenal (purely descriptive), predictive, or sketches that elucidate the causal structure of the system.

(P2) But the aim of a model is neither to describe nor just to predict, but to account for the ontic structure of the world.

(C) Mechanistic explanations, because of their structure – that of explaining a phenomenon as it is produced by its components – are ontically adequate because they are able to account for the ontic structure of the world.

There are two interpretations of the Ontic Conception that Craver alludes to –

(OC1) The explanation is an objective part of the world, and causal explanations are out there in the world. This interpretation not only claims that mechanistic explanations are good models of explanation, but they are explanations of the phenomenon as it exists in the world. As per this interpretation, mechanisms are not representations but causes. Craver notes –

> "…the term explanation refers to an objective portion of the causal structure of the world, to the set of factors that bring about or sustain a phenomenon (call them objective explanations)... Objective explanations are not texts; they are full-bodied things. They are facts, not representations."[69]

In this interpretation, the aim of giving a mechanistic explanation is to allude to the actual causes or the mechanisms. Mechanisms are causes themselves, not representations.

(OC1) is not only concerned with giving explanations that scientists can use to explain the phenomenon, but to situate it in the causal fabric of the world. There are two claims that can be associated with this thesis –

1. *Identity claim* – that explanations are identical to real mechanisms that exists in the world. That explanations are out there in the world.
2. *Non-representational* – that explanations are not representational. Representations are by their definition non identical. Representations (mental, diagrammatic, models, etc.) stand in some sort of morphic relationship with their target system such that there is an asymmetry between the representation and the represented. The represented (target) is represented by the representation

[69] Craver (2007), p27.

(model, text, diagram) but not vice versa. Due to the identity claim about explanations being identical to the real mechanisms, the explanations and mechanisms have a symmetrical relationship and therefore non-representational.

Claim (2) is a direct consequence of claim (1). Therefore, to maintain the (OC1) conception is to claim than mechanisms themselves are explanations and a good explanation must lay bare the mechanisms as they exist out there in the world.

The deniers of (OC1) conception of mechanistic explanation might point out that the practices of science do not show that this needs to be true. Biologists routinely provide non-mechanistic, non-causal explanations – they can ask why questions about a phenomenon to which eluding to the mechanism is not a sufficient answer. A standard example of this is the explanation of why when a bunch of sticks are thrown up in the air with sufficient spin, will generally show more sticks in the horizontal orientation as compared to vertical if we freeze the frame at any random moment during their fall. The explanation to this is that there are more horizontal orientations in which the sticks can be in as compared to vertical orientations. The explanation appeals to probability distribution of the system rather than giving a causal explanation. Closer to home, in biology, many authors have pointed that non-mechanistic explanations do exist, and they are explanatory. These explanations consist of showing structural aspects [Huneman (2018)] of the phenomenon, or the design aspects [Wouters (2007)]. Proponents of (OC1) might reply that while non-causal, non-mechanistic 'explanations' foster a better 'understanding' of the phenomenon, they are nonetheless not explanatory[70]. Following this line of reasoning inevitably leads to a semantic

[70] A significant debate about this the difference between explanation and understanding. Since Hempel (1965), philosophers have made this distinction. The gist of the distinction is that while "explanation may be viewed as a two-term relation between an explanans (for example, a theory) and an explanandum (the

exploration of what is meant by explanation and if it requires us to take cognizance of pragmatic interests of the explainer, if explanations should have generality to them [Sheredos (2016)], or it needs a linguistic analysis of 'explanation' [Wright (2012)]. I do not want to address these debates, and instead stay within the purview of scientific practices. Under such constraints, I therefore leave these questions posed by (OC1) conception aside and analyse the other conception of ontic explanations – that of norms for what models can be explanatory.

2.2.1: Ontic Conception as Norms for Modelling

(OC2) This interpretation is not about mechanisms as causes but a way to evaluate models. For Craver, a good explanation is one that is exhibit how the phenomenon 'fits into the causal structure of the world'. Therefore, the mechanistic explanation has to be mapped onto the real world – i.e. it should allude to the components and their organisation that are *actually* causal for the phenomenon. Craver is ready to accept that modelling mechanism is an epistemic exercise, but that the real-world phenomenon that we aim to explain places some constraints on what this model must do. First and foremost, it must allude to real causal structure and components that *actually* bring about the phenomenon.

phenomenon), understanding is always a three-term relation involving explanans, explanandum, and a subject" [de Regt, Leonelli, and Eigner (2009b), p.3]. While some have taken it to mean that understanding is a purely subjective matter [Trout (2002)], others have pointed out that there can be criteria that differentiate between mere Feeling of Understanding and Genuine Understanding [de Regt (2009)]. In this respect, I agree with Diéguez (2013) who argues that there are objective criteria for understanding through models, although they might not be universal.

(OC2) does not rule out that other kind of explanations – like phenomenal, or predictive explanations exist, but that a good explanation is the ontically adequate one – i.e. is constrained by ontic considerations – or what actually exists; and as per Craver, only mechanistic explanations are capable of doing that. And that the representations of the real mechanisms that exist out there in the world are only good if they satisfy the ontic commitment. Craver writes –

> "ontic explanations… make an essential contribution to the criteria for evaluating explanatory communications, texts (models), and representations. Good mechanistic explanatory models are good in part because they correctly represent objective explanations."[71]

Craver gives us an example of what he means. According to Craver, there is a difference between how-possible and how-actual models. How possible models tell us if the phenomenon is possible within our understanding of the world. So, for example, an equation $(x – h)^2 + (y – k)^2 = r^2$ tells us that circles can exist, but not exactly that one actual circle exists – that is the aim of how-actual models. Craver points out that there are models that explain how a phenomenon is possible; for example – the Hodgkin-Huxley (HH) equation of action potential tells us how it is possible for the action potential to be created and propagated in the neurons. But since HH is a set of equations that does not allude to the components and their actual activities, it is not a how-actual model. In theory, the HH equation can be realised by many different causal structures – i.e. may real-world mechanisms can exist (or be created) that satisfy the equation. Thus, the HH model is a how-possible model. The how-actual explanation of the action

[71] Craver (2014), p40.

potential in neurons was first elucidated not by HH equations, but by identification of the components and their organisation.

Essentially, Craver holds that explicating mechanistic explanations is first and foremost an examination of the ontic structure of the world. In both (OC1) and (OC2) a good explanation is either one that shows the causes in re[72], or if representational then must allude to components that exist in re. But both these conditions are question-begging.

2.2.2: The Ontic Account and its Problems

(OC2) is not a condition about the causal structure of the world but of the explanatory models. It requires that the ontic considerations constrain the explanatory representations. Therefore, while Craver may grant that explanations are an epistemic endeavour and are marred with our epistemic limitations. But he holds that a good explanation is the one that has ontic content in it – and since for Craver, an ontic explanation is mechanistic and out there in the world, any representation must refer to the structure of the mechanism. This is a more reasonable position but one that fails when we consider that the scope of mechanistic explanation, at least the epistemic kind, is rather limited. As we have seen with the Nieuwkoop centre example, the Craverian mechanisms have a problem when it comes to biological systems where redundancy and reorganisation are at play. But Craverian mechanisms have other problems too. Many authors [Brigandt (2015), Gross (2015), Théry (2015)] have pointed out how

[72] I take this term from Wright (2015), where it is defined as "located within being or in the matter of" (p21).

mechanistic explanations cannot account for various features of a biological system – be it spatio-temporal diffusion, non-object entities, or dynamicity.

Let's consider dynamicity. Biological systems show time-dependent phenomena. They are not only spatially organised but also temporally organised. This dynamicity of biological systems is essential to their functions. Moreover, the components of a system might change over time or even come into or go out of existence. Take the protein synthesis for example. In recent studies[73] based on a whole-cell model simulation of translation in *S. cerevisiae*, it was observed that the oscillation in mRNA density leads to varying effects on the translation process. The oscillatory nature of mRNA determines which gene codons are translated and expressed. This oscillatory nature of the mRNA population is very much a part of what proteins and in what composition are they present in the cell. While it would be easy to give a mechanistic story about how mRNA is formed and how they are translated, it is difficult for mechanistic literature to account for this time-dependent oscillatory phenomenon. Many a time, in such situations, the scientists thus use differential equations to map the variable that may correspond to some components of the system. In the case of the whole-cell model simulation of translation in *S. cerevisiae*, the variables in question are the quantities of free ribosomes that are present and the transcription rate of the genes. In such cases, time-dependent equations are better tools for modelling the system that exhibits the phenomenon. Craver would grant that models other than mechanistic models exist. Still, he would maintain that these cannot be qualified as 'good models' based on his ontic considerations – this kind of model, as per him, is phenomenal.

[73] Zarai and Tuller (2019)

The defenders of dynamic models might argue that dynamic models still explain in the virtue of elucidating the properties of the system in question. Moreover, the tools that the mechanists use for analysing mechanisms and their manipulability – i.e. Woodward's Interventionist account – apply to equations dealing with variables despite them not being 'mechanistic'. What these models, as per their proponents, do is to analyse a system in a way that mechanisms cannot – i.e. the dynamicity. They would be correct in pointing this out. As Woodward (2013) points out, there are specific conditions that a system must have to be studied by the interventionist framework –

1. *Stability* – that the relationship between the components is stable throughout changes in background condition. The more stable this relationship, the more we can generalize it.
2. *Modularity* – the condition of modularity captures the idea that the relationship between components is not affected by changes in the relationship between some other sets of components – i.e. the when a specific subset of components is intervened upon, it should not lead to change or disruption of the relationship between another set of components.
3. *Organisational sensitivity and fine-tunedness* – this concerns with how readily can the system be decomposed for mechanistic analysis. As per Woodward (2013) some systems show better spatio-temporal demarcation than others and hence are much better contenders for decomposition while others are not.

Woodward's criteria show how interventionism might fail in cases that do not show stability, modularity and fine-tunedness of organisation. As Craver himself notes in his mutual manipulation criteria, many systems that show redundancy or reorganisation, will not satisfy these criteria. In our example of axis development, the relationship

between the Nieuwkoop centre and axis formation was not stable (as we can remove the Nieuwkoop centre and still have axis development go on unhinged, but the introduction of an addition Nieuwkoop centre leads to two axes being developed). It was not modular (since the remaining embryo reorganises itself in the absence of Nieuwkoop centre to proceed on to develop the axis). And it was not fine-tuned (as the Nieuwkoop centre is not spatially demarcated from its surrounding). All this should make us think that the axis formation phenomenon is then only minimally mechanistic (in Craverian sense).

But, at the same time, many differential equations will show the conditions of stability, modularity and fine-tunedness, and hence can be intervened upon while satisfying Woodward's conditions for ideal intervention. Woodward's framework is an epistemological toolkit that can be used to analyse various explanatory models. Mechanisms use the Interventionist account for analysis, but it is not sufficient condition for an explanation to be called mechanistic. These conditions are concerned with how the mechanistic model explains. Bechtel (2008), for example, writes that mechanistic explanations are necessarily engaged in decomposition into parts and thus "emphasizes the contributions made by parts of a mechanism to its operations"[74]. Building on these conditions, Hochstein (2016) notes that for an explanation to be called mechanistic, it should elucidate –

> "(1) The parts of the system.
>
> (2) The way in which these parts are spatially and temporally organized within the system.
>
> (3) The operations that go on between the relevant component parts.

[74] Bechtel (2008), p21.

(4) The resulting phenomenon produced by the system."[75]

And while Woodward's account of causation provides a useful epistemic guide to how to achieve (1) through (4), it is not the crux of mechanistic explanation[76]. Mechanistic explanations explain by decomposition of the system into parts and elucidating how these parts bring about the phenomenon. But it does not explain the ontic constraints on the explanation that Craver enforces. This framework can be construed as a purely epistemic exercise where recognition of parts and their operations is dependent on the modeller. For example, the modeller can still construe the Nieuwkoop centre as a demarcated part (at some stages) through processual partition frame [Winther (2006a)] Craver will need more than this to justify (OC2).

Kaplan and Craver (2011) argue that there is a condition that can guide our intuition about what successful models must do. He calls his conditions "models to-mechanism mapping" (or 3M) –

> "(3M) In successful explanatory models… (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism."[77]

[75] Hochstein (2016), p1393.

[76] Albeit, the failure of a system to be analysed by Woodward's interventionist account would make it more likely to be not explicable in mechanistic terms given that it will not satisfy the conditions of stability, modularity and finetuned-ness and hence would be difficult to decomposed into components.

[77] Kaplan and Craver (2011), p611. It must be noted that Kaplan and Craver (2011) identifies mechanistic explanations as successful.

The (3M) condition, along with Woodward's manipulationist account gives us a good idea of what it is meant to be a mechanistic model. The model should represent the constituents and causal relations between them. Although Kaplan and Craver do not provide a proper definition of what 'correspondence' means, but a reading of the text suggests that since they are dedicated to explanations being as close to mechanisms out there in the world (mechanism realism, as they write), the more accurate a representation the better it is. But, as noted above, this excludes a lot of other models – like topological models, dynamic equations, etc. – from the purview of being able to provide mechanistic explanations. A downside of this is that any important information coded into these models can essentially not be incorporated into the mechanistic explanation. Important phenomena – like oscillations, network analysis, etc. – are thus, outside the scope of providing mechanistic explanations.

Levy and Bechtel (2013) point out this problem. They argue that complex biological mechanisms need us to exclude fine-grained information about the structure of the mechanism. So, say, we want to analyse how a particular system is organised, we need to use abstraction. When, for example, gene regulatory networks are studied, the focus of the modelling is on how a network is constructed rather than what the exact node does. Network motifs[78] like feedforward loops[79] are constructed which describe the working of networks without much attention to the functional or structural features of

[78] I will focus on network motifs and graph theory further in the next chapter when I will discuss Systems Biology and how network analysis can provide us with valuable models. For now, it will suffice to say that biologists have discovered that network motifs have a ubiquitous presence in biological systems and are largely independent on what the nodes are. Hence, we can find them in gene regulatory networks through metabolism to ecology.

[79] A feedforward loop is a network motif in which a transcription factor (X) for the node Z regulates another transcription factor (Y) for Z. And both X and Y together regulate Z. Such loops have been found in *E.coli* gene regulatory network. See Chapter 3 for more details.

the nodes. Levy and Bechtel (2013) note that to be able to study such networks in biology, especially ones that are dynamic and time-dependent, we need to abstract away from the structural features of the individual components and their specific interactions and analyse the network's architecture. They write–

> "to understand organization, one often needs to abstract from the structural specifics of a mechanism and represent it in a skeletal, coarse-grained manner."[80]

Levy and Bechtel present an important point. Even if we grant that the nature of causation is mechanistic as per (OC1), leaving out explanatory models that are not mechanistic will be a mistake as the phenomena that these models explain are relevant to explain several ontic features of the world. We do see oscillation in protein concentrations due to change in the number of available mRNAs. Quantitative, dynamic and network models contribute to explanations of the phenomenon.

On the other hand, Piccinini and Craver (2011) have suggested that models that are not mechanistic as per (OC2) are still relevant as "sketches of mechanisms"[81] in which some structural aspect of mechanistic explanations are omitted. They hold that these models are not full-blown mechanistic explanations but "place direct constraints on any structures that can possibly process such information—on how the different states of the system can be constructed, combined, and manipulated—and are in turn constrained by the structures to be found in the system"[82]. That is, explanatory models that do not

[80] Levy and Bechtel (2013), p241

[81] Sketches and Schema are described as descriptions of mechanisms which are not explanations. Sketches are partial descriptions of mechanisms that leave out some relevant components and organisational detail, while Schemas are more complete descriptions of mechanisms with the relevant entities, activities, and organisation included. See Craver and Darden (2013).

[82] Piccinini and Craver (2011), p296. Piccinini and Craver talk about functional analysis of cognition as information processing systems, specifically. But the idea can be extrapolated to apply to other models too. Consider the network motifs for example. Network motifs can be constructed as logical gate systems

abide by (OC2) are still relevant because they can explain details about mechanisms by supplementing mechanistic models with information that is relevant to flesh out mechanisms in further details. These are, as per Piccinini and Craver (2011) further constraints on models and are in turn, constrained by the mechanism. Furthermore, these explanations can be turned mechanistic once we fill in the 'missing aspects' that were omitted.

The (OC2) condition can now be altered to reflect this point of view. A good explanatory model is one that provides relevant information about the mechanism, even if it does not allude to the structure of the mechanisms itself. Let us call this **(OC2')**.

I think that (OC2') is a much more agreeable position for ontic constraints on explanatory models. In the next section, I show that the middle ground approach to normative constraints for explanatory models can satisfy this condition. But before that I analyse the epistemic account of mechanism which argues that mechanistic explanations are first and foremost modelling exercise aimed at representing the target system in a certain way. Following the epistemic account, I analyse the middle ground approach and show why (OC2) is unsuitable for it and why (OC2') is suitable. Lastly, I show that mechanistic explanations must be construed as multi-model idealizations and that mechanistic explanations can be derived from non-mechanistic models that do not aim to represent the difference making factors. The latter will be done using the example of whole cell modelling.

with information processing capacities. Therefore, we can have network motifs that act like AND gate. In such cases, Piccinini and Craver's argument holds that the motif places some constraints over what kind of entities can realise an AND gate.

2.3: The Epistemic Account of Mechanisms

Bechtel and Abrahamsen (2005, 2009) take the epistemic approach to mechanistic explanations. For them, the features of a mechanistic explanation are largely explanatory –

> "[E]xplanation is itself an epistemic activity, what figures in it are not the mechanisms in the world, but representations of them. These representations may be internal mental representations, but they may also take the form of representations external to the cognitive agent—diagrams, linguistic descriptions, mathematical equations, physical models, and so on. Generically, one can refer to these internal and external representations as models of the mechanism."[83]

Bechtel and Abrahamsen (2005, 2009) focus on how mechanistic models are constructed. For them, mechanistic explanations are engaged in –

1. *Decomposition* – Identifying the components that take part in bringing about the phenomenon,
2. *Localisation* – assigning distinct causal roles to these parts; and
3. *Recomposition* – explaining how these components, when taken together, explain the phenomenon in its environment.

They note –

> "In order to develop a mechanistic explanation. It is necessary to decompose the mechanism into parts and/or the operations they perform. Doing so often requires specialized instruments and techniques… Many of the advances in modern

[83] Bechtel and Abrahamsen (2005), p425.

> biology have stemmed from the development of particularly effective means for decomposing particular classes of biological systems. Scientists have not explained a phenomenon, however, until they have recomposed the mechanism - determined how it is organized and interacts with its environment."[84]

Bechtel and Abrahamsen are interested in what can be called Epistemic Constraints on mechanistic models. The problem, as they identify, is that both decomposition and recomposition require a certain epistemic rigour in the face of the challenges that the system that is being studied presents. Decomposition might be problematic because components might not be localised and easily spatially demarcated. Recomposition, on the other hand, may present its own problems of integration and effects that are more systematic. They write –

> "As challenging as is decomposition, recomposition often is even more difficult. Parts and operations are often highly integrated with one another so that as one part of the mechanism changes, numerous other parts are affected and change their operations…Ambitious modelling projects, sometimes using the available tools for analyzing dynamical systems, often are required to understand the consequences of organization and particularly the orchestration of multiple operations in real-time."[85]

Bechtel and Abrahamsen recognise the limitations that mechanistic models might face while reorganising. Systemic integration and system-wide effects might be challenging to a modeller because they yield unexpected results. Parts may affect other parts. One might encounter feedback loops and cyclic operations. In such conditions, they argue

[84] Bechtel and Abrahamsen (2009), p179.
[85] Bechtel and Abrahamsen (2009), p179.

that other techniques (like dynamic modelling using equations) can be used instead of standard mechanistic modelling. Bechtel (2011) is committed to what he called 'dynamic mechanistic models' – where the recomposition exercise is conducted with the help of computational models. Their view is consistent with his views of mechanisms being an epistemic exercise. Bechtel is interested in modelling the given system as well as it can be done within the constraints of the particular model. Each model has its own heuristics and limitations. As per this view, mechanisms are but one explanatory model. Levy (2013) calls this conception of mechanism *Strategic Mechanism* to differentiate it from Craver's ontic mechanism (which he calls *Explanatory mechanism*). According to Levy, the driving force behind the strategic mechanism is that some biological mechanisms can be represented with mechanistic models owing to their structure, but others may be more difficult to do so. In essence, and as Bechtel and Richardson (1993) note, mechanistic models have the analogy of actual machines in mind. This does not mean that they see mechanisms as machines, but that the heuristics of localisation and decomposition require certain modularity that is paradigmatically present in machines. Mechanisms are fruitful if there is enough modularity in the target system, but they will fail to be explanatory if this modularity is absent. But there is another motivation for building mechanistic explanations for the strategic mechanists – that of the epistemic conception (EC) of explanatory models –

(EC) Explanatory models are human activities and are bound by epistemic limitations. The normative constraints on explanations are primarily epistemic – such that models are built on understanding and communicating as an exercise in gathering knowledge.

As per (EC), the essence of the model is not to explore the ontology of the system, but to model it in the best way to be communicative and representational. Therefore,

proponents of (EC) attach much higher significance to mechanistic sketches, diagrams, and schemas than the proponents of (OC) – mechanistic sketches, schemas and diagrams are important models of representation and communication. Darden (2006), for example, describes mechanistic schema as having "specify roles, black boxes, at varying degrees of abstraction with more or less detail specified... [t]he schema terms can then be filled with the mechanism's entities and activities as they are hypothesized and discovered"[86]. Proponents of (EC) also argue that idealization and abstraction play an important role in the construction of mechanistic models for three reasons –

1. To explain a complex system, the modeller, with her epistemic limitations, must deliberately make a choice between relevant and irrelevant factors. Abstraction is not just a feature of a model but a requirement for the exercise of modelling.
2. Since the paradigmatic examples of mechanisms are machines, it holds that the kind of modularity needed to model mechanisms must sometimes be introduced to make the model more fruitful. In such cases, idealizations might be employed to make a 'fictional model'[87] that can be studied using mechanistic representation.
3. Especially in the case of sketches and schema, abstracting away and idealizing from the finer details might be more fruitful for communicative convenience.

For example, the Nieuwkoop centre can still be considered as a part in the EC by introducing a fine-tunedness which may not exist in the real system. This can be justified as a modelling requirement and the assumptions about the demarcation of the Nieuwkoop centre can be stated while noting the failure of decomposition and localisation as properly applicable heuristics in this case. But even with these

[86] Darden, (2006), p86-87.
[87] See Bokulich (2016)

limitations, some aspects of Nieuwkoop centre can still be explained. For example, by construing it as a spatially demarcated part with causal significance for the primary organiser we can explain the grafting experiment. On the other hand, the removal experiment can be accounted for by highlighting the failure of decomposition and localisation. Taken together, the two ways to represent the Nieuwkoop centre shows that localisation and decomposition might not be the best heuristics for this case. But as mentioned in Chapter 1, the fallibility of heuristics is itself explanatory [Bechtel and Richardson (1993)].

The proponents of (OC2) might agree with these modelling techniques. But might hold that ontic constraints are normative and primary for mechanistic models. Others, like Illari (2013), argue that both epistemic and ontic constraints are equally important. As per Illari (2013) –

> "It seems that the most sensible conclusion to draw is that neither aim of mechanistic explanation [ontic or epistemic] is prior to the other. Ontic and epistemic constraints are both ineliminable, as both aims must be met, to generate a successful mechanistic explanation:
>
> - Describe the (causal) structure of the world: to be distinctively mechanistic, describe the entities and activities and the organization by which they produce the phenomenon or phenomena.
> - Build a model of the activities, entities and their organization that scientists can understand, model, manipulate and communicate, so that it

> is suitable for the ongoing process of knowledge-gathering in the sciences."[88]

While Illari's argument rings true as a possible resolution to what mechanistic explanations should be like – the proponents of OC should agree that model building is guided by epistemic considerations and the proponents of EC should agree that the models must have some ontic content to be useful to elucidate mechanistic explanations. But it must be noted that this middle-of-the-road conception does not hold for the OC1 as it circumvents the identity claim. Therefore, mechanisms and explanations do not need to be the same thing (although they might be in some cases. Illari notes that the debate, at least the one relevant to her conception is about normative constraints). Therefore, explanations can very well be representational given that they represent at least some aspects of the mechanism. A thing to note though is that this not a trouble for (OC2), and if ontic conception is to be interpreted as constraints on what kind of explanations and models count as mechanistic then Illari's middle ground approach is appropriate resolution to the problem.

But this would also imply changes to how the onticists construe the models. Love and Nathan (2015) note that while it is generally argued that the difference-making aspects of models are not idealized [Strevens (2008)][89], it might not be true. In their example of protein synthesis, they note that idealization of the causal features appears in several ways. Firstly, the binding sites complexity is reduced in these models; secondly, the concentrations of different causal factors, which can itself be causal is not

[88] Illari (2013), p250.

[89] Strevens (2008) notes – "The content of an idealized model, then, can be divided into two parts. The first part contains the difference-makers for the explanatory target ... The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target" [Strevens (2008), 318)].

considered. While a reply can be that mechanistic models should be construed as mechanistic schemas, the problem is that difference-making factors are significantly idealized in these models. The idealization thus used, may obscure difference making factors. Love and Nathan (2015) instead argue that explanatory and discovery exercise in model making is inherently epistemic and needs idealization and abstraction of factors that are either not the focus of the study or are too complex to model. They write–

> "The idealization of causal relations demonstrates that these models do not depict how the mechanism actually works. If actual difference-makers are represented in such a way that they are not difference-makers, according to what is already known about the mechanism, mechanistic explanations appear to fail according to their own criteria."[90]

In such cases, the idealization of relevant difference-making factors is not only warranted but necessary. In this regard Model to mechanism mapping (3M) criteria for good mechanistic model is too stringent.

Another consideration is that models that appeal to different kinds of explanations (non-causal, design) can still have relevance for discovery of mechanisms. In Chapter 3, I argue that mathematical models of systems biology must be construed as minimal models that show why certain causal factors are irrelevant in explaining a phenomenon. These are optimal models that idealize and abstract away from various causal factors. Rice (2015) notes that optimal models do not aim to provide causal explanations and are holistically idealized. What is meant by holistically idealized is that the explanatory

[90] Love and Nathan (2015), p.768.

power of the model is derived from idealizations and they cannot be just decomposed into idealizations about difference-making and non-difference-making factors [Rice (2019)]. But they can still aid mechanistic explanations in form of constraint-based reasoning. Similarly, design explanations [see Wouters (2007), (2013)] explain by comparing a given system to another system that lacks the given trait. The comparison provides a counterfactual answer to a 'why' question that biologists aim to answer – why the trait is present and what advantage (in terms of life chances) it provides. As Wouters (2013) writes –

> "The how-questions are answered by describing features of the system that produce the relevant abilities, activities or characteristic… The why-questions are answered by describing features of the organism, its environment or its way of life due to which the characteristics to be explained are advantageous to the organisms that have it. Biologists typically call this type of answers to how-questions 'mechanistic' or 'causal' explanations and this type of answers to why-questions 'functional' or 'ecological' explanations"[91]

But still, the design explanations can have consequences for mechanism discovery, despite being not framed as mechanistic models, by providing 'constraint based' reasoning about mechanistic features.

If we consider the plurality of scientific practices, then the ontic constraints on epistemic mechanistic modelling seems misguided on the stronger view. I believe that this is reason enough to discard the 3M criteria as too stringent. On the other hand, (OC2') provides a good criterion for determining the ontic constraints on mechanisms.

[91] Wouters (2013), p.463.

As per (OC2'), models need not be accurate representations of mechanisms but need to provide, at the least, some information about the mechanisms in form of constraints-based reasoning. In such cases the model can be considered a mechanistic sketch. I think this ontic constraint, albeit a very diluted version of Craver's original ontic conception is best reflected in our scientific practices.

2.4: Mechanistic Explanations and Multi-model Idealizations

Love and Nathan (2015) provide another interesting argument that I think has some bearings on how we construe mechanistic explanations (and explanations in biology in general). As per them, mechanistic explanations are generally considered to be minimalist model idealizations, instead they should be considered as multi-model idealizations. The concept of minimalist and multi-model idealization comes from Weisberg (2007a). Weisberg (2007a) differentiates between three kinds of idealizations–

1. *Galilean idealization* – the practice of introducing distortion in theories with the aim to simplify them and making them computationally simple and can be used both for constructing experiments and theories that may explain the phenomenon.[92]
2. *Minimalist idealization* – the practice of constructing models that only account for the core features of the system. Constructing minimal models include abstracting away from irrelevant details, modelling partial systems, and making models

[92] Weisberg (2007a), p640.

simpler for communication and understanding by only focusing on the difference-making factors.[93]

3. *Multi-model idealization* – the practice of making several incompatible models of the same phenomenon. These models describe different aspects of the phenomenon. The motivation behind building multiple models with their different idealizations is that models have different goals and different standards of accuracy, precision, generality and simplicity. Furthermore, these models are also motivated by understanding that modelling complex systems is difficult. Multi-model idealization, therefore, work as a way of overcoming our cognitive limitations while studying complex systems through trade-offs of what a single model can represent.[94]

Weisberg notes that even though the multiple-models with their particular idealizations are not good contenders for scientific realism, taken together, these models can provide various representations of a system (unlike minimalist models) which in turn can provide various dimensions to understand the real world. In essence, multiple-models are effective in overcoming the epistemic limitations that plague our modelling methods, not by having a particular dedication to realism, but by representing the various aspects of the system in different models, thus providing a better picture of the world when taken together.

Love and Nathan (2015), thus argue that mechanistic explanations consist of multiple models with their own idealizations such that different models are idealized according to the pragmatic considerations of the modeller (i.e. what factors they want to study and

[93] Weisberg (2007a), p.642-3.
[94] Weisberg (2007a), p666-7.

reason about) and the demand of the explanations. Therefore, imposing ontic constraints of accurately representing the mechanistic features on each model is not only a stringent condition, but counterproductive. While some mechanistic models can be minimalistic idealizations, other's (especially those which are not engaged in giving direct mechanistic representations) might not be. But taken together, they may imply useful information for mechanistic discovery, representation, and explanation. Furthermore, as noted by many proponents of EC [Levy and Bechtel (2013), Bechtel and Abrahemsen (2005, 2009)], our heuristic strategies of constructing mechanistic models might fail in various cases. But if mechanistic models are supplemented with other kinds of models, such that each model has its own goals, idealizations and representations, we might be better suited to satisfy (OC) as a whole when it comes to scientific practices.

Therefore, for (OC2), the aim of giving mechanistic explanations is much better served by constructing various models that have their own explanatory goals. In this respect, I believe that (OC2') is better suited to be compatible with the goals of a multi-model system when it comes to exploring mechanistic explanations.

Hochstein (2016) has given an example of how this is possible. Hochstein argues that mechanisms are not present in a singular model but are distributed amongst various models. As an example, he considers the action potential of neurons as a phenomenon to be explained. He notes –

> "During the twentieth century, various different modelling techniques were used to generate different kinds of models of the action potential, yet each was only

> partial in its characterization of the underlying mechanism, with no single model able to represent the various aspects of it simultaneously."[95]

Hochstein notes how different models provided only partial explanations. Hodgkin-Huxley model (1952), although provided a mathematical model for generation and propagation of action potential, was not sufficient enough to provide the structural features of the system. Similarly, the single-channel recording, although it could represent sodium and potential channels, was insufficient to provide a causal structure. Another model – the Fluid Mosaic Model [developed by Singer and Nicholson (1972] focused on modelling the cell membrane by representing the organisational structure of lipids and proteins but was not able to model the dynamicity that the electrical models like Hodgkin-Huxly could.[96]

Hochstein (2016) believes that this shows that the desired mechanistic explanation is distributed amongst different models. Some models contribute towards our understanding of some features of the mechanism, while others do the same for other features. At first reading, Hochstein's claims seem pretty similar to Piccinini and Craver (2011)'s idea of different explanatory models being supplementary to mechanistic explanations. But this will be a limited reading of his example. It is not only that the mechanistic explanation is distributed amongst different models, but that if we aim at scientific explanations to be realist, then the ontic information about the system is indeed distributed amongst several different models of the same system. These models together form the basis for our understanding of the world as it actually is. These might be mechanistic in some cases, but not in others.

[95] Hochstein (2016), p1397.
[96] Hochstein (2016), p1397-8.

Therefore, I hold a stronger reading of Love and Nathan (2015). Explanations simpliciter are multi-model idealizations. According to this reading, explanations (mechanistic or otherwise) are composed of multiple models, each constructed with its own goals, within their own theoretical framework, and with their own heuristics, idealizations, abstractions, and assumptions. Taken alone, these models only provide piecemeal explanations regarding the system, but when taken together, they can provide a better understanding and fuller explanation of the system's behaviour and thus the conditions for (OC1) must be applied to the multi-model system rather than every model individually. As I will show in Chapter 5, this is the case for bacterial chemotaxis in *E. coli*, where no one model can capture the complete behaviour of the chemotactic system. Therefore, this reading is guided by the analysis and understanding of scientific practices in biology.

So, how can we get the most out of mechanisms? Mechanistic explanations are an important part of scientific exercise. As mechanists like to note – scientists do study and describe the system as mechanisms. The heuristics of localisation, decomposition, and the study of entities and activities provide a way for philosophers to explain how scientists explain the target systems, its functioning and its organisation. In many cases, mechanistic explanations are best suited to explain the phenomena, and in many cases, they might be a good window into the ontic structure of the world. But I believe that some mechanists have taken this to mean that mechanistic explanations are in some way more suited for scientific practices than non-mechanistic explanations. I believe that this is a mistake. Mechanistic models and explanations are an excellent epistemic tool for representing the causes that constitute phenomena. But they need other representational models to be ontically adequate. In this respect, Bechtel's epistemic

view is better suited for being integrated with other models, and despite not being as dedicated to the stronger ontic claims of Craver, it is better suited to analysing the ontology of biological systems.

2.4.1: Whole Cell Model

I believe a final example might explain the commitment to modelling as an epistemic exercise which eventually leads to explaining some ontic features of the system of study might make this claim clearer. For this I think a good example is that of whole-cell models. Whole-cell models have been a recent achievement of inter-disciplinary studies in molecular biology. Karr, et al. (2012) developed a whole-cell computational model for *Mycoplasma genitalium*. Based on various studies and models of the organism, it takes the mechanistic model as its starting point. Various studies of *M. genitallium* have already decomposed the cell into its various components and localised their causal contribution to the sustenance of the cell. The challenge of the whole-cell model is to be able to recompose these components in such a way that it can simulate the functioning of the whole cell. The way that Karr, et al. (2012) do that is by first dividing the cell into 28 modules and modelling them individually. *Figure 2.4* shows these modules.

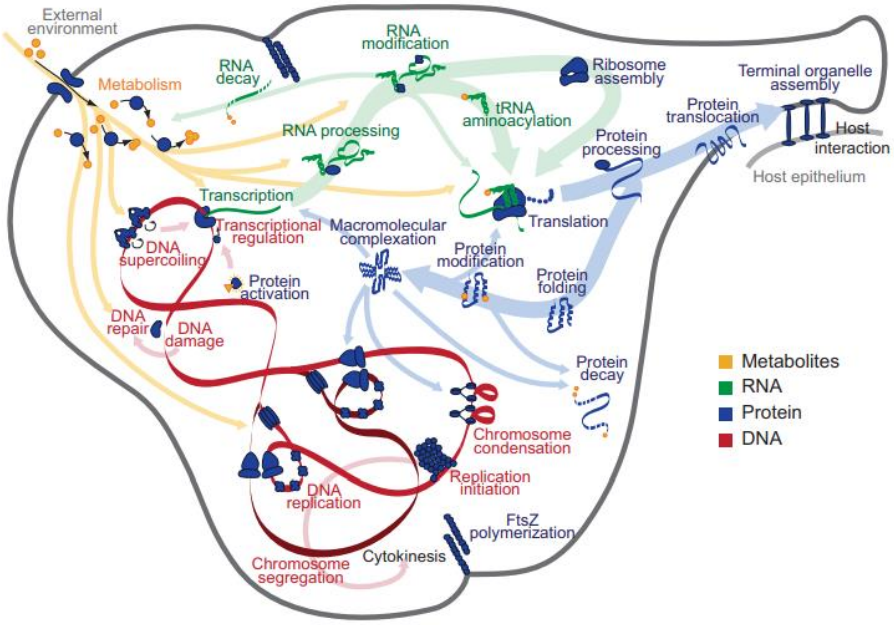

Figure 2.4 – Whole-cell model of M. genitallium showing 28 modules. Each module is composed of various components and mathematically modelled. From Karr, at al (2012), p390.

They further modelled each module individually using previous data and models (submodels). For each submodel, various variables were identified for this modelling. In all, they identified 16 variables (see *Figure 2.5*). Finally, the modules were individually modelled using different methods, for example, "metabolism was modelled using flux-balance analysis, whereas RNA and protein degradation were modelled as Poisson processes"[97]. These are highly idealized models that show dependencies between the variables but do not accurately represent either the components or the causal processes.

[97] Karr, et al (2012), p390.

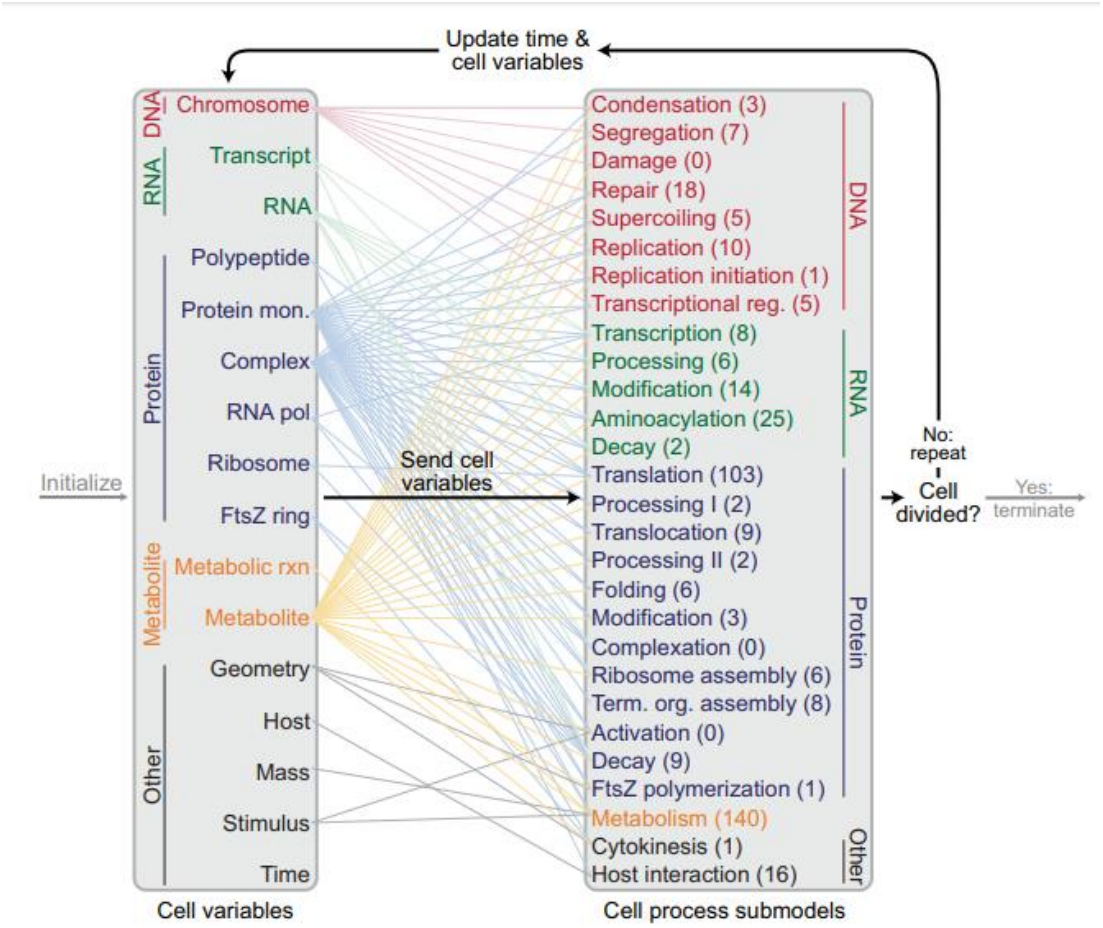

Figure 2.5 – Cell cycle simulation: on the left-hand side is a list of all the cell variables, the right-hand side is a list of all cell submodels. The arrows indicate the dependency of the submodels of the variables. The simulation is run over and over until the cell is divided. Karr, et al (2012), p390.

Finally, they tried to integrate the sub-models to simulate the whole cell. They realized that the cell variables play a part in various modules – thus leading to the interdependency of the sub-models on each other. Example – replication needs nucleotides which are produced during metabolism. Hence, they needed not just to integrate the individual modules into one assembly (as if being independently obtaining the inputs) but temporally integrate them such that each module must obtain its input from other modules and thus be dependent on their functioning. For this, they introduced

an assumption that each sub-model was independent in its function for a time less than 1 second but were dependent on other sub-models for larger timescales. They write –

> "Simulations are then performed by running through a loop in which the submodels are run independently at each time step but depend on the values of variables determined by the other submodels at the previous time step."[98]

By introducing this assumption, they introduced interdependency of the sub-models on each other, albeit only on a larger timescale, and hence a partial temporal recomposition. Of course, the assumption that these sub-models are independent on timescales less than 1 second is not accurate, but it is due to constraints of what the computer simulation can do.

Finally, they ran the models for different variables and different sets of gene configurations. They use their model for predictive purposes, but one of the things they found was that the cell to cell variation of the cell cycle was comparatively lower than expected. But at the same time, the variation in the timescale of some modules (especially replication initiation, and replication) was much higher. That is, that even though in some gene configurations, even when the difference between the timescales of these modules were significant, the variation between the time taken for the cell cycle to complete was much less. The order of magnitudes was also not non-significant. Replication initiation showed a variation of 64.3%, replication showed a variation of 38.5%, while the cell cycle showed a much less variation of 9.4%.

On further investigation, they found that the replication submodel functioned on two different speeds subject to availability of deoxyribonucleotide triphosphate (dNTP).

[98] Karr, et al (2012), p390.

dNTP is produced by the replication initiation model and is used up in the replication model. They found that there was an inverse relationship between the two modules, such that if the initiation model takes longer time, then the replication model takes shorter times. They write –

> "longer initiation times led to shorter replication time… cells that require extra time to initiate replication also build up a large dNTP surplus, leading to faster replication."[99]

Thus, the overall cell cycle remains relatively invariant. The temporal recomposition of the whole-cell hints towards how cells can impose control on their organisation due to constraints imposed by different modules.

What the whole-cell model shows is that we can build on top of other models and add new representations of the system, which in turn help us understand the features of the original models. What Karr, at al (2012) do is to build upon mechanistic models that already exist. These models have been built over time and have elucidated various features of the structural and causal organisation of the target system. When Karr, et al. (2012) build upon these models, they utilise computational techniques to recompose these models in a certain way so that new explanations about the system as a whole. This computational model is then used by them to explain features of the system that were not captured by any individual mechanistic models. This exercise leads to a better understanding of the mechanistic models. The casual structure of replication has been modelled in the past. These models have modelled how dNTP is causal for replication. But Karr, et al. (2012) are able to enrich this model by elucidating how this mechanism

[99] Karr, et al (2012), p393.

of replication is affected by another sub-mechanism which leads to different rates of replication – thus, adding to our understanding of difference-making factors for replication. Besides, they show how this may affect the overall phenomenon of interest (the cell cycle). By doing this, they have added to our understanding of the mechanism of replication. But they have done it without the considerations that (OC2) requires. They did not set out to give a mechanistic model of the whole cell. They intentionally modelled the modules with idealized equations while abstracting away from the details. They did not try to make a mechanistic model of the whole cell, but how the sub-models are affected by certain variables and how these variables change over time. Their exercise was purely epistemic with the aim of providing a whole-cell computational model. But they were able to add mechanistic explanations back into the original model through this study.

Therefore, Craver's insistence in (OC2) is unfounded when analysing individual models. Ontic considerations must be imposed on multi-model systems, without being imposed on individual models. This more comprehensive approach to ontic considerations will help us model the system in various ways – guided mostly by epistemic considerations. In many cases, it will finally add to our ontological understanding of the system. Mechanisms are a part of such modelling exercise that must co-exist with other explanatory tools – and for that, they need to be communicative and relevant, rather than too stringent about their respective structure. In this respect, Bechtel is closer to scientific practices than Craver.

2.5: Conclusion

In this chapter, I have argued that mechanistic explanations play a significant role in the scientific process of discovery and understanding by providing a framework to model causal relationships between parts and processes that form a system. But contra Craver, these explanations are not primarily guided by ontic considerations but by epistemic considerations. Our epistemic goals impose certain constraints on how we model a system, and mechanistic models are no exception. The ontic constraints are secondary as no one model can account for the complete ontic picture of the system. For this, we need multiple models with their individual epistemic commitments. The ontic picture is much clearer when we analyse what these models have to tell us about the system when taken together.

CHAPTER THREE

Systems Biology and Mathematical Models

In the last chapter, I analysed mechanistic models. I argued that mechanistic models should be considered as epistemic exercise into the discovery of causal relations that bring about a phenomenon with the relaxation of the ontic conception. In this chapter, I look at mathematical models as used in the field of systems biology. The chapter aims to analyse what do mathematical/computational models explain and how do they explain. The structure of the chapter is as follows. First, I analyse what Dynamic Mechanistic Explanations are and following Issad and Malaterre (2015), I argue that the explanatory work in dynamic mechanistic explanations if done by the mathematical formalisation rather than by the mechanistic model. After which, following brief analysis of the field of systems biology, I analyse three methods of mathematical/computational modelling –

a) network motif analysis,
b) large-scale network analysis, and
c) landscape analysis.

I analyse the modelling techniques used to construct such models. Following which I analyse how mechanistic-minded philosophers have argued that mathematical models can be used to elucidate dynamic features of the system. I argue that they are correct that mathematical models can be used to analyse dynamic features of mechanisms and for mechanistic discovery, but that this does not explain how mathematical models

explain. Instead mathematical models explain by showing dependencies that hold for a certain class of systems.

3.1: Dynamic Mechanistic Explanations

As noted in the last chapter, mechanistic models sometimes fail to account for the dynamics of the mechanisms that they aim to explain, especially in the cases where non-trivial time-dependent behaviour or quantitative dependence are seen. In the example taken from the last chapter, Zarai and Tuller (2019) show how the mRNA population shows oscillation such that transcription rate depends on the number of free RNA molecules present at any given time. While mechanistic models are able to account for how RNA is transcripted into mRNA, the quantitative nature of the mechanism is not explained by the causal process alone. This is because while the mechanistic model accounts for the causal properties of RNA – how it binds with the DNA strand, and how it copies the corresponding codons; how the time-dependent quantity of free RNA affects the rate of transcription is not part of the causal analysis. But quantities play a vital role in molecular explanations. Nathan (2014) argues that the quantities of entities can play a crucial role in explaining how a phenomenon is brought about – causation by concentration – as he calls it. There are three reasons why it might be true for different systems –

1. A system is not composed of fixed stocks of entities. Entities are generated and degenerated with new entities of the same type replacing the previous stock.
2. Activities that show a binding relation might need a concentration of particular substrates to maintain a steady state of reaction. For example, as Nathan (2014)

notes – "(w)hen a molecule binds to DNA, it does not remain there indefinitely: proteins fall off and reattach all the time"[100]. Therefore, a concentration of similar entities is needed in the vicinity to latch on when one of them falls off.

3. The cell might need to maintain different compositions at various times to facilitate a particular phenomenon.

In other cases, the mechanism might show non-sequential, non-linear dynamics that cannot be captured by a simple mechanistic model. An example of this comes from bistability shown in various biological processes where the system can exist in two different states, and the switching from one state to another shows an all or nothing behaviour. That is, the system will stay in one state until a particular signal persists for a particular duration of time or at a particular intensity. A mechanistic model might be able to diagnose the causal relations that bring about the changes when a specific signal is applied but fails to answer why only at a certain threshold would make a difference to the system's state.

As a remedy to both these quantitative and qualitative behaviour of systems, proponents of mechanistic explanations have offered what has been called 'Dynamic Mechanistic Explanations' (DMEs, henceforth). DME offers a hybrid model such that while it holds that a mechanistic model is required for understanding the causes of a phenomenon, it emphasises the importance of mathematical and computational tools to understand the organisation and behaviour of the system not captures by a simple mechanistic model [Bechtel and Abrahamsen (2010, 2013), Levy and Bechtel (2013)].

The motivation behind DME is that mathematical and computational models are capable of modelling the dynamic behaviour of the mechanistic model, albeit in abstract form.

[100] Nathan (2014), p.196.

Furthermore, such strategy not only explains the dynamic nature of mechanism but can help in mechanism discovery [Baetu (2015a)]. Baetu (2015a) cites the example of the spike of gene expression in leukocytes which is known to be produced by a negative feedback loop mechanism (known as NF-κB regulatory mechanism). But, as he notes, the response of generic negative feedback loops does not show a spike-like behaviour but an oscillatory behaviour. Since this response does not match the observed behaviour, it shows that the mechanistic model is missing some parts that explain the spike-like behaviour. The solution to this came in the form of the observation that there are three isomers of IκB, and only the production of one of them is regulated by a negative feedback loop. At the same time, the other two are expressed continuously. Through mathematical modelling, it was thus observed that the spike seen was due to the additive effect of the expression of the two isomers – one regulated by the negative feedback loop and another expressed continuously. The process, thus, proceeded by showing that the NF-κB regulatory mechanism realises an organisational feature (negative feedback loop) which shows an oscillating behaviour. This oscillating behaviour, when combined with the continuous expression of other components, produced the spike that was observed.

Levy and Bechtel (2013) further argue that organisational features of mechanisms can be understood using mathematical/computational models. To do so, they argue, one must move away from fine-grained mechanistic details to a coarse-grained view of how the components of the mechanism are organised. Abstraction, such that the specific causal details are omitted, thus plays a significant role in aiding the understanding of 'organisational principles'. The model thus abstracts away from the causal details of the mechanistic model and focuses on how the entities and activities are organised in the

system. By doing so, they concentrate on connectivity models. For example, a particular mode of organisation is utilised in the bacterial arabinose regulation system as a *persistence detector*. A persistence detector has the organisational model as shown below –

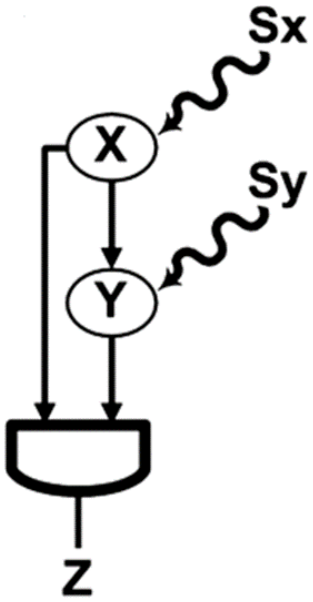

Figure 3.1 – Logic diagram of persistence detector [from Mangan and Alon (2003), p.11980].

The persistence detector has certain generic organisational features, such that – the node X activates the node Y. And collectively, X and Y affect activate the functioning of Z. Thus, as per the structure – Z works as a logical AND gate such that the output of Z depends on both of the input from X and Y; such that when the input signal S_x is active, it regulated Y to propagate the signal S_y. Z, which is regulated by both X and Y, will be only active when both signals S_x and S_y are present. Let us assume that it takes time t_1 for X to activate Z directly when S_x is present, and t_2 for X to activate Z through Y (indirectly). Given that $t_2 > t_1$, Z will only remain active if S_x is persistently present for a time higher than t_2. Therefore, the output of Z is dependent on S_x being active for a

long time. If S_x is not active for long enough, Z will not be active. It is called a persistence detector as Z will be active only when S_x is persistently present. Thus, this organisational structure explains the dynamic behaviour of components without appealing to the molecular details of how the activation of Y or Z takes place. It keeps the focus on the organisational structure of X, Y and Z and features like concentrations as it varies over time. The persistence detector like structure can be found in many systems like the arabinose regulation system amongst others, especially where the signal from X can be noisy.

Thus, as per the proponents of DME, abstract organisational models and their computational models can explain the time-dependent functioning of the mechanism. These computational models, generally based on mathematical derivations, are explanatory when combined with the mechanistic models to present a complete picture of the mechanism and how both its description in terms of entities and activities, and organisation brings about the explanandum phenomenon.

But, Issad and Malaterre (2015) argue that the explanations in DME are not derived as they are in simple mechanistic models – by rehearsing a causal story about the entities and activities and how they bring about a phenomenon – but by mathematical models that necessitate specific outcomes. Dynamic mechanistic models explain by a phenomenon (P) produced by the system (S) by –

1. Constructing a mechanistic model consisting of entities and activities in S.
2. Constructing a mathematical model such that at least some of the variables and functions in the mathematical model correspond to the entities and activities of the mechanistic model.
3. Solving the mathematical model such that it can be shown that P is obtained.

Issad and Malaterre (2015) note that in such a scenario, contra simple mechanistic models, the explanatory work is being done by the computational/mathematical model rather than the mechanistic model. They argue that mechanistic models draw their explanatory power form constructing a mechanistic model M and rehearsing a causal story (CS) about how M realised P. But in case of DME, the explanatory power does not come from the CS, since the complexity of the system might not make a causal story attainable (at least not in the mechanistic form). Instead, the explanatory power of DME comes from the mathematical derivation. For example, in the leukocyte NF-κB case, the spike is not explained by the entities and activities per se but by additive result the mathematical derivation of a negative feedback loop and the other isomer's continuous response. Both these responses are best represented as mathematical equations without any causal story of what the actual entities and activities are and their specificity. Issad and Malaterre (2015) further argue that even when a causal interpretation of the computational model can be provided, it is not what is explanatorily important because the explanatory force lies with showing how the computation model fits with the observed phenomenon. It is necessitated by the solution of the mathematical models. They further present a formalised model for the explanation which they call Causally Interpreted Model Explanation (CIME) –

> "A causally interpreted model explanation (CIME) explains a phenomenon (P) produced by a system (S) in virtue of the following three necessary and sufficient conditions:
>
> (M) Displaying a model of S that includes variables and functions defined over these variables,

(CI) Providing a causal interpretation of these variables and functions (for instance by means of a manipulationist account of causation),

(D) Deriving, through a mathematical treatment (be it analytical or numerical), the explanandum phenomenon P from the model."[101]

According to these conditions, the Mechanistic model is a limiting case of CIME which satisfies conditions (M) and (CI) to construct a mechanistic model of the system (S) where the variables and functions can be supplemented with an exhaustive causal interpretation, such that they represent entities and activities in (S). In addition, the mathematical derivation is simple enough and therefore is conducive to telling a causal story thus, providing a complete mechanistic model of (S). Dynamic Mechanistic Models, according to them, too form a limiting case of CIME where the model (M) is relatively complex due to non-linear dynamics (such that the mathematical derivation is not straightforward), the causal interpretation of the variables and functions is not as straightforward, and some variables do not map onto entities and activities in (S). In such cases, since not all variables and functions are causally interpreted, and the mathematical derivation is more complex, a causal story is elusive. Thus, Dynamic Mechanistic Models –

1. Are not mechanistic models in the traditional sense as they do not lend themselves to complete causal interpretation.
2. Do not derive their explanatory power from rehearsing a causal story but through mathematical/computational models that define the relationships between the variables and functions that are non-linear.

[101] Issad and Malaterre (2015), p.286-87

Issad and Malaterre (2015) thus argue for the importance and explanatory force of mathematical/computational models in biology. But the question remains – how do mathematical/computational models explain? In this chapter, I analyse some methods of mathematical/computational modelling in biology. These examples will be from the nascent field of systems biology. Before moving on to the examples, I take a brief look at systems biology – its goals and methods.

3.2: Systems Biology

While it is difficult to define systems biology, when pushed for a definition, most systems biologists and philosophers define it by explaining how it breaks away from traditional molecular biology. Some common characteristics thus appear.

Firstly, systems biologists note how this new field has its base in systems thinking. System biologists argue that traditional molecular biology focuses on isolated pathways of genomic expressions that are limited in their scope. Thus, they many times do not study how these pathways are situated in a system. Systems biology, on the other hand, they say, is an attempt to understand the systematic behaviour of the biological systems wherein these pathways are situated. This can be achieved through novel techniques, which (and as a second distinction from traditional molecular biology) involve modelling systems through mathematical and computational techniques.

As O'Malley and Dupré (2005) note, the shift in focus from mechanistic qualitative understanding to the understanding of genome-wide effects needed a dynamic framework. The rise of data-driven analysis facilitated this transformation where quantitative analysis took precedence over minute mechanistic details. Systems

Biologists, thus, look not at individual entities but vast quantities and their topology to draw conclusions. As Bechtel and Abrahamsen (2012) note, "When complex dynamics are not present, or are present but not addressed, simple stepwise sequences of operations can be specified on paper or even in the researcher's head. Achieving an account of complex dynamics such as oscillations, though, requires the tools of computational (mathematical) modelling"[102]. The prevalence of computational modelling can be seen in Gene Regulatory Networks (GRNs). Most gene expressions are facilitated by complex networks made up of various components. These components interact in various ways and are highly interdependent. As Ideker, et al (2001) note – the transmission and expression of genetic information, thus has the following features–

- "It operates on multiple hierarchical levels of organisation.
- It is processed in complex networks.
- These information networks are typically robust, such that many single perturbations will not greatly effect them *(sic)*.
- There are key nodes in the network where perturbations may have profound effects; these offer powerful targets for the understanding and manipulation of the system."[103]

To sum up, biological systems are more complex than a mechanistic framework, that is incapable of taking dynamics into account, is capable of modelling. The hypothesis-driven research of molecular biology, while being good at isolating and studying simple pathways, is not sufficient to account for the nature of biological systems which are inherently dynamic and causally entangled. This dynamic nature is best captured by a

[102] Bechtel and Abrahamsen (2012), p.709.
[103] Ideker, et al (2001), p.345.

systemwide analysis that shifts focus away from individual pathways and accounts for (a) concentrations, (b) distributed causality, and (c) non-linearity of the interactions – features that they hold are not adequately accounted for by the traditional methods of molecular biology. How much these methods achieve their goals is yet to be fully seen. But there are certainly some promising examples that I will discuss throughout this chapter.

3.2.1: The Evolution of Systems Biology

Westerhoff and Palsson (2004) argue that systems biology has its roots in the 19th-century research on whole-cell embryology and network analysis. But as they note, it would be wrong to say that it is a paradigm shift from molecular biology. Integrative methods studying both individual entities and higher-level mapping did exist in molecular biology[104]. Instead, the availability of large amounts of omics data has facilitated a further step in this process. The rich data gathered about genomes, proteomes, transcriptomics, and metabolomics – has made it possible to do data-rich science – where, with the help of computers, it is possible to do multiple analysis. Westerhoff and Kelly (2007) highlight how advancement in genomics lead to a better understanding of genome-wide expressions[105].

Furthermore, high-throughput sequencing makes it possible to study when and where these genes are expressed in the form of proteins. This provides a huge library of multi-faceted information about cellular structures and functions. As Conesa and Mortazavi

[104] Westerhoff and Palsson (2004), p.1249

[105] Genome-wide expressions studies which genes are expressed in particular situations – cells, environments, and even diseases.

(2014) note – the rise of systems biology can be found in large scale analysis of the properties of genomes – which aimed at gaining functional data of the genomes. This similar approach was then used to gain the functional data of other biomolecules like protein (proteomics) and metabolites (metabolomics). These developments changed biology to form a "data-poor into a data-intensive field" and "motivated the development of novel computational, machine-learning and other quantitative methods for genomic analysis that attracted a large number of engineers, physicists, and mathematicians into biology"[106].

But what is the relationship between data and systems biology? The advancements in genomic data accumulation have opened the avenues for analyses. Conesa and Mortazavi (2014) note that systems biology works as an analytical tool for data. The data generated by high throughput techniques needs quantitative methods for analysis. But many bio-informatic techniques that are capable of such quantitative analysis[107] are not sufficient to tell us about the systems. Methods of systems biology are helpful here. Conesa and Mortazavi (2014) provide a nominal definition of systems biology as –

> "… the study of a given biological system by (a) the perturbation of a property of that system, (b) the measurement of resulting gene, protein, and pathway responses, (c) the integration of these data, and (d) the ultimate modelling of these data to describe the system as well as its response to perturbation."[108]

[106] Conesa and Mortazavi (2014), p.2.
[107] For example, mapping of reads that assign a genomic location and expression level of the sequence that has been read. During data creation, the position of the sequences is not read. Instead, it must be mapped onto the genome using analysis like BLAST.
[108] From Ideker, Galitski, and Hood (2001).

As Conesa and Mortazavi (2014) note, the functional genomics data is used by various methods of analysis to model biological systems. Functional genomic data is a collection of data gathered from large-scale gene expression measurement methods. These are global expression data of thousands of molecular features, without any "a priori importance"[109] assigned to any set of data.

Nonetheless, the methods of systems biology are more complex than what Conesa and Mortazavi (2014) propose. As Westerhoff and Palsson (2004) note, the history of systems biology is an amalgamation of two strands of biological research – that of traditional molecular biology, which aim at studying individual macromolecules like single gene regulatory networks; and a system theoretic approach that aimed at studying more general systemic properties like self-organisation and non-equilibrium dynamics. The molecular biology strand developed into data-rich omics research aimed at studying these macromolecules simultaneously, and the systems theory became prominent as the need for theories and models in this data-rich field became necessary. The system-theoretic root, due to its engagement in large scale analysis of functional units and organisational principles has been useful in guiding the study of molecular pathways when considered to be functioning in tandem with each other.

Similarly, Huang (2004) distinguishes between what he calls the 'localist' and 'globalist' branches of systems biology. While the localists focus on pathway centric models that are aimed at the discovery of new pathway behaviour or underlying mechanisms, the globalist view focuses on system-wide effects and finding generic design features. The localist view is an extension of molecular biology but accounts for features that have been neglected in traditional molecular biology like cross-talk

[109] Although see Leonelli (2014, 2019).

between pathways. These studies have been mostly successful due to high throughput data that help build computational models that can account for these behaviours. The globalists, on the other hand, have their roots in systems thinking. They are attracted to understanding how system-wide effects are generated and how properties emerge from lower-level organisations. They tend to study network architecture and topological properties of the system. They are interested in understanding the generic features of biological organisations.

O'Malley and Dupré (2005) divide systems biology into *pragmatic* and *systems theorist* corresponding to Huang's localists and globalists, respectively. They identify the pragmatic approach with bottom-up methodology – that of starting with molecular details and builds up to higher-level functions, while system theorists start from top-down – from higher-level generalities to more detailed models.

From these analyses, two aims of systems biology become clear –

1. To understand the complex nature of biological mechanisms using computational models combined with high throughput data, such that the previously uncaptured features of molecular pathways – like protein-protein interaction, crosstalk, and dynamicity can be explained.
2. To give general principles for biological organisations based on system theoretic tools like network analysis and complex system studies.

As O'Malley and Dupré (2005) note, the reconciliation and integration of these two approaches have been difficult. In what follows, my attention would be focused on the former strand of systems biology, albeit I believe that the delineation is not that straightforward. Localists might use the concept of generic architecture that explains why a certain pathway shows some specific qualities. Globalists, like Huang, on the

other hand, may apply system theoretic principles to study pathways and how they are situated in the broader cellular context. Thus, their methodology for both goals can transcend the differences. As a field, the current state of systems biology is to construct mathematical/computational models to explain features of the target system previously not captured by traditional molecular biology. For this, they combine their models with high throughput data to draw insights into the dynamics and properties of the system. As O'Malley and Dupré (2005) put it "Both systems biologies [pragmatic and system theoretic*]* are currently less about systems (in a theoretical sense) than about aspirations towards systematic and thoroughgoing approaches to the phenomena of interest. The field could, therefore, be described as an epistemological commitment to a general approach that foregrounds mathematical modelling in order to capture system dynamics and transcend piecemeal analyses of interconnected biochemical processes"[110].

3.3: Models in Systems Biology

In this section, I consider three approaches to modelling in systems biology –

1. Network Motifs Analysis
2. Large Scale Network Analysis
3. Attractor Landscape Analysis

I aim to provide a brief overview of these modelling methods and cite examples which will be used later on to facilitate a better understanding and analysis of models of systems biology.

[110] O'Malley and Dupré (2005), p.1273.

3.3.1: Network Motifs

Network analysis as a study gained attention following the observations that many networks across various fields share the feature of being scale-free. Scale-free networks (in contrast with random networks – that have roughly the same number of links between the nodes) have relatively fewer connection between most nodes while a few nodes – called 'hubs' are densely connected (see *figure 3.2*). World wide web, for example, typifies a scale-free network.

For biological systems, in a network (for example – a gene regulatory network, see *figure 3.3*) each node represents a molecule and its concentration across the system in question (example – a cell). So, for example, a node might represent DNA polymerase and its concentration. The edges represent the causal interactions between the molecules. The result of these causal interactions is the increase or decrease in the concentrations of these molecules.

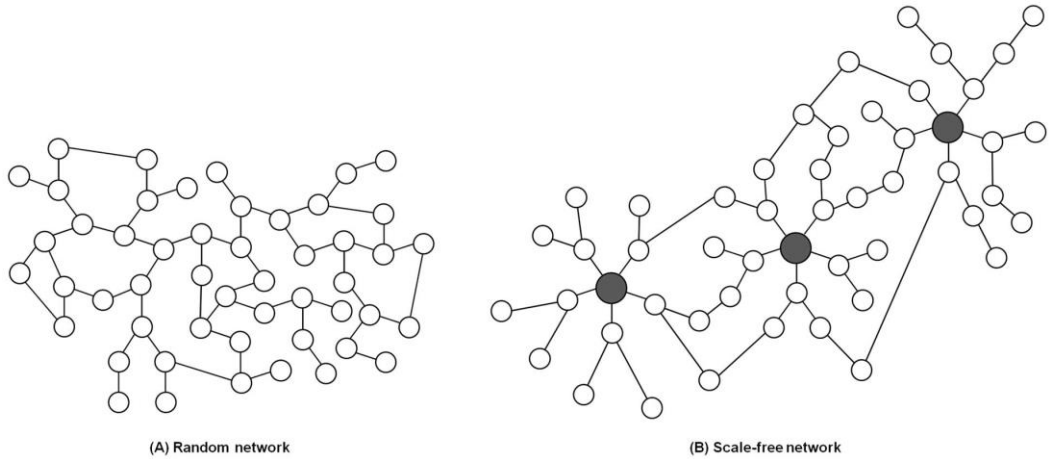

Figure 3.2 – Comparison between (A) Random and (B) Scale-free network. The scale-free network contains highly connected hub nodes and relatively scarcely connected non-hub nodes [from Seo, Kim, Lee, and Youn (2013), p.1739].

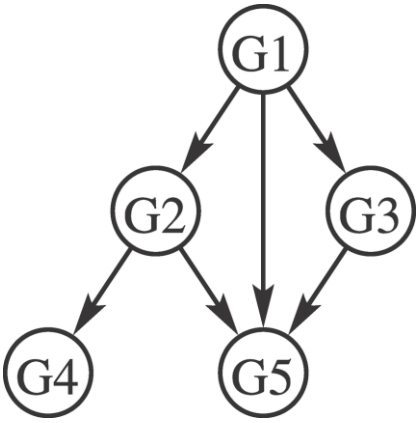

Figure 3.3 – A prototype Gene Regulatory Network (GNR). Where each node represents a molecule (in this case genes) and the arrows represent the regulatory interactions between the genes. In this example - gene G1 regulates G2, G3, and G5, gene G2 regulates G4 and G5, and gene G3 regulates G5. [From - Needham, Bradford, Bulpitt, and Westhead, (2007), e129]

Given the construction of these networks, we can then identify distinctive patterns of connected nodes. These patterns are called *motifs*. As Alon notes in his seminal work –

> "Our goal will be to define understandable patterns of connections that serve as building blocks of the network. Ideally, we would like to understand the dynamics of the entire network based on the dynamics of the individual building blocks."[111]

Alon and his collaborators [Mangan and Alon (2003); Mangan, Zaslavar, and Alon (2003); Alon (2007a); Alon (2007b)] elucidate the concept of motifs by their studies of transcription networks of unicellular organisms (especially *E. coli*). The general approach of motif identification is that of a top-down reverse modelling[112] perspective. One starts with a complete structure of the network and then identifies recurring patterns of functional importance. As Alon (2007a) notes, the hope is that these motifs will act

[111] Alon (2007a), p27.

[112] Reverse modelling starts from experimental data with the aim to seeks potential causalities through a study of correlations observed in the data. Forward modelling starts from known causal structures and studies higher level behaviours that are expected from such structures. [Gunawardena (2014)].

as building blocks for the network and its function as a whole. For example, Mangan and Alon (2003) identify a class of three-node motifs called feed-forward loops (FFLs). FFLs is composed of molecule X, which regulates molecule Y, and both X and Y together regulate Z. The regulatory action can be either positive (activation) or negative (repression) between the individual nodes. Therefore, it is possible to construct eight possible FFLs. These eight possible configurations can be divided into two subcategories according to the direction of the edges between them – coherent FFLs (where the direct effect of X on Z is same as the indirect effect of X on Z via Y); or incoherent FFLs (where the direct effect of X on Z is opposite to the indirect effect of X on Z via Y).

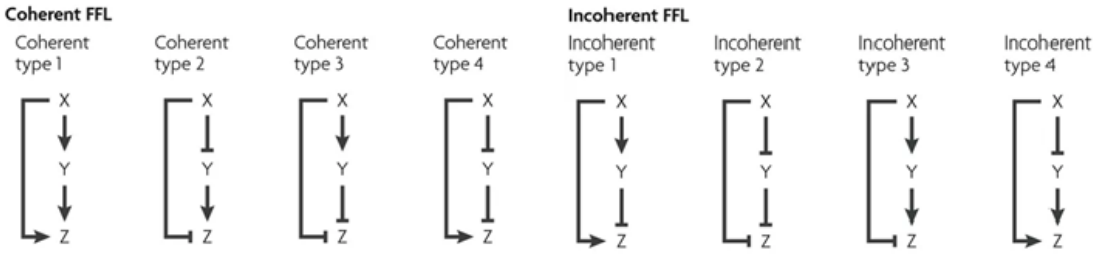

Figure 3.4 – Eight configurations of FFL – Four coherent and four incoherent. Arrows at the end of edges represent activation while dashes at the end of the edge represent repression. From Alon (2007b), p.452.

Motifs like FFLs can be used to model the regulation of Z with respect to X and Y over time using ordinary differential equations. Tyson and Novak (2010) show two different ways in which (for example) incoherent type 4 acts. When X is set as a stepwise increasing signal over time, it induces a pulse-like response in Z over time, such that Z(t) pulsates at every interval when X(t) is stepped up but returns back (or nearly back) to its original level. They call this *sniffer* – as it is commonly found in sensory systems. In the other, when X is set as a square wave (modulating between two

levels), the incoherent Type 4 FFL works as *cock and trigger*. When X(t) is at the high level, Y(t) and Z(t) remain in a low state, but when X(t) is set to a low value for enough time, Y(t) rises. At this point, if X(t) is set again to the higher value, Y(t) decreases, and Z(t) shows a pulse which then settles back to its original level. They compare it to a gun being cocked when X(t) is low and is triggered when it is set to high again.

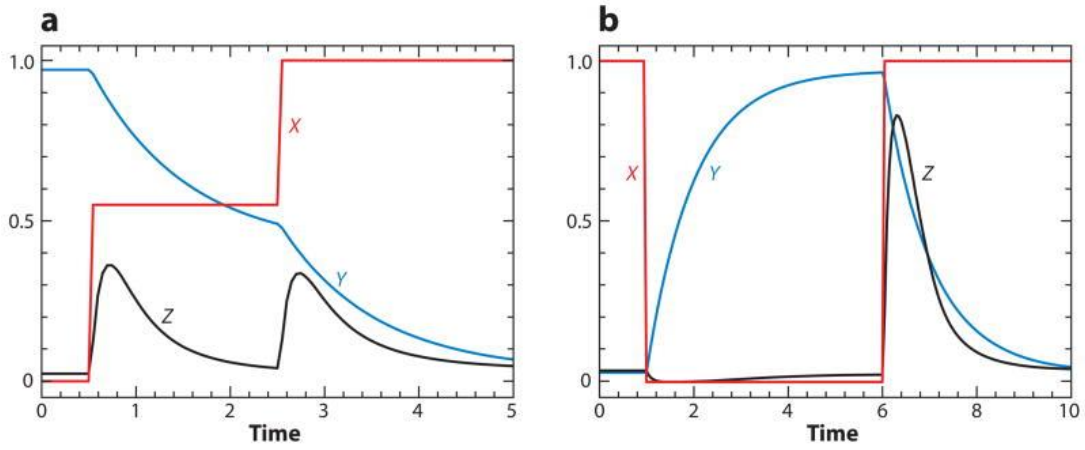

Figure 3.5 – Incoherent type 4 FFL responses. (a) sniffer and (b) trigger. [From Tyson and Novak (2010), p231]

FFLs are but one type of motifs. Many other motifs have been identified. For example, an *autoregulation motif* consists of a single node with the edge looping back to itself. It represents the capability of the molecule to regulate itself – for example, a transcription factor regulating the transcription of the same gene that it is a product of. Similarly, a *bifan motif* is a four-node motif where the two upstream nodes both regulate the two downstream nodes. Of course, the chances of motifs existing increases exponentially with the number of nodes. But as Alon (2007a, 2007b) points out, only a small number of motifs can be found in biological networks. For example, out of 13 possible motif architecture that can be constructed out of three nodes, only FFL is the

only significant motif found in the *E. coli* transcription network. Similarly, out of 199 possible four-node motifs, only two show up.

Therefore, the question arises as to why only a few network motifs are found to overpopulate biological networks. To answer this question, let us first understand how motifs are identified. The search for network motifs starts once a gene network is constructed using experimental data. In such a network, then, the frequency of the occurrence of motifs of different sizes is measured using computing methods. These occurrences are then compared to a network which is generated by randomising the network in question. The randomised network is a network which is similar to the given network in its architecture – the number of nodes and edges, but the connections between the nodes are assigned at random. Hence, the directions of the edges between the nodes are randomised[113] [114]. Significant motifs are the ones which have a much higher frequency in the real network as compared to the randomised network. Alon (2007a, b) argues that this method ensures that only biologically significant motifs are thus selected, and therefore must be of significance. But what significance? Alon (2007a) argues that since these motifs are much more frequent in real networks as compared to randomised networks, they must have been selected for the advantage that these motifs provide to the network and thus the organism they are present in –

> "… *edges in network motifs must be constantly selected* in order to survive randomisation forces. This suggests that if a network motif appears in a network much more often than in a randomised network, it must have been selected based on some advantage it gives to the organism. If the motif did not offer a selective

[113] Shen-Orr, Milo, Mangan and Alon (2002)

[114] For example, if node A points to node B, and C points to D, then during randomisation edges will be switched such that A points to D and C points to B.

> advantage, it would be washed out and occur about as often as in randomised networks."[115]

Thus, while motifs are statistically identifiable, their presence must be attributed to the selected advantage that they provide to the networks they are present in[116]. So, for example, FFLs can act as a 'sign sensitive delay element'[117]. Similarly, a negative autoregulator can reduce the variation in protein levels[118]. These functional advantages that motifs provide (when compared to topologically similar random structures); thus, according to Alon, et al., must have been selected for. As noted in Chapter 1, this is a fallible heuristic assumption. Nonetheless, as I will argue later on, the functional utility can be expressed as a present looking affair using the concept of 'design principles' [Green, Levy and Bechtel (2015)].

Having laid the foundation for what motifs are and how they function, the task in front of Alon, et al. is to explain the functioning of a network in terms of these motifs and their functions. Alon (2007a) notes that in *E. coli,* four families of motifs can be observed – autoregulation, FFLs, single-input modules (SIMs)[119], and dense overlapping regulons (DORs)[120]. These four families account for almost all the functions of the *E. coli* system.

[115] Alon (2007a), p.29.
[116] See Chapter 1 for a criticism of this assumption.
[117] Mangan, Zaslaver, and Alon (2003). Also called a persistence detector.
[118] Becskei and Serrano (2000)
[119] SIMs are motifs in which a single transcription factor X, controls multiple operons Z1, Z2…. Zn. X is usually autoregulatory. Shen-Orr, Milo, Mangan and Alon (2002).
[120] DORs are motifs where several regulators control a set of genes in combinations. In *E. coli*, DORs have been found to form up various systems like carbon utilisation, stress response, etc. Shen-Orr, Milo, Mangan and Alon (2002).

Shen-Orr, Milo, Mangan and Alon (2002) note that DORs in *E. coli* function like computational units. And are composed of the other motifs such that FFLs, auto-regulators, and SIMs are integrated inside DORs. They also observe that no DORs regulate other DORs. Therefore, noting that there is are modular sub-routines that form the whole network. As Alon (2007a) notes –

> "Overall, the rather simple way in which the network motifs are integrated makes it possible to understand the dynamics of each motif separately, even when it is embedded within larger patterns."[121]

The search for network motifs shows a novel approach to discovery in biological systems. The method proceeds by first constructing a network of the molecules by using data from experimental techniques like microarrays[122]. Then these networks are compared to randomised networks with similar architecture but different edges. From this comparison, motifs are identified that stand out. Then these motifs are studied for functions. The study of the function of motifs proceeds by analysing their internal structure. The functioning of these motifs is studied in isolation as a signalling system which maps their inputs onto their outputs. The study of network motifs does not need any molecular details of the molecules and their activities. Of course, these details can be supplemented. Edges carry signs that can correspond to molecular activities like activation, repression, etc. But other than that, motif analysis does not need any detailed knowledge of how these activities are performed, or that of the structure of the molecules that form the nodes (other than their concentrations). The study of DORs in

[121] Alon (2007a), p.90

[122] Microarrays are used to study the expressions of a large number of genes at the same time.

E. coli, further shows that it is possible to reconstruct the network using the motifs such that they form functional submodules.

The explanatory significance of motifs, thus, is derived from being able to study them in isolation and then showing how they can be used to reconstruct the system. In this respect, motifs share the strategies of mechanistic explanations – that of decomposition, such that the smaller elements that form the system can be then studied for their properties; and eventual recomposition, such that these elements and their properties can account for the functioning of the system when taken together. Though, due to the methods of statistical analysis, motif identification process expands the scope of mechanistic methodology. By scanning vast networks for motifs, provides a way to mechanistically explain the functioning of the system without the knowledge of the molecular details.

3.3.2: Large Scale Network Analysis

While motifs present a good analysis of networks through decomposition into functional units, it does not take full advantage of what network analysis has to offer. Graph-theoretic methods are used to study network architecture. The network is represented as nodes connected to each other. How these nodes are connected to each other determines the topological properties of the network. We can analyse the network architecture using different parameters. Firstly, we can determine the minimum number of edges that need to be traversed to reach from one node to another. The average of all these minimum path lengths between all the pairs of nodes is called the mean shortest path length (MSPL). MSPL determines how fast signals from one node travel to another in a given

network. The other is the clustering coefficient (C) of a node. Clustering coefficient determines the degree to which the neighbours of a given node are connected to each other. Thus, a node with a higher clustering coefficient is highly connected to the nodes in its neighbourhood. Finally, we can determine the degree of connectedness of each node in a network. The degree *k* represents the number of nodes that each node is connected to.

Using these parameters, we can analyse various types of networks. For example, Random networks (which have nearly the same number of edges per node. See *Figure 3.2*), have nearly the same clustering coefficient and degree of connectedness for each node. But not all networks show similar architecture. Small world networks, for example, are networks in which most nodes are not neighbours of one another, but the neighbours of any given node are likely to be neighbours of each other. Thus, they exhibit both small MSPL and high clustering [Watts and Strogatz (1998)]. Thus, they show a high degree of coordination throughout the network where nodes can be reached with a small number of steps. Scale-free networks exhibit an unequal distribution of edges per node, such that there are few densely connected nodes and many scarcely connected nodes, show a non-uniform clustering coefficient and node connectivity. As it turns out, scale-free networks happen to be the most prominent architects in real-world networks. As noted, before, power-grids, world wide web, and various biological networks show a scale-free architecture.

The scale-free architecture shows a high propensity for robustness [Watts and Strogatz (1998)]. Due to the topology of the graph makes it, so that intervention on most of the nodes does not alter the functioning of the network radically. This is because most nodes are not densely connected. Thus, perturbations do not seriously hamper the ability

of the network to function if they do not affect the hub nodes. As an example, let us consider a GRN consisting of many genes (as shown in *figure 3.6)*. During knockdown experiments, if we intervene on gene/gene product A, then the network is not severely impacted as there are multiple other pathways to its neighbouring nodes (A' and A"). Therefore, even if A is affected by the intervention, its neighbouring nodes are not severely affected. They can still perform their usual function. But if the hub gene/gene product B is intervened upon, many of its neighbouring nodes (example B', B", B''') will not be accessible as they do not have any other edges connecting then to the rest of the network other than the ones through B. Therefore, if a hub node is affected due to intervention, then a lot of other nodes are rendered non-functional along with it. Since hub nodes like B, are relatively few as compared to non-hub nodes, most interventions on scale-free networks will not be detrimental to the functioning of the whole network. Thus, scale-free networks show the property of robustness to a larger number of interventions, then (say) random networks.

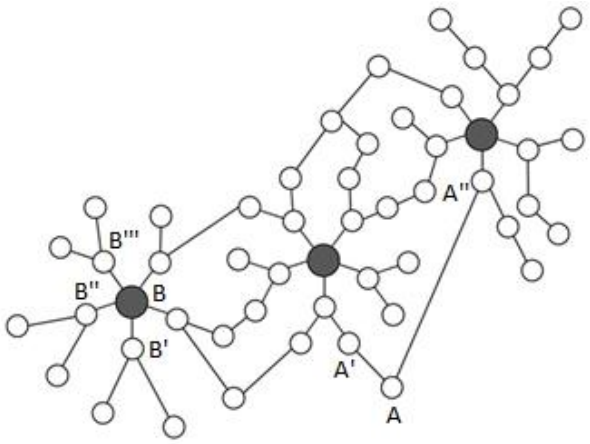

Figure 3.6 – A scale-free network. Intervention on A will only disrupt the pathway A'-A-A", while an intervention on hub node B will disrupt all the connected nodes and pathways associated with them.[adapted with modifications from Seo, Kim, Lee, and Youn (2013), p.1739].

The explanation of how intervention on differently connected nodes affect the scale-free network's functioning and how some perturbations affect the network severely while others don't, is what Huneman (2010) calls topological explanations. Topological explanations explain the phenomenon by addressing the structural organisation of the system. Huneman (2010) defines topological explanation as –

> "an explanation in which a feature, a trait, a property or an outcome X of a system S is explained by the fact that it possesses specific topological properties T_i."[123]

Where T_i are properties which "specify invariance (of S) under some continuous transformations, and which will determine equivalence classes between all structures S' homotopic to S"[124]. Or alternately, T_i defines an "equivalence class, and distinguish S from other graphs S* not having those properties"[125].

Thus, the property of robustness when explained through the scale-free network architecture counts as a topological explanation, where the property of being robust (X) of the network (S) is explained by S possessing the topological property of being scale-free (T_i). This topological property T_i will not be shared by networks which show a different architecture (for example a random network).

Huneman (2010) further notes how topological explanations are different from mechanistic explanations. While the explanandum of mechanistic explanations is the causal relations between activities and entities, the explanandum of topological explanations is the organisational structure without tracking the activities and entities and their causal relations. So, for example, while mechanistic explanations explain

[123] Huneman (2010), p.216.
[124] Huneman (2010), p.216.
[125] Huneman (2010), p.216.

robustness by appealing to the stability of relations between entities and activities as invariant under intervention, topological explanations explain them as a result of the network belonging to a class of topological kind. He writes –

> "Suppose that two systems S_1 and S_2 have the same associated shape S' in abstract space but that the relata of links in S_1 are J_1 and K_1, and in S_2 are J_2 and K_2, with their distinct associated activities. Yet the topological properties of S', realised identically by S_1 and S_2, will not be affected by this difference of activities and properties."[126]

Robustness is one amongst many properties that can be explained using topological explanations. Evolvability, for example, is another feature of a system that can be explained. In their paper, Isalan, et al. (2008) studied the effects of adding new links to *E. coli* GRN to explore the network's evolvability due to events like gene duplication. They constructed 598 strains of *E. coli* carrying artificial gene constructs consisting of different transcription or σ-factor genes (genes which code for transcription initiation proteins), thus, creating new pathways in the network and altering existing pathways. They expected that the radical changes in topology might not be tolerated by the bacteria. But they found that ~95% of the new networks were tolerated, and some networks survived selection pressures better than the wild variety. They conclude that new links in networks are rarely a barrier for evolvability and many times can confer fitness advantages.

But most importantly for our purposes, when analysing the effect of such changes on hub genes, they note - "…hub genes could have been less resilient than less-connected

[126] Huneman (2010), p.223

genes, but the bacteria can compensate. Therefore, at least when it comes to altering regulatory inputs, the hub genes do not appear to be the Achilles' heel of the network"[127]. They conclude "bacteria can both tolerate and exploit radical changes in their circuitry"[128].

This study shows that –

1. Networks can respond 'globally' to perturbations to dampen their effects and to survive to radical changes. Thus, robustness as a property of the network, cannot be defined by studying a sub-part of it.
2. Evolvability of a system depends on the response of the whole network to perturbations. Network growth can change the topology of the whole network and can sometimes be the drivers for evolution.

Example 1: Bowtie Structure of Human Immune System

Kitano and Oda (2006) explain the vulnerability of human immune system when attacked by human immunodeficiency virus (HIV) in terms of the network architecture. They note that the HIV binds with the CD4 receptor sites, eventually depleting the CD4+ T-cells, which eventually leads to immune system failure in later stages of HIV. Kitano and Oda (2006) want to address the failure of the immune system as a response to the attack on CD4+ T-cells. They suggest that the reason for this vulnerability "are deeply embedded within the architectural features of the immune system"[129]. They observe that the network shows a bowtie like structure (*figure 3.7*). The CD4+ T-cell

[127] Isalan, et al (2008), p.840.
[128] Isalan, et al (2008), p.844.
[129] Kitano and Oda (2006), p.6.

forms the central node of the structure such that various pathways converge on CD4+ T-cell and various pathways are activated by the CD4+ T-cell. Thus, an attack on CD4+ T-cell disrupts all these pathways, hence leaving the whole immune system deficient. Kitano and Oda (2006) write –

> "the bow-tie network in the immune system at both intracellular and intercellular levels is fragile against attacks on non-redundant elements within its core. Removal of such an element results in immunodeficiency."[130]

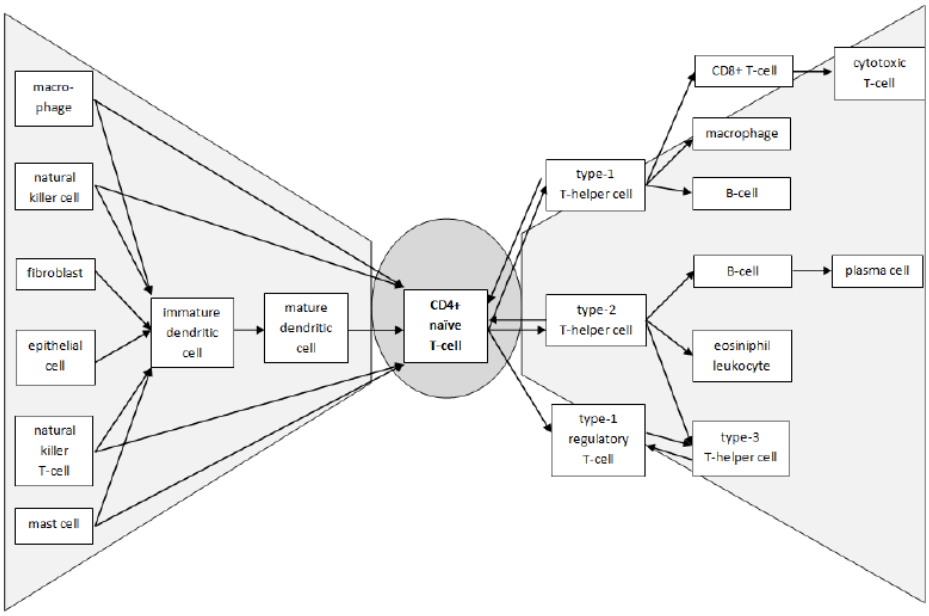

Figure 3.7 – A simplified version of the Bow-tie architecture of the immune system. [From Jones (2014), p.1138.]

The bowtie structure explains the vulnerability of the immune response owing to its network structure [Jones (2014)]. The CD4+ T-cell node forms a hub which, when

[130] Kitano and Oda (2006), p.2. From Jones (2014), p.1139.

disrupted has the potential to disrupt the whole network. Therefore the vulnerability of hub nodes of scale-free networks can be observed with the bowtie architecture of the immune system.

Thus, topological explanations can be used to address the global properties of a network. Unlike network motifs, which decompose the network into sub-modules and analyse them for functions, large scale network models proceed by analysing the whole network for properties that are dependent on the architecture of the network. Huneman (2010) maintains that topological explanations are not mechanistic because they do not model entities and activities, but only the structural organisation.

3.3.3: Attractor Landscape

The attractor model, which forms a part of Haung's globalists approach (see *section 3.2.2*), aims to capture high order properties of biological systems as opposed to localist approaches like Network Motifs. In his 2004 paper, Huang notes that localist approaches are unsatisfactory as they are "rooted in classical molecular biology"[131] and are "shaped by decades of devotion to the study of individual cellular pathways that represent to them linear causal relationships"[132]. He notes that a proper understanding of gene function "requires that one reaches beyond the narrow focus on individual pathways"[133].

[131] Huang (2004), p.281.
[132] Huang (2004), p.281.
[133] Huang (2004), p.281.

Huang's motivation for a globalist approach is rooted in the observation that wholes (like cells) tend to show a systematic behaviour that is much less complex and robust than the individual pathways that form them. He writes –

> "Cells in multicellular organisms exhibit a simple, coherent whole-cell behaviour which may precisely reflect a higher-order dynamics of the global network: the switching between cell fates. This strictly regulated, rule-based systems behaviour is robust and remarkably simple compared with the complexity of the underlying molecular network."[134]

In Huang's view, therefore, one must study the dynamic behaviour of the whole cell. This approach of "cell fate" is exemplified by studying the behaviour of the cell-based on dynamic interactions of the components instead of topological properties.

In Huang's model, a network is represented as a landscape where the cell state at any moment is represented by a state vector - S(t). The landscape itself is the representation of the states of all the individual components of the cell. The landscape is projected onto a two-dimensional plane with the third dimension depicting the energy[135] of the vector adopting a particular state. The attractors are the low-lying areas that are relatively stable and are most likely for the system to adopt. In mathematical terms, attractors are the solutions to set of equations of the system dynamics. In thermodynamic terms, they represent the equilibrium towards which the dynamic system will tend to move once it is in the 'basin of an attractor'[136].

[134] Huang (2004), p.291.
[135] In thermodynamic terms, the energy represents the free energy of the system in different states.
[136] Basin of an attractor can be defined as the initial conditions from which a system tends to move towards an equilibrium condition (attractor). Enver, et al (2009), p.388

An attractor thus, represents the stable state that a dynamic system will eventually adopt once it is in the basin of the attractor. Any given state space can have many attractors – i.e. stable states that represent equilibrium. As Enver, et al. (2009) note, with certain modifications (like overexpressing a gene, modifying interaction strengths, or changing external signals) the system state can be 'lifted up' from one attractor and moved to another (*figure 3.8*). Furthermore, it is not necessary that there be only a single pathway from one attractor state to the other, there can be multiple (in *figure 3.8* red and green arrows represent two different pathways from one attractor to the other).

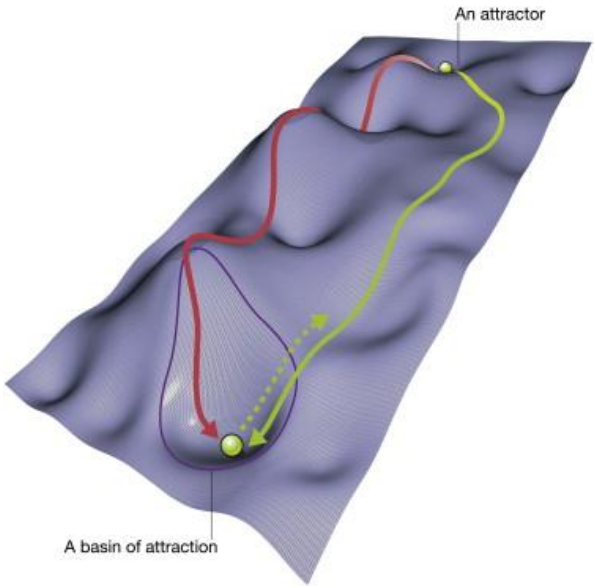

Figure 3.8 – Example of n landscape in cell differentiation. We can imagine the state vector as a ball rolling on the landscape. The attractors are represented as depressions. The lines represent the path of the state vector over the landscape. Red and Green solid lines represent two different pathways from the initial state to the final state. Green dotted line represents the pathway of how the cell can possibly leave the basin of the attractor. From Enver, et al. (2009), p.389.

What makes the attractor approach useful is that it can define various features of the system. For example, an attractor state can represent a particular state corresponding to a set of gene expressions. Similarly, the basin of the attractor can account for the robustness of the system in that attractor state as the larger the area of the basin, the

more difficult it will be for the system to move from out of the attractor state; thus making it less vulnerable to perturbations and noise. At the same time, it can explain how perturbations, external signals, and noise can push a system out of a particular attractor state, thus explaining plasticity. As a result, attractor models and cell fate models have seen widespread use in explaining cell differentiation, stem cell states, and modelling of cancer cell trajectories in recent years.

As a way of fleshing out landscape model, I take two examples – one of neutrophil differentiation and the other of blood cell development from common myeloid precursor (CMP) cell. These examples show how landscape models can be used to explain cell differentiation and cell fate decisions.

Example 2: Neutrophils and cell differentiation

Neutrophils are a kind of white blood cells. Huang, et al (2005) study how triggering human promyelocytic HL60 cells *in vitro* in various means can lead to the formation of neutrophils. To achieve this, they induced the differentiation of promyelocytic cells by using two stimuli - the solvent dimethylsulfoxide (DMSO) and the hormone all-trans-retinoic acid (atRA). Both these stimuli have been observed to trigger the same cell fate – that of triggering HL60 cell to develop into neutrophils. Huang, et al (2005) used microarrays to study the genome-wide mRNA levels over a time period. In the landscape model perspective, we can imagine the initial state of the HL60 cell as that of a being in a stable attractor. With the introduction of the two different stimuli, the state vector S(t) can then be imagined as climbing up the valley and landing in the basin of attractor that is the stable state of neutrophil. Huang, et al (2005) assume that since

the action of the two stimuli is widely different, they are unlikely to target the same set of genes. In such a case, they note, the initial trajectory taken by the HL60 cells introduced with the two different stimuli will be different. But since both stimuli lead to the formation of neutrophils, these trajectories must converge eventually. They note–

> "The convergence of trajectories from different directions across a large number of gene dimensions is a necessary condition for a high dimensional attractor state and cannot be easily explained by the existing notion of a specific, unique ''differentiation pathway'' as the common target of the two drugs."[137]

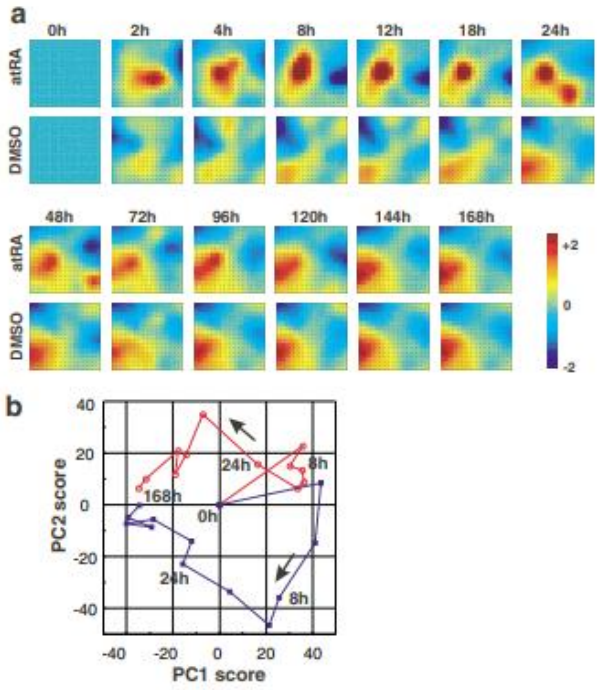

Figure 3.9 – Neutrophil differentiation – (a) The self-organising maps for atRa and DMSO induced neutrophil development at different time stamp showing an initial divergence followed by convergence. (b) The state-space vectors for atRa and DMSO induced neutrophil development at different time stamps. From Huang, et al. (2005).

[137] Huang, et al (2005), 1-2

The microarray results for both the stimuli were studies using self-organising maps over seven days. The results showed that after initial divergence, the two microarray data sets eventually started converging until they started showing a nearly similar pattern (*figure 3.9*).

The neutrophil cell differentiation study elucidates two important points –

1. Neutrophil differentiation can occur by different mechanisms. Hence there is multiple realizability in neutrophil differentiation
2. As Huang, et al. (2005) note, cellular differentiation is not best understood by differential pathways but by understanding the push of the system towards acquiring a dynamic equilibrium (a stable attractor state).

Example 3: CMP and cell fate

In blood cell differentiation, it has been found that common myeloid precursor (CMP) cell acts as an intermediate branching point between two different pathways. It either continues into erythroid/megakaryocy lineage or shifts to myelomonocytic lineage depending on the transcription factors - GATA1 or PU.1, respectively (see *figure 3.10*). Furthermore, it has been found that GATA1 and PU.1 suppress the expression of each other.

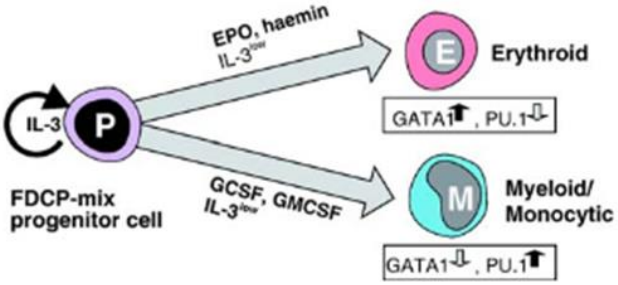

Figure 3.10 – A model for CMP differentiation. When GATA is high, CMP continues on the Erythroid lineage. When PU.1 is high, CMP continues on the Myeloid/Monocytic lineage. From Huang, et al. (2007), p.696

The CMP cell differentiation forms what is called a binary switch, such that the system can switch between two states from a metastable state depending upon various factors like gene expression, transcriptome concentrations, and external stimuli.

The binary switch, thus, acts as a cell fate switch where the cell exists in a metastable state till a stimulus is applied. At which stage, it switches to one fate or the other. In the case of blood cell differentiation, CMP can be considered as a metastable state that continues when the levels of GATA and PU.1 are relatively stable, thus suppressing each other. But when one increases, it may facilitate the 'dropping' of the cell into one of the stable states. Huang, et al. (2007) study how this cell fate decision is made.

They first theorise the structure of the network that may facilitate this binary switch. In the process, they consider two architectures – a simple mutual inhibition network for GATA and PU.1, and a mutual inhibition network coupled with autoregulation (see *figure 3.12*).

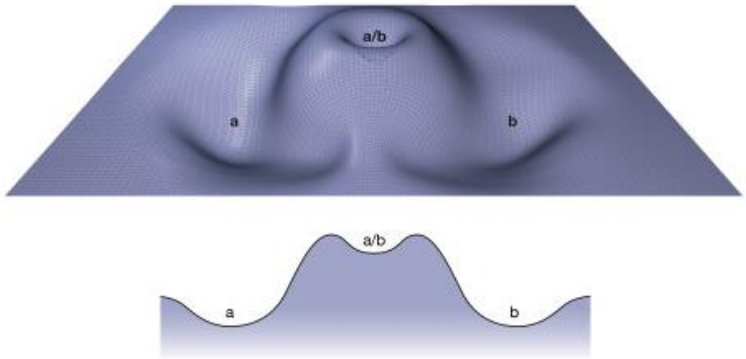

Figure 3.11 – A binary switch in n dimension landscape. (a) and (b) represent stable attractor states corresponding to different cell states, while (a/b) represent a meta-stable attractor. From Enver, et al. (2009), p.390.

The topology is then formalised with the following differential equation –

$$\frac{dx_1}{dt} = a_1 \frac{x_1^n}{\theta_{a_1}^n + x_1^n} + b_1 \frac{\theta_{b_1}^n}{O_{b_1}^n + x_1^n} - k_1 x_1$$

$$\frac{dx_2}{dt} = a_2 \frac{x_2^n}{\theta_{a_2}^n + x_2^n} + b_1 \frac{O_{b_2}^n}{\theta_{b_2}^n + x_1^n} - k_1 x_2$$

(Where x_1 and x_2 are GATA1 and PU.1 concentration, respectively; a_1 and a_2 are relative strengths of self-induction of GATA1 and PU.1, respectively; b_1 and b_2 capture the relative strength of cross inhibition; and k_1 and k_2 represent the rate of the first order deactivation. The parameters θ represent the strength of the regulatory interaction, and n is the Hill's coefficient.)[138]

[138] Huang et al (2007), p689.

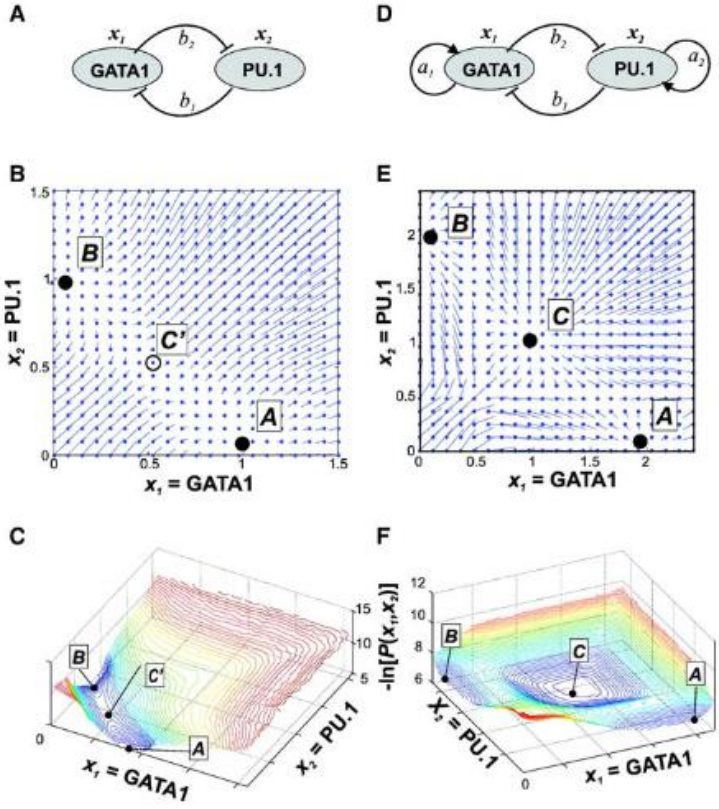

Figure 3.12 – Two possible topologies for CMP cell differentiation. A, B, and C correspond to topology showing only mutual inhibition, its state-space vector diagram, and its landscape. D, E, and F correspond to topology showing mutual inhibition clubbed with autoregulation, its state-space diagram, and its landscape. In the first case (only mutual inhibition) the metastable attractor (C') is absent. In the second case (both mutual inhibition and autoregulation) the meta-stable state C is well defined. From Huang, et al. (2007), p.697.

The first term in the equations represents the contribution of autoregulation to the change in concentration of GATA1 and PU.1, respectively. The second term represents the contribution of cross inhibition, and the third represents the unregulated decay. In the first topology (without the autoregulation) a_1 and a_2 can be set as zero to remove the influence of autoregulation and thus the first term disappears.

Solving the equation, they notice that in the second case (with autoregulation), a metastable attractor state is present that corresponds well with the CMP state, which is

absent in the network without the autoregulation. Thus, establishing a possible network topology for the binary switch as mutual inhibition with autoregulation [(D) in *figure 3.12*].

Next, Huang, et al (2007) investigate how the system might leave the metastable state and settle into either the erythroid or the myelomonocytic lineage. They envision two possibilities, both involving a change in the landscape (bifurcation) –

1. *Instructive* – An asymmetric parameter change that enlarges one of the basins of the attractor on either erythroid or myelomonocyte side. This can be achieved by decreasing a_1/a_2 or increasing k_1/k_2 to deplete the GATA1/PU.1. Thus, facilitating the drop from CMP metastable state to the attractor state with a larger basin. It is called instructive because it changes the landscape in favour of one of the two states.
2. *Stochastic* – A symmetric parameter change that changes the landscape topology symmetrically for both the attractors. This can be achieved by changing x_1 and x_2 simultaneously to disturb the meta-stable state. The behaviour can be envisioned as raising of the hill of the metastable state such that the system will fall into either of the attractor states due to a random symmetry breaking event – like fluctuations in transcription factors, or due to lineage bias of the cell.

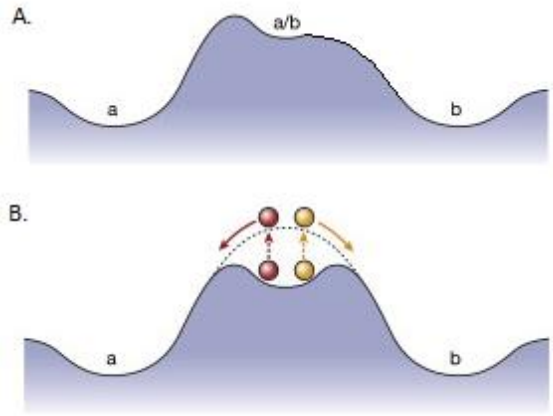

Figure 3.13 – (A)Instructive and (B) Stochastic bifurcation. In the case (A) the basin of the attractor enlarges (in this case b) to facilitate the drop from the meta-stable state (a/b) to b. In the case (B), the meta-stable state converts to hill-top thus, drop to either stable states a or b. Adapted (with changes) from Enver, et al. (2009), p.390.

Huang, et al. (2007) ran models for both instructive and stochastic cases – by introducing asymmetric signal and symmetric signal respectively. They compared the results to microarray experiment data for real cells. In the symmetric case, they found a curve in the simulation case that matched closely with the experimental microarray data. Thus, they note that the model of stochastic signalling matches with experimental data. Hence, ruling out instructive signalling.

What the binary cell fate decision model shows is that it is possible to model cell fates without knowing the exact biochemical mechanism underlying it. These models can also imply the underlying network topology that might be present, but Huang, et al. (2007) also caution against the idea of deterministic network interference. As they note, these inferences are drawn from *in vivo* experimentations using microarray data. Thus, instead of implying the structure of a singular network motif that might be present, they instead rely on the whole-cell dynamics to draw their conclusions. It is thus, challenging

to imply the underlying mechanism or network structure without further experimentation.

3.4: Explaining with Mathematical Models

Having considered examples pertaining to modelling in systems biology two questions arise about the significance of these models (in biology) –

1. What do mathematical/computational models explain?
2. How do mathematical/computational models explain?

In this section, I will explore the answer to these questions. While mathematical models have been explored concerning their formal structure, the major focus in biology has been on how mathematical models compliment mechanistic models. Therefore, instead of focusing directly on mathematical models, I will start with the question of how mechanistic explanations defines, uses, and explains mathematical models, and then move on to the two questions posed above.

The structure of the section will be as follows. First, I explore how mechanistically minded philosophers have used mathematical models to argue for a further understanding of mechanisms and their organisational qualities. Following this, I argue that the underlying idea behind these arguments is that mathematical models provide dependency relations that hold for a class of systems.

3.4.1: Mathematical Models and Mechanistic Explanations

Most mechanistically minded philosophers who have focused on mathematical modelling, have done so within the aim of providing comprehensive mechanistic explanations. Mechanistic models, as discussed in the last chapter, decompose the system into parts and activities to explain the phenomenon of interest. The decomposition helps modellers determine the causal contributions of different components and how they are organised. The account thus provides an explanation of how a phenomenon is produced by explaining how the different components contribute causally to it. As has been discussed, Woodward's interventionist account provides a theory of causation to most of these accounts [Craver (2007), Glennan (2002)]. The interventionist approach shows how component X causes Y when we can establish that the relationship between them can be generalized under some conditions. All these generalizations taken together (G_i), explain all the dependency relationships between all the causally relevant factors. Therefore, mechanisms explain by showing how X's causal influence in the form of activity φ contributes to the phenomenon Ψ.

But as has been noted in the last chapter, the biological systems can show varying degrees of dynamic behaviour. As a result, many philosophers [Levy and Bechtel (2013); Brigandt (2015); Green, Levy and Bechtel (2015); Brigandt, Green and O'Malley (2016)] have argued that the dynamicity and organisational qualities of mechanisms are best captured using mathematical and computational models. Others [Green, et al. (2018)] have argued that mathematical/computational methods can be used to help in mechanism discovery. In what follows, I first discuss the use of mathematical/computational models as a strategy in mechanism discovery, which will

be followed by analysing their use in elucidating organisation and dynamicity in mechanisms.

3.4.1.1: Discovering Mechanisms through Mathematical Models

One of the utilities of mathematical models as pointed out by several authors is about how they can be used as heuristic strategies to aid discovery of previously unidentified mechanisms, or mechanistic components and activities. Green, Levy and Bechtel (2015) for example, cite the example of a study by Eldar, et al (2002) in which the gradient of bone morphogenic protein (BMP) was studied. BMP gradient is found to be important for early pattern formation in *Drosophila*. The study aimed to explain why the gradient was robust to changes in various components of the system. Eldar, et al. (2002) constructed various computational models representing the possible architecture of the system that can explain the robust BMP gradient. By running simulations, they were able to eliminate the models that did not show the required robustness. The models that were left showed some similar features. They showed a design which correlated well with a "shuttling-degradation mechanism"[139], which was further confirmed by experiments.

The modelling exercise by Eldar, et al (2002) shows that computational models can limit the possible solution space for mechanistic discovery by eliminating possible designs that do not fit well with the observed behaviour of the system. By doing so, these models can work as a heuristic strategy to find out the underlying mechanism and their component parts. This is especially useful when decomposition is difficult to

[139] Green, Levy and Bechtel (2015), p.26.

achieve and the contribution of each component difficult to be ascertained individually. In the case of robustness, the application of mutual manipulation criteria becomes difficult. In such cases, computational models can limit the possibilities by modelling robustness into the computational framework, thus optimising the discovery process.

Green, Levy and Bechtel (2015) call these *design principles*. Design principles work by elucidating that given a specific behaviour, only specific designs can be expected where designs are organisational patterns that, by virtue of their function, bring about the desired behaviour. Green, Levy and Bechtel (2015) further define design principles as explicating the role-function – i.e. the contribution it makes to the functioning of the system. They contrast it with the selected effect function – that of being selected for a specific effect. They note that while these two functions can be thought of being dependent on each other – that the role function is the effect for which the selection took place, the way in which design explanations proceed is not to elucidate the selection history. For example, the design of the *persistence detector* is to reduce the effects of noise. This is the role function of the system. The design of the persistence detector – the coherent FFL explains this role function. This might also be the effect that it was selected for – that to reduce noise in the system, but it is not necessarily true. The persistence detector might have been selected for other reasons that it no longer serves, or it might not have a selective history conducive to environmental reasons. But to be able to draw that conclusion, we must be able to study the selective history of the system that the persistence detector is a part of.

Nonetheless, in the present context, the persistence detector, due to its design, has a specific role function – that of reducing the effects of noise in the system. This, they call the *thin notion of design*. The thin notion of design, as they note, is a *present-looking*

affair – i.e. how the design contributes to the overall functioning of the system here and now[140]. This also counteracts the heuristic assumptions of motifs as evolved units. Since design in the thin sense is a present looking affair, we do not need to ascribe their presence to evolutionary forces. Instead we are pointing out that these motifs appear in the system and can have functions ascribed to them due to their structure. The question of 'if they were selected for' can thus be circumvented.

Another aspect of the discovery of design is that of optimisation. Design is optimised for the function as compared to other possible design. That is, given a design (D_1) explains a function (F) better than design (D_2), where D_1 and D_2 are both possible designs capable of realising F. In the Eldar et al. (2002) study, the design is optimised for robustness. Similarly, designs can be optimised for other features. Optimising assumptions provide a way of narrowing down the possible models which can then be studied and confirmed using further experimentation like knockout and knockdown experiments. The optimising assumption might prove to be correct or incorrect based on experiments. Nonetheless, these assumptions are essential as computational limitations make it difficult for each model to be considered.

In conclusion, mathematical/computational models can help in the discovery of previously unknown mechanistic details that bring about an observed behaviour by modelling the behaviour and optimising it for certain aspects. By doing this, it limits the possible space for realisation, which can then be explored further using traditional experimental techniques.

[140] Green, Levy and Bechtel (2015), p.24.

3.4.1.2: Modelling Dynamic Behaviour

Another aspect of mathematical/computational model with regards to mechanistic models is that of explaining how the mechanism in question is organised such that a particular behaviour can be realised. As discussed in already, mechanisms can show dynamic behaviour that is difficult to capture within a purely mechanistic framework.

Brigandt (2015) argues that mathematical models are indispensable for some kinds of explanations in biology, especially where quantitative analysis and temporal dynamics are needed to be accounted for. This is a view held by many mechanistically minded philosophers like [Bechtel and Abrahamsen (2010, 2013), Levy and Bechtel (2013). Brigandt maintains while mechanistic explanations remain a cornerstone for elucidating causal features of the system, mathematical models can explain what counts as explanatorily relevant to the model in question. Brigandt defines explanatory relevance as –

> "(ER) A component of an account representing causal features (including a mathematical equation) is explanatorily relevant, if omitting it or changing it results in an account from which the particular explanandum does not follow any longer. Features that are not explanatorily relevant for the explanandum at hand (and the criteria of explanatory adequacy) are to be excluded from the explanation."[141]

Brigandt (2015) thus, argues that the ER determines what is relevant to the explanation. The criterion thus defines an epistemic guide to gauge the relevance of various variables for an explanation. Unlike Craver (2007), this is not an ontic criterion about what

[141] Brigandt (2015), p.143.

constitutes an explanation of Craver's ontic adequacy criteria. Therefore, as per the explanatory requirements, the model can exclude features of the system as per ER. The inclusion of the ER provides a novel way of explaining a mechanistic phenomenon.

For example, Levy and Bechtel (2013) argue that to understand the behaviour of the mechanism, it is essential to abstract away from the causal details and study the mechanism's organisation using tools like graph theory. They write –

> "Our contention is that it is often the connectivity, treated abstractly, that explains why a mechanism exhibits the particular behavior it does—especially when the behaviors in question concern non-trivial dynamic patterns."[142]

The object of study thus, becomes not the exact causal relationships between the components, but the way the organisation of these causal connections brings about a behaviour. Organisational features are essential to mechanistic explanations. The study of the structure of causal connections thus, explains the integration of components by displaying their particular interactions with each other (or a subset thereof). The persistence detector, for example, is an example of an abstract organisation that shows how the output Z is dependent on a persistent input X being present. What matters to this organisation is not the exact nature of how X brings about Y and how X and Y together bring about Z, but that X acts as an activator for Y and together X and Y act as activators for Z. Instead of assessing the causal interactions between the components, the persistence detector furnishes us with temporal details about the organisation. If X is withdrawn before Y is activated, then output Z will stay as zero because of the absence of signal from Y. If X is withdrawn just after Y is activated, the output of Z will still

[142] Levy and Bechtel (2013), p.245.

stay as zero because of the absence of X. The output of Z will be one only when X stays active for enough duration for signal X and Y both to be present. This shows the temporal character of the organisation as simple Boolean functions without any particular details about their characteristics and causal relationships other than the specification of activation and repression. Thus, abstracting away from causal details to represent simple organisational features can help explain the temporal dynamics of the system in question.

As it happens, the persistence detector like structure can be found in the bacterial arabinose metabolism system where spurious signals of cyclic adenosine monophosphate (cAMP) are observed [Levy and Bechtel (2013). In such a case, we can provide two kinds of explanations –

1. How the cAMP contributes causally to the arabinose system by explaining the causal structure of how cAMP triggers a specific transcription factor CRP through binding, or how CRP-cAMP binding activates the protein AraC. These two activities together are responsible for the activation of the arabinose operon. We can explain how the specific binding takes place such that CRP-cAMP complex is formed, or how AraC activates the operon. Thus, through our description of these components and the activities they bring about through their properties and specificity, we can explain how the arabinose system works – by combined effects of CRP-cAMP complex and AraC.
2. We can explain how the operon functions when both CRP-cAMP complex and AraC are present. But since the cAMP signals are spurious, we can further explain how the arabinose system weeds out the spurious signals and only works when cAMP is continuously present. To do this, we need not go into the

> causal details. What we need to show is that the system works like a persistence detector where the CRP-cAMP complex activates AraC and the operon needs both CRP-cAMP signal and AraC signal to be present. Therefore, when cAMP is short-lived, the operon is not activated.

This explanation does not depend on the details of how CRP-cAMP binding happens, or how CRP-cAMP activates AraC, but that these relations hold. The rest of the explanation is dependent on showing how the organisational features bring about the behaviour. The explanation thus proceeds by showing that a persistence detector configuration is realised. And therefore, a particular behaviour is to be expected given the configuration. By abstracting away from the causal details and focusing on the organisation, we can thus, explain how the system works to minimise the effect of noisy signals.

Both of these cases offer different explanations. We can also see how ER holds for these explanations. In both cases, the explanandum requires only some features of the system to be cited. In the former, causal details matter because we aim to give an exhaustive explanation of how each component interacts causally with each other (or a subset thereof). In the latter, the explanandum requires us to explain not the causal features but the organisational features that explain the temporal dynamics of the system.

Another aspect (as discussed in the previous section) is that of optimisation. Optimisation is a process of selecting the best possible model for a given set of constraints such that a function can be optimised. While we have considered the example of Drosophila pattern formation, another example might help clarify the need

for optimisation further. Itzkovitz, et al. (2012)[143] studied the intestinal crypt development by a small number of stem cells. As they note – during development, the crypt stem cells can either divide symmetrically – producing two stem cells or two somatic cells; or asymmetrically – producing one stem cell and one somatic cell. Itzkovitz, et al. (2012) modelled the replacement of epithelial cell as three distinct options –

a. An early phase of only asymmetric division that produced only somatic cells which later gave rise to more somatic cells
b. An early phase of only symmetric division that gave rise to a large population of stem cells, which later produced somatic cells through asymmetric division.
c. A mixed symmetric and asymmetric division option that produced both somatic and stem cells throughout the process.

Itzkovitz, et al. (2012) then solved the models such that the time taken for the maturation of the crypt is optimised – i.e. they find the solution where the crypt development takes the minimal possible time. Solving the mathematical model with the optimising assumption, they found that the best model was the one that showed dynamic switching between symmetric and asymmetric division – which happened to correspond with the case where the early phase only consists of symmetric division, thus increasing the population of the stem cell and later phase of purely asymmetric division. Here the model was optimised for the minimum time taken for crypt development.

The optimisation assumptions have a bearing on what the model explains. As Rice (2015) notes, the optimised model explains because –

[143] Example taken from Green, Levy and Bechtel (2015)

a. It accurately represents the constraints and trade-offs of the system in question,

b. Makes accurate optimisation assumptions.[144]

This, of course, depends on the explanatory relevance of the model. In the above example of crypt development, the explanandum was the formation of the crypt, given the constraint of time taken per cell division and the varieties in which cell can divide (symmetrically and asymmetrically). The optimisation assumption was that of minimising the time taken for crypt development. Thus, the best model – that of the early symmetric division followed by switching between symmetric and asymmetric division.

Lastly, idealization in mathematical models plays a significant role. Idealization is the intentional distortion of some aspects of the system such that even though they are false but are still explanatory. In any scientific modelling exercise, idealization plays an important and ubiquitous role[145]. Mathematical models idealize some aspects to be explanatorily relevant. For example, the persistence detector is idealized as a Boolean circuit where the functioning of Z is represented as a logical AND gate. A consequence of this is that when X and Y are active, Z is considered to be active instantaneously, even though in real networks resembling a persistence detector, this will not be the case as the causal processes that lead to output will be time taking. Moreover, the output in the idealized case will be instantaneously either 0 or 1, while in real cases it might build up to full output over some time. Similarly, when considering networks of pathways, the complexity might be reduced by idealizing the intermediary steps while considering a specific input-output response. Since the pathways are complex and many

[144] Rice (2015), p.592

[145] See Chapter 1.

intermediate components and activities take place at different rates, an idealization is introduced where the intermediate nodes are considered to be already in steady state[146]. Thus, the complexity of the intermediate steps is idealized as being in steady state while they might not be in real models. This helps to discount the intermediate steps while considering the relationship between the selected input and output nodes. In both cases considered above, the models idealize some of the features of the system by introducing falsehoods.

As an example, let us consider pattern formation in mice teeth development [Brigandt (2015), p.483]. Brigandt (2015) cites the modelling exercise undertaken by Salazar-Ciudad and Jernvall (2002), where during the model construction of how cusps are formed, the molecular details of how several regulatory networks are abstracted away from. Instead, the model relies on deriving a basic activator-inhibitor structure from the said interactions such that a simple model of the reaction-diffusion system is realised. Reaction-diffusion systems consist of the topological structure, as shown in *figure 3.14b*.

[146] Ilsley et al. (2009), p.1284.

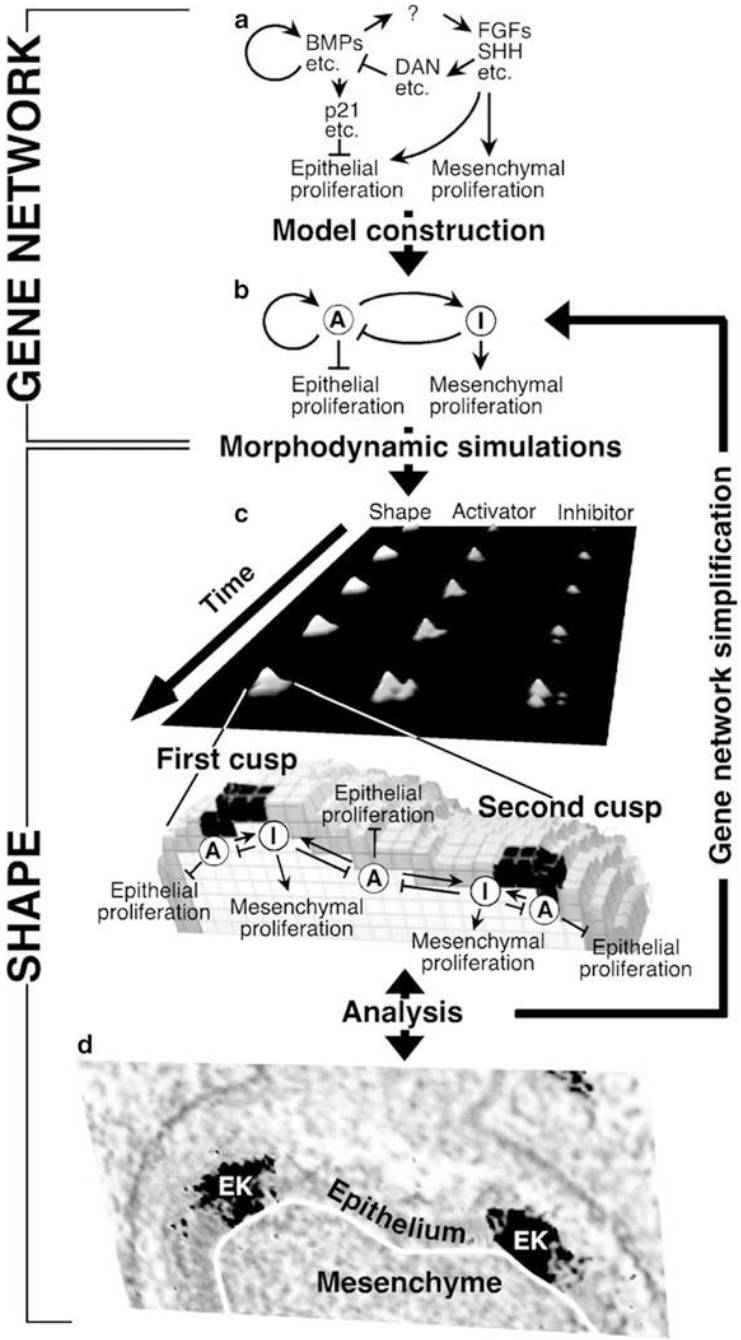

Figure 3.14 – From Brigandt (2015), p.147. (a) A mechanistic schema of cusps development in mice. (b) Network model of showing the activator-inhibitor topology with downstream effects. (c) Morphodynamic simulation of the model showing cusp development. (d) comparison of the model results with empirical data.

The activator increases its own concentration and also enhances the production of the inhibitor. In turn, the inhibitor suppresses the expression of activator in its immediate

region. This leads to the creation of stable patterns of concentrations of the related downstream components. In the mice teeth formation example, the activator has a downstream effect of inducing enamel knot differentiation while the inhibitor has the downstream effect of suppressing the enamel knots. When the model is simulated, it shows the cusp shape over time, corresponding to the activator and inhibitor distribution. The model's predictions were then confirmed using in vivo data. There are two things to note here –

First, the model abstracts away from causal molecular details in a substantial way. The gene regulatory networks, the intermediate steps between the activator-inhibitor interactions and how the activator and inhibitor both produce their downstream effects are all abstracted away from such that no causal story about how these effects were brought about is provided. Moreover, the explanatory force of the model lies and not in elucidating the causal story.

Secondly, the model is not a particular model of teeth formation in mice per se, but that of the reaction-diffusion system. That is, Salazar-Ciudad and Jernvall (2002) mathematical model is not particularly of teeth generation in mice, but a model of general reaction-diffusion systems which applies to various cases in biology[147], where parameters of the model can stand for similar dynamics between different entities of different systems that the model can be applied to.

As discussed earlier in the chapter, Issad and Malaterre (2015) argued that the steps involved in constructing and solving the mathematical model are not open to rehearsing a causal story. In the example above, for instance, the model proceeds by abstracting

[147] Examples include models of ecological invasions [Holmes, et al. (1994)] and tumour growth [Hormuth, Eldridge, et al. (2018)].

away from the mechanistic model of the cusp development such that the topological model thus constructed shows the influence of activator and inhibitor on each other without displaying the intermediate causal steps. Furthermore, the phenomenon of interest is derived by solving and simulating the mathematical equations derived from the activator-inhibitor topology. The causal story about how the activator activates or the inhibitor inhibits, or how the downstream effects are causally realised are not the focal point of the explanation. Therefore, a significant model fitting is needed to establish the corresponding causal relationships through experimental verification.

In this section, I have argued that mathematical/computational models in biology do not explain by tracking causal details; instead, they abstract away from the causal details, idealize aspects of the mechanisms, and optimise the model as per the need of the model. I agree with Levy and Bechtel (2013) that the mathematical models can elucidate organisational properties, and with Brigandt's ER condition. But both these conditions require model fitting to show that the suggested model can be used to explain the features of the system in question. The crux of the argument is that the mathematical model, while can be used for elucidating the organisational aspects of mechanism, does not derive its explanatory power from the mapping but from the structure of the model itself. Whether it fits the system to be explained requires model fitting and parameter estimation.

In what follows, I argue that mathematical/computational models explain not by tracking mechanistic details or by imposing ER considerations, but primarily by showing how certain dependencies exist between the features of the system. Furthermore, they delineate the possibilities for the system by showing how it belongs to a particular class of systems. In this process, they provide particular affordances for

the system's behaviour. Finally, I consider an account of collaborative explanations by Fagan (2015) and argue that it is best suited to characterise how mathematical models are utilised in mechanistic explanations.

3.4.2: Explaining with Mathematical Model: Dependency, Universality and Formal Constraints

Let us take the example of CMP and cell fate. The strategy followed by Huang, et al. (2007) is to construct two competing models (one without autoregulation and one with autoregulation) and consider which model fits the observed behaviour better. While the model without autoregulation does show the corresponding states, the model with autoregulation shows a more stable attractor for all the three stable states (progenitor, erythroid and myelomonocytic). They thus, along with experimental data, establish that the second model corresponds better with the observed behaviour. Furthermore, the optimal model thus established provides some possibilities for the realisation of the behaviour in the form of parameter change that alters the landscape (instructive or stochastic). But the model does not specify one to be more likely than others. To establish that the stochastic parameter change takes place, they conducted experiments that revealed the behaviour.

There are three steps to this modelling exercise –

1. Constructing various models that can explain the behaviour in question (in this case – one with autoregulation and the other without), and establishing the parameters such that the variables in the model have a physical interpretation corresponding to the real system. In this case, Huang, et al. (2007) explicate

the parameters as influences on the nodes without establishing the exact mechanism about how these influences are brought about.

2. Running simulations that show how the solutions of the models correspond with the experimental data. In this case, the two model simulations were compared for the presence of tri-stable attractor states based on their global solutions. In this process, one was eliminated. This stage can be called model fitting.
3. Predict how the model will behave in different conditions, thus providing a range of possibilities for the behaviour of interest to occur, followed by further model simulation to establish how these different realisations can be obtained. Finally, the different possible simulations are compared to experimental data to establish the actual realisation of the behaviour of the system.

For (1) and (2), the model in question is determined to be optimal by the method of model fitting the data. This can be done through various regression techniques. Furthermore, it requires parameter estimation from the available data. The technicalities of these techniques are out of the scope of this chapter. Nonetheless, the available models show specific characteristics. The model with autoregulation, for example, shows the characteristic of having two stable attractors with a meta-stable attractor. Although what these attractor states represent in the target system must be determined by fitting the model to the data, the characteristic of the attractor states can be understood by stimulating the model without any particular model fitting and parameter estimation. What the model explains is that a topology of a certain kind shows a specific behaviour. In the case of the first topology (with only the nodes mutually inhibiting each other) a landscape with only two stable attractor states is observed. While in the case of the topology with both mutual inhibition and autoregulation, a landscape with two stable

attractor states and one metastable attractor state is observed. This is not established by modelling the system with exact parameter values, but by dependencies established within the mathematical framework. Therefore, the explanation of why the bi-stable behaviour was observed can be explained by the topology and the mathematical derivation.

For (3), the two possibilities of bifurcation or state jumping such that the system switches from the meta-stable state to either of the stable states is again defined by the affordances that are defined by the mathematical model. For example, for an asymmetric bifurcation can be brought about by an increase in k_1 or decrease in a_1 such that the meta-stable attractor collapses. The physical interpretation of the processes that may lead to an increase in k_1 or decrease in a_1 is insignificant for the modelling. The conditions for asymmetrical instructive bifurcation can be derived from the equation itself. The interpretation, though it helps establish the molecular mechanism that may be underlying the target phenomenon, does not feature in the explanation as to why the bifurcation happens. It is a condition defined by the model itself.

What the CMP example shows is that the explanatory force of the model is derived from the mathematical formulations and not from studying the underlying specific details of the system. What the mathematical model thus, shows is that the system that it applies to belongs to a certain class of systems where specific dependency relationships hold. In the CMP case, the dependency relationship holds between a particular topology showing a tri-stable attractor landscape. That, this is the case for CMP is established by auxiliary steps of model fitting and parameter estimation.

Similarly, in the case of topological explanations as explicated by Huneman (2010), the scale-free architecture of a network explains its properties like robustness and

vulnerability. Hence, while explaining why the attack by HIV cells on CD4+ T-cell leads to immunodeficiency, the explanation must cite how the immune system realises a particular architecture (bowtie architecture) such that the CD4+ T-cell occupies a central hub. The explanatory burden thus lies on the topological property realised. The dependency is between architecture and robustness.

Similarly, in the case of motifs, ordinary differential equations that express the relationship between the inputs and the output are used to explain their behaviour. For example, a negative autoregulation motif works by suppressing itself. Thus, say, a protein P is negatively autoregulated. The change in concentration X of P over time can thus be described by the ODE–

$$\frac{dX}{dt} = f(x) - \delta.X \tag{1}$$

where $f(x)$ represents some dependency function, and $\delta.X$ represents the degradation rate of X. In case of negative autoregulation, the function $f(x)$ can be described in terms of the production rate of X such that it remains constant - σ till a certain threshold concentration Y is reached after which, the production rate becomes 0. Therefore

$$f(X < Y) = \sigma \tag{2}$$

$$f(X \geq Y) = 0 \tag{3}$$

Therefore, as soon as X reaches the concentration threshold Y, the production rate of X becomes zero and the concentration of the protein degenerates. But as soon as the concentration falls below Y, the production starts again. As a result, if the protein is negatively autoregulated, given enough time will settle for the threshold value Y and will be regulated to stay at this concentration.

The details of how the protein can achieve a steady-state concentration in the negative autoregulation example are not dependent on the causal analysis of how the protein is produced through the process of translation, or how it is inhibited, but on the mathematical derivation. Of course, the rate of production and degeneration, and threshold values are variables that will be different for different cases. Still, the dependency relations between these variables are necessitated by the mathematical derivation. Therefore, the mathematically necessitated structure of negative autoregulation motif explains why particular behaviour is observed in cases where such a structure is realised.

Green and Jones (2016) note that such explanations form a broad explanatory strategy that they call *constraint-based reasoning*. They formalise the structure of constraints-based explanations as follows –

1. "A claim that systems within some class **C**, differing with respect to some range of mechanistic detail, share some organizational or structural property O and some dynamic or functional property D.
2. A demonstration that formal constraints **F**, applicable to systems within **C** despite their heterogeneous mechanistic detail, define or limit the range of possible dependence relations between O and D.
3. An inference that all systems in **C** exhibit the permitted O-D dependency relations, regardless of differences of mechanistic detail among members of **C**."[148]

Where formal constraint (F) "*is a general principle, often conceptualized in mathematical terms, that characterizes a dependency relation among some of a system's*

[148] Green and Jones (2016), p.365 (emphasis in original text).

(natural or hypothesized) features in a way that limits and affords a certain set of possibilities for our models of the system"[149].

The argument for constraints-based explanations is that the dependency relationships between the structural properties and functional properties of a system can be captured by formal constraints of the model. These formal constraints form the basis of explaining the generic features of the system. Furthermore, they constrain the possibilities of the behaviour of the system. A negative feedback loop will show oscillatory output no matter how it is realised by the underlying mechanism. Scale-free network architecture will show robustness or vulnerability based on which node is intervened regardless of the underlying entity.

Batterman and Rice (2014) argue that systems that show similar macroscopic behaviour despite being realised by different microscopic details form what they call a *universality class*. This universality class also contain various model systems. The argument thus is that since the model systems and the real-world systems show similar behaviour and dependencies despite their realisers being different, therefore, the universality class can be described using model systems. Batterman and Rice (2014) dub these models that are used to explain the real-world system in the same universality class as minimal models[150]. As per Batterman and Rice (2014), minimal models explain not by the model to system mapping, but by showing how certain causal factors are irrelevant when it comes to explaining macroscopic behaviour of various systems in the same universality class.

[149] Green and Jones (2016), p.346 (italicised for emphasis).

[150] Not to be confused with Minimal model idealizations Weisberg (2007a), where minimal models are models that only account for core causal features.

An example of this is Fisher's explanation of sex ratio. R.A. Fisher's (1930) model shows that the sex ratio of many species settles for the equilibrium of 1:1, i.e. the same number of male and females. Fisher shows that it does not matter what the initial population sex ratio was, but that with enough time, natural selection will optimise a particular sex ratio based on substitution cost of producing one male instead of a female and vice versa. Since the substitution cost in both cases is the same (i.e. equal amount of resources is required a male or to produce a female), the population sex ratio will eventually settle for 1:1. As Rice (2015) notes, the model makes idealizing assumptions about many features of the actual causal process. Some of these idealizing assumptions are that the population is of infinite size, the resource distribution is equal, and that mating is random. The model explains, not the actual causal trajectory of the behaviour of the population, but that despite the initial conditions and the causal history, the system will show a particular behaviour because of the formal constraints on the system that necessitate certain dependencies.

Sober (1983), while analysing the explanatory power of Fisher's explanation, notes that –

> "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."[151]

Equilibrium explanations, of which Fisher's explanation of sex ratio is an example, explain not by showing how the initial conditions and the trajectory of the system realise

[151] Sober (1983), p.202.

the equilibrium, but by showing that given certain features of the system, equilibrium will be obtained.

As Sober (1983) notes –

> "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... Equilibrium explanations are made possible by theories that describe the dynamics of systems in certain ways."[152]

Thus, the explanatory power of equilibrium explanations does not derive from tracing the actual trajectory or the causes of why equilibrium obtained, but by the showing how systems with different microscopic details will achieve them nonetheless if they show certain dynamics. Thus, equilibrium explanations explain by showing how the system (possible or real) belongs to a universality class. This universality class shows the behaviour of being in equilibrium when specific dependencies are realised.

The dependency relationship for equilibrium can be seen by analysing systems that show equilibrium using mathematical modelling [Gross (2015)]. Any system which shows a particular production and degeneration rate can be expressed with the ODE –

$$\frac{dX}{dt} = \sigma - \delta.X \qquad (4)$$

[152] Sober (1983), p.207.

where the change in concentration of X with time is expressed as its production rate σ minus its degeneration rate $\delta.X$. When solved for the concentration of X as dependent on its initial concentration X_0 (at time t = 0) and time t, we get –

$$X(t, X0) = \left(X0 - \frac{\sigma}{\delta}\right) \exp(-\delta t) + \frac{\sigma}{\delta} \tag{5}$$

As can be seen from the equation, at time t = 0, the solution comes out to be X_0. But if we let sufficient time pass (at $t \rightarrow \infty$), the first term becomes negligible. Therefore -

$$X(t, X0) \rightarrow \frac{\sigma}{\delta} \qquad (for\ t \rightarrow \infty) \tag{6}$$

Hence when sufficient time is passed, the concentration of X does not depend on the initial concentration X_0, but on the values of σ and δ, which are constant for a given system. Thus, the concentration X is in equilibrium. Furthermore, even if the concentration is changed such that the system is not in equilibrium any longer, the concentration will eventually return to the equilibrium state. Thus, the equilibrium can be explained by dependency relations which are independent of a causal story about how the system got to the state [Gross (2015)]. The initial concentration and the trajectory followed does not feature in these dependency relations. The relation holds for any system which shows production and degeneration of a component such that its concentration is dependent on the rate at which these events take place. The formal constraints on systems that show equilibrium, despite what entities realise it, is dependent on some features (the rate of change in concentration) and not on others (the initial concentration). Equilibrium is thus, explained by the dynamic feature of the system – that it attains equilibrium as dependent on structural features of the system (that there is a constant rate of production and degradation), and not the causal details of how the system proceeds to realise these.

The dependency relations are still open to manipulation à la Woodward[153]. We can, for example, explain why the concentration at equilibrium is (say) X_E and not X_E' by showing how changing the rate of production or degeneration will change the steady-state concentration. But we have not altered the formal constraints on the system. Instead, we are explaining that the dependency relations explain why in a particular case equilibrium is obtained at a specific concentration. The model still explains why equilibrium is obtained and how it is dependent on the structural features of the system (that of showing production and generation explains the dynamic quality of equilibrium). To understand how the structure of the system determines the property of having an equilibrium, let us consider a case where the degeneration rate is set to zero. In such a case, the equilibrium point disappears as the concentration X at any point becomes dependent only on the production rate. Therefore –

$$\frac{dX}{dt} = \sigma \tag{7}$$

Solving the equation, we get –

$$X(t, X0) = X0 + \sigma . t \tag{8}$$

At time $t \rightarrow \infty$, the equation becomes –

$$X(t, X0) \rightarrow \infty \tag{9}$$

Thus, the system does not possess any equilibrium state anymore. Essentially, what we have done by completely disrupting the degeneration is to change the system

[153] Woodward (2018) notes that the invariance condition holds not just for causal claims, but for mathematical and conceptual claims too. Therefore, showing that the equilibrium concentration shows counterfactual dependence on the production rate and degeneration coefficient over a certain range is still possible under the manipulationist account.

structurally from one that shows equilibrium to one that does not. We have changed the formal constraints in such a way that some structural property of the system is altered.

This is not to say that a particular instance of equilibrium cannot be explanatory. We can still provide a causal story as to how the system attains equilibrium. Therefore, it is critical in understanding how the trajectory of the system as it moves towards equilibrium is specified by parameters that are not essentially the part of the aforementioned dependency relationships. The initial concentration still plays a part in defining the trajectory as can be seen from equation (5). Thus, the trajectory of different systems towards equilibrium will be different and dependent on certain variables that are not formal constraints on the model. We can, in principle, give a mechanistic explanation of the trajectories and how the system follows the trajectory to its final equilibrium position, and manipulate them to obtain different trajectories and equilibrium states. Thus, by intervening on the value of initial concentration (X_0), we can obtain a different trajectory to the equilibrium state. For example, in the neutrophil differentiation case (see *Example 2*) different trajectories with different causal factors will still lead to neutrophil development due to presence of an attractor.

Similarly, we can change the rate of production and degeneration to alter the trajectory. But the formal constraints on the system remain the same – that a system that shows production and degeneration of a component will eventually settle in an equilibrium state. The model still explains minimally why equilibrium will obtain.

A lot of explanations in biology show the system of being in dynamic equilibrium or steady state. The state of the system in steady state acquires a property where the concentrations of the entities remain constant over a period of time. The steady-state equilibrium, unlike chemical equilibrium, is not characterised by the absence of work

or influx of material, but by a system that performs work in order to maintain the steady-state. For example, attractor states may correspond with stable or unstable dynamic equilibrium[154]. For example, in the CMP example, the meta-stable state forms an unstable equilibrium which can be disturbed by changing some aspects of the system – either the steady-state disappears, and thus the system does not remain in dynamic equilibrium anymore or driving the system out of unstable dynamic equilibrium and towards a stable dynamic equilibrium.

Therefore, minimal models explain by citing formal constraints on the system such that the system is shown to belong to a particular universality class. Mathematical models in systems biology too follow a similar explanatory structure. As shown in the CMP example, the modelling exercise proceeds by showing how a topology consisting of two nodes which autoregulate their own concentrations and mutually inhibit each other's productions, show a particular global behaviour necessitated by the mathematical derivation. This model is shown to 'fit' with the systems whose behaviour is to be explained showing that the system will show the behaviour because it realises the topology and thus, belongs to a specific universality class. The parameter estimation, which is essential to provide a physical interpretation of the minimal model, does not proceed by tracking causal trajectories. Many steps in the mechanistic model might be abstracted away from and idealized. Therefore, while the mathematical/computational model can work as a heuristic for mechanism discovery and as a useful method for understanding dynamic nature of a mechanism, its explanatory power is not derived

[154] This is not strictly true as there may different kinds of attractors like limit cycles and torus attractors that can show different kinds of stability. In addition, there can be strange attractors which show chaotic behaviour. I am keeping my analysis limited to point attractors as described in the examples that I have considered.

from representing the mechanism but through the mathematical derivation and application to the target system through subsequent model-fitting.

In the next section I examine another account of how models of systems biology explain within a mechanistic framework. This account construes the two different kinds of modelling (mechanistic and mathematical) as complementary – i.e. the two models explain different aspects of the mechanisms. I argue that this account is better suited to account for the utility of mathematical modelling in mechanistic explanations, but still has some shortcomings.

3.5: Collaborative Explanations – A Reconciliation?

A possible way to reconcile mechanistic explanations and mathematical/computational explanations has been presented by Fagan (2015) in form of collaborative explanations. Fagan (2015) notes, mechanistic explanations (mEx) consist of two different strands – causal mechanistic explanations (mEx_{causal}) and constitutive mechanistic explanations ($mEx_{constitutive}$). While most of the literature has focused on mEx_{causal}, the $mEx_{constitutive}$ has been ignored or conflated with the causal. Fagan (2015) states that while mechanistic explanations have the goal of explaining the phenomenon of interest by describing the underlying mechanism in terms of organised interactions of its parts, the difference between the two approaches is about how their explanandum is interpreted. While mEx_{causal} is concerned with the causal aspect of the mechanistic explanation, i.e. how the phenomenon is brought about due to complex interaction between parts, the $mEx_{constitutive}$ has its explanandum as the working of the whole due to the complex interactions between the parts. The explanans for both the case though remains the same

– that of complex interaction between parts. Thus, mEx_{causal} is concerned with the cause and effect relationship between the mechanism and the phenomenon. In contrast, $mEx_{constitutive}$ is concerned with the mereological relationship between the mechanism as a whole and its parts.

The difference between mEx_{causal} and $mEx_{constitutive}$ brings forth another avenue for the analysis of mathematical/computational models. As Fagan (2015) notes, when mechanists talk about constitution, they generally limit their analysis to constitutive relevance – i.e. what parts and activities need to be considered as a causally relevant to the bringing about of the phenomenon [Craver (2007)].

Instead, Fagan (2015) shows how methods of systems biology can provide these constitutive explanations (especially in complex systems), with an example of the development of stem cells into mature cells. The process utilises various techniques of mathematical/computational modelling, which we have been discussed in previous sections. The investigation into how the stem cells develop into mature cells of certain types can be divided into six steps (*figure 3.15*)–

1. Providing a mEx_{causal} model which specifies the parts of the mechanism and the causal relationships between them.
2. Building a 'wiring diagram' consisting of nodes and edges corresponding to the components (like genes, proteins, etc.) and relationships (activation, repression, etc.). The nodes represent not only the components but their concentration (where applicable) as a time-dependent state variable. The collection of all the possible values of these state variables forms the state space for the whole system.

3. Formalised the wiring diagram into mathematical equations expressing the change in state variables over time. The equations can be solved for both local variables which show the change in the state variables of individual nodes over time, and for a global solution showing steady states of the system.
4. The local solutions are expressed as vectors in the state space, which converge towards the system's steady state (attractors).
5. Constructing landscape model that represents the state space as a multi-dimensional state space. Fagan (2015) notes that the landscape is not an effect produced by the molecular mechanism, instead it is the representation of the network considered as whole changing over time[155].
6. Providing correspondence between the landscape and the cellular components. The landscapes 'top' side represents the explanandum – that of development of cell into distinct mature states as branching pathways ending on attractor states, while the 'bottom' side represents the molecular contributions of the components of the network.

[155] Fagan (2015), p.74.

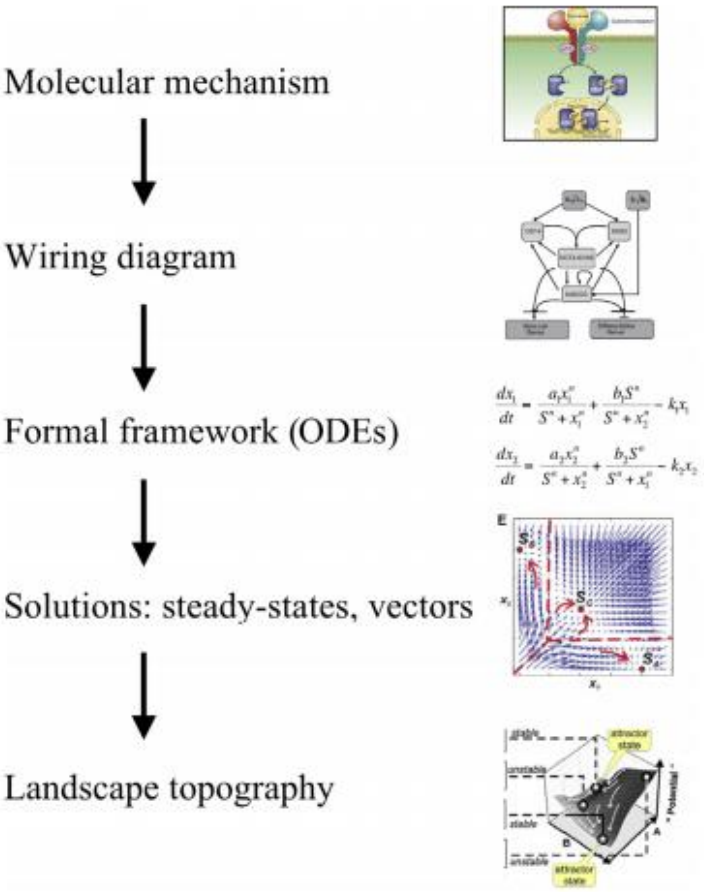

Figure 3.15 – The different steps in constructing mEx$_{constitutive}$. From Fagan (2015), p.73.

Fagan (2015) thus, explains what can be called joint mechanistic explanation. A joint mechanistic explanation works by explaining how the parts of the mechanism jointly explain the phenomenon –

"Components $x_1,\ldots x_n$ jointly Ψ as system M if and only if:

i. each x_i has properties that mesh with one or more other x_i,
ii. $x_1,\ldots x_n$ are spatially organized and their activities φ_1-ing,… φ_m-ing causally organized in virtue of their meshing properties,
iii. $x_1,\ldots x_n$ and $\varphi_1,\ldots \varphi_m$ so organized constitute system M Ψ-ing, and

iv. $x_1,...x_n$ and $\varphi_1,\ldots \varphi_m$ not so organized do not constitute M Ψ-ing."[156]

As we can see, the meshing properties are doing the explanatory work in this account. Meshing properties enable us to explain how complexes between two or more components are formed such that their joint existence is necessary for the mEx as they explain the behaviour of the whole. Fagan (2015) cites two examples – that of complexes formed during transcription and translation of proteins, and that of enzymatic action. In the former, we have complexes like DNA sequences and specific mRNAs, mRNAs and large ribosomal subunits, tRNA and specific amino acids. In the latter, enzymatic action is explained in terms of enzyme-substrate binding at specific active cites explains the complex's catalytic activities, these thus form the wholes that together explain the mereological dependence of the whole on its parts owing to their meshing together in a specific way.

Motivated by the account of $mEx_{constitutional}$, Fagan (2015) provides an account for 'collaborative explanations', where "given a set of components and their associated activities, and an overall system exhibiting some behavior, an explanatory model of the latter should describe:

(i) the properties of components that allow them to interact in specific ways

(ii) the spatial and causal organization determined by these interactions

(iii) how the overall system behavior is constituted by the organized components, such that

(iv) components' organization makes a difference to the overall system behavior"[157]

[156] Fagan (2015), p.75
[157] Fagan (2015), p.75.

Collaborative explanations thus explain by elucidating two properties – dependency between parts and whole such that the whole is dependent on the parts organised in a certain way and properties of the components that enable them to 'mesh' with one another.

Fagan (2015) is correct in pointing out that mathematical/computational models can provide constitutive explanations. For example, the landscape derived from the formal framework does show a global solution that can be interpreted as the combined result of the whole network's action. Moreover, Fagan's account of collaborative explanation elucidates the need of both – abstracting away from the causal details of the system to explain organisational properties, and of the use of mathematical modelling in cashing out the said properties. Also, it aims to fill a gap in the mechanistic philosophy – that of the explanatory questions related to whole and parts. But I think there are some caveats that need to be spelled out and show how the mathematical models achieve this in order to show how there is a limitation to this idea.

Firstly, while Fagan (2015) illustrates the role of abstraction in the construction of the wiring diagram from mEx_{causal}, the consequences of such abstraction are not considered. As Fagan (2015) notes, while constructing the wiring diagram from mEx_{causal} the "(n)odes do not correspond exactly to components in the mechanism description. Complexes of multiple components, functionally distinct forms of one molecule, or the same molecule in different cell compartments are often represented as distinct nodes"[158]. Thus, the components cited in mEX_{causal}, and those that are utilised by $mEx_{constitutional}$ might not be the same. Thus, the compatibility of mEx_{causal} and $mEx_{constitutional}$ might not be established by default and may need further analysis and

[158] Fagan (2015), p.73

some pre-requisite knowledge about how some complexes are formed. The second problem is that of multiple realizability. For example, in neutrophil development case [Huang, et al. (2005), see section 3.3.3 *Example 1*] multiple pathways realise the same result. Therefore, many different mechanisms might be matches when the abstraction steps lead to similar topological structures.

Similarly, how do the idealization steps required for mathematical explanations affect the $mEx_{constitutional}$? Specifically, what bearing do the idealizations in each step have on the next step? For example, what components form the nodes will have a bearing on how state spaces are defined and eventually on what local and global solutions are derived. Furthermore, in many cases, the steady state itself might be an idealization. As Heinrich and Schuster (1998) note "(t)he concept of steady-state is a mathematical idealization that can describe real situations only in an approximative way, due to fluctuations of different nature. Every steady-state can be considered as a quasi-steady-state of a subsystem embedded in a larger, non-stationary system"[159]. Tracking idealizations is essential because it can have a bearing on how different models connect with each other.

The bearings of abstraction, idealization, and optimisation on $mEx_{constitutive}$ is significant. For one, the methods of systems biology (especially mathematical modelling) proceed by constructing models that are highly idealized and abstracted from the system in question. Therefore, the correspondence step needs varied experimentation to show how the model thus constructed maps onto the system in question. Secondly, these idealizations and abstractions have a bearing on what is explained and how it is explained. While Fagan (2015) argues that the model explains

[159] Heinrich and Schuster (1998), p.65.

the specific meshing properties of the system's components, that might not be the case. The meshing properties, once modelled are explained in a very abstract and idealized form such that the model shows that the components of the system do 'mesh' together but does not necessarily describe the meshing property in all its specificity. Again, what these meshing properties are must be established independently of the model. For example, the persistence detector as applied to the bacterial arabinose metabolism system indicates that CRP-cAMP complex occupies the node X such that their binding activates AraC. Still, the abstract nature of the modelling does not define the nature of the binding.

As I have noted, mathematical models explain by situating the system in a universality class. Therefore, while it is true that the model will minimally elucidate the dependencies of the system such that some inferences about the structure can be drawn (for example, a system which belongs to the same universality class as an equilibrium model will show production and degeneration of the components and thus an equilibrium) but the specificity of the organisation must be established through experimentation into what the components are and how they mesh specifically.

Nonetheless, I think that Fagan's account is a step forward. Firstly, because it specifically illustrates the difference between constitutive and causal mechanistic explanation, which has received much less attention in the literature. Secondly, it shows how modellers can reason about one form of explanation while using models that are not constructed with the same explanatory framework in mind. The generality offered by mathematical models by constructing minimal models places the mechanistic models that can be derived from it in a universality class and shows at least some dependencies that will be realised by the $mEx_{constitutive}$. This can thus form the first step towards

deriving the specificity of the mEx, and to test certain hypothesis about how constitutive phenomenon can be achieved. In this respect, the abstraction, idealization, and optimisation help the modeller to understand the formal dependencies and work both top-down and bottom-up to –

a) ascertain that the mEx_{causal} cited is accurate enough to realise the said behaviour. If the construction of the state-space model does not correspond well with the behaviour that is being studied, then we can infer that some constitutively relevant aspect has been missed [Baetu (2015a, 2015b)]; and

b) to establish at least some dependencies that can be further explored to understand how the specific meshing properties are established. This is because the casual mechanistic details will be constrained due to O-D dependency relationships of the class the system belongs to. Therefore, there is merit to think of mathematical models as complimentary to mechanistic models, albeit attention must be paid that the specificity of the mechanism cannot be directly drawn from the mathematical model, but must be derived from the knowledge of the formal dependencies that are discovered by the said model.

3.6: Conclusion

In this chapter, I have argued that mathematical models of system biology are constituted by diverse modelling techniques that can be construed in an explanatory framework. The explanatory powers of these models is not derived from rehearsing a causal story about the phenomenon, but by showing how the system that realises the phenomenon belongs to a universality class and thus shows certain dependencies

despite the difference in realisations. This is achieved by the process of optimisation and abstracting away from and idealizing the causal details. As a result, the models of system biology are minimal models [Batterman and Rice (2014)]. Nonetheless, I agree with Brigandt (2015), Levy andBechtel (2013) that these models have explanatory import for mechanistic explanations. I also agree with Fagan (2015) that these models can help reason about the constitutive mechanistic explanation albeit in an abstract way. The mathematical/computational models of systems biology can elucidate mechanistic features of the system by explaining their dynamic nature, organisational features, and by aiding discovery of mechanisms. But the causal features of the mechanisms cannot be deduced directly from the mathematical models and must be interpreted separately. Instead the mathematical models explain by showing how a certain class of systems realise certain dependencies. This is because the mathematical models idealize and abstract away from mechanistic details and construct optimal models which show only structural dependency relations and not causal relations. Nonetheless, the models can provide a 'constraint-based reasoning' for mechanistic explanations such that they constrain the possibilities for the realisation of the phenomenon by placing it in a universality class and showing what minimal dependencies hold between the features of the system.

CHAPTER FOUR

Systems Theory and Closure of Constraints

In the last chapter, following O'Malley and Dupré (2005), I drew a distinction between pragmatic models in systems biology and systems theoretic models. In that chapter, I concentrated on the former strand of systems biology. I argued that the best way to understand mathematical/computational models is to understand them as explaining of dependency relationships between features of the system that can be represented as formal constrains. While the framework offers a generality of types of systems where certain mathematical formalism can explain certain dependencies, it fails to offer general principles for biological systems, especially the characteristics of biological systems in terms of an organised whole capable of autonomous existence. By autonomy, what is generally meant is that the system produces its own conditions of existence.

In this chapter, I focus on the second strand – systems theory. Systems theory aims to provide general principles for what should being called a biological system must entail. The tendency for defining these general principles is to study higher-level organisational relations that are necessary for the realisation of biological systems as compared to systems that do not fall under the biological domain. The structure of the chapter is as follows – I start with a general introduction to the idea of self-determination through both historical and contemporary ideas. The aim of this exercise is two-fold. Firstly, the analysis offers some basic ideas about biological organisation. And secondly, it enables us to study a particular model – that of the closure of constraints as presented by Mossio, Moreno, Montévil, amongst others [Mossio, Bich, and Moreno

(2013); Moreno and Mossio (2015); Montévil and Mossio (2015); Mossio and Moreno (2010)] (Mossio et al., henceforth). The basic idea of their argument is that biological systems achieve a causal closure at a certain level such that they are able to determine their own conditions of survival and achieve a minimal autonomous existence.

My aim in this chapter is twofold. Firstly, to analyse and further the idea of closure of constraints as a viable theory of biological organisation. As philosophers like Mossio, Montévil, Bich have noted – biological systems differ from other physical systems due to the property of closure of constraints. This helps the biological system maintain itself in a far from equilibrium stasis and avoid the perils of the second law of thermodynamics. For this, they need to have a conception of constraints as causally efficacious. To do this, they need a concept of emergence that can explain the idea of such an emergent organisation within the realm of biology. I note that there is an oversight in their theory about what configurations that realise constraints are, which is related to the idea of emergent organisation. I argue that a better picture can be provided by supplementing their idea of configurations with Wilson's (2010) concept of emergence in terms of reduction, restriction, and elimination of degrees of freedom. This helps explain the nature of constraints as emergent configurations that assert causal influence over its surroundings, thus help sustain the system by directing the flow of energy. Secondly, following this, I show how the constraints so realised can explain certain features of biological systems - like regulation and robustness - as general principles for systematic organisation. This helps us construe the insights from the closure as constraints view as a model for general principles for biological organisation I combine the idea of constraints with that design explanations [Wouters (2007)] to show that the ubiquity of regulation and robustness in biological systems can be

explained by modelling them as self-maintaining organisations maintained by the constraints. This requires us to understand the nature of the constraints, how and why control is exerted, and how the systems, thus, are able to maintain themselves and exert second-order control on the system. I end the chapter with an example of glycemia regulation in humans as a case study to show how these concepts fit together to model the regulatory features of a system and thus answer certain design questions.

4.1: Systems Theory: Defining Biological Organisations

The first observation, while defining any biological organisation, is to realise that biological systems are open systems that exchange energy and matter with the environment. This observation seems essential when we consider that like all physical systems, biological organisations are bound by the second law of thermodynamics. The second law states that in an isolated system (a system where no exchange of energy and matter takes place with the surrounding), the entropy always increases. For example, a chemical reaction always tends towards the state of equilibrium where the forward and backward reaction reaches the same rate, thus decreasing the energy of the chemical system to its lowest level. The reaction cannot proceed beyond this as it would violate the second law of thermodynamics unless it is acted upon by outside forces. For example, adding more components will shift the system out of its equilibrium until the chemical reaction proceeds to attain equilibrium again. Therefore, in the absence of disturbances from the surroundings, the system will maintain its equilibrium. As von Bertalanffy (1968) notes, at first sight biological systems seem to be in equilibrium. Cells tend to maintain a particular ratio of compositions, and in the case of perturbations, tend to return to their original composition or achieve a new stable composition. But the

appearance of stability in biological systems, while is analogous to equilibrium, is not one of true equilibrium. This is because the equilibrium is a feature of an isolated system left to its own devices. But the biological system is not an isolated system. It is an open system which exchanges energy and matter with its surrounding. Left to its own devices, the exchange of energy and matter will lead to the disruption of the equilibrium. Another aspect is that a system in equilibrium does not require energy and no energy can be derived from it. But the biological system requires energy and performs work – it consumes nutrients, it expels waste, it moves through space. It is thus, not in equilibrium. The appearance of equilibrium, in such cases, is not a true equilibrium, but a steady-state maintained by the system through a constant exchange of matter and energy with the surrounding. This means that work has to be put in to maintain the system in dynamic equilibrium. As von Bertalanffy (1968) notes, the entropy change in isolated systems is always increasing such that –

$$dS \geq 0$$

Where the entropy (dS) always increases with time. But this is not true for open systems, where–

$$dS = dS_e + dS_i$$

While processes inside the system – like chemical reactions, energy dissipation will lead to an increase in entropy (dS_i), the system can import matter to decrease their entropy (dS_e). Organisms do this by acquiring food which can be broken down to fuel the system. In the process, the net entropy can be negative such that $dS_e > dS_i$. This is how a system is able to maintain negative entropy. Schrödinger in his seminal work 'What is life' (1944) alludes to this when he writes that the 'organism feeds on negative

entropy'. What it means is that biological systems, due to their openness to the surroundings, can import matter and energy and can perform work such that they can maintain themselves at a far from equilibrium dynamics.[160]

Kauffman (2000) expands on this idea by introducing the concept of autonomous agents and work constraint cycle. For Kauffman, an autonomous agent "is a reproducing system that carries out at least one thermodynamic work cycle"[161]. Where work cycle is defined as a coupling of endergonic (where the energy of the products is higher than the reactants or the reaction absorbs energy) and exergonic (where the energy of the product is lower than the reactants or the reaction expel energy) reactions. To conceptualise this, consider a Carnot engine (*figure 4.1*). The Carnot engine converts thermal energy into mechanical energy. The engine consists of a piston inside a cylinder with a working gas. The movement of the piston expands or compresses the working gas. In addition, the engine consists of a heat source and a sink. A work cycle in Carnot engine consists of the 4 steps during which the movement of piston corresponds to the volume and pressure of the working gas (see *figure 4.1*), to complete a work cycle.

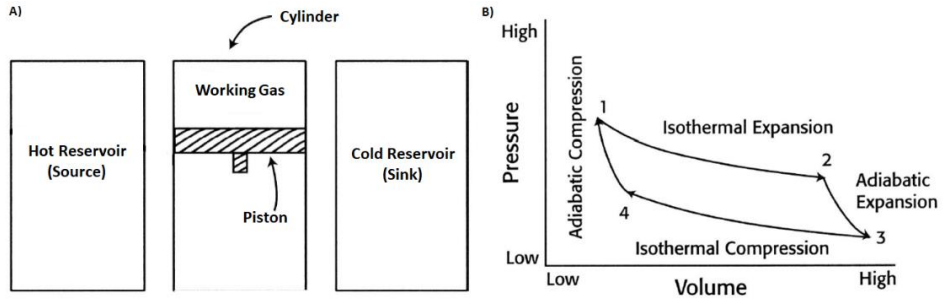

Figure 4.1 – a) The construction of a Carnot engine. b) the volume-pressure diagram of a Carnot cycle showing Isothermal Expansion (1-2) where the working gas is heated due to contact with the source thus

[160] von Bertalanfy further elucidates the use of isomorphic models to explain systems across the sciences (including both biological and otherwise) based on higher order similarities at an abstract level – an idea similar to what has been discussed in Chapter 3. I will not be discussing this further here. See Green and Wolkenhauer (2013) for further analysis.

[161] Kauffman (2000), p.100.

leading to expansion which moves the piston outwards thus maintaining a (more or less) constant pressure; Adiabatic Expansion (2-3) where the contact between the source and the gas is severed but the gas continues to expand, pushing the piston outwards thus decreasing pressure; Isothermal Compression (3-4) where the gas is cooled down due to contact with the sink, and at the same time piston is moved back in thus reducing the volume but maintaining the pressure; and Adiabatic Compression (4-1) where contact between sink and gas is severed but the movement of piston continues therefore increasing the pressure. At the end of this process the assembly is back to its original configuration [from Kauffman (2000), p.54].

There are two aspects to be noted here –

1. The engine couples endergonic and exergonic processes to perform work. If the engine only consisted of a source and not a sink, then the work done to push the piston back into the cylinder so that it could reach its original position would be as much as the work done by the piston during the isothermal and adiabatic expansion.
2. The constraints in the form of the cylinder's wall and piston play an essential role in deriving useful work from the engine. If the cylinder has non-rigid walls with no moveable piston, then the pressure changes and expansion of volume in one direction (that of the piston's axis) will not be obtained. Instead, the cylinder will lose energy through all of its walls, thus dissipating the energy gained from the source without providing any useful work.

Thus, Kauffman (2000) institutes what he calls a work-constraint cycle. As per the concept, the constraints on the system are essential to harness the system's energy towards what can be called useful work. Therefore, to maintain a system such that it can sustain its characteristic of doing work, the constraints must be maintained. In the case of the engine, the constraints are maintained by putting work into the system at the end of the expansion phase.

Kauffman (2000) proceeds to apply the same idea to biological systems. He conceives of an auto-catalytic set. The simplest example of an autocatalytic set is of a couple of reactions that catalyse each other – if A catalyzes the formation of B from B' and B'' and in turn B catalyzes the formation of A from A' and A'', then the two catalyzed reactions form an autocatalytic set. The set realises what can be called a 'catalytic closure' such that they form a system – consisting of the reactions that form A from A' and A'', and B from B' and B'' catalysis itself.

But the auto-catalytic set is still not a completely autonomous system. The autocatalytic set left to its own devices will eventually reach equilibrium, and therefore the reaction will stop. In an isolated system, this will be the end of the process. But since a biological system is an open system, it can be driven away from equilibrium into steady-state dynamics by providing free energy to the system. The thermodynamic grounding that Kauffman (2000) aims to provide to the process is something similar to a Carnot engine which completes a work cycle by coupling exergonic and endergonic processes. A biological system, according to Kauffman (2000) realises a similar work-constraint cycle by coupling the auto-catalytic set with energy imported from outside the system. As an example, let us consider Kauffman's hypothetical system where two DNA trimers ligate to form a hexamer, such that the hexamer catalyzes its own synthesis while being an endergonic reaction since the hexamer has higher energy than the trimers (*figure 4.2*).

When left to its own devices, and in the presence of access trimers, the system will reach the equilibrium condition. But, when it is coupled with the exergonic breakage of pyrophosphate (PP) into two monophosphates (P+P), which provides free energy to the trimer ligation an increase in the concentration of hexamer beyond equilibrium

concentration can be obtained. But once the concentration of PP and P+P reaches the equilibrium concentration; the reaction will stop. This is true unless the endergonic reaction of the monophosphate combining to form the pyrophosphate can be coupled with another exergonic reaction. Kauffman (2000) argues that this can be done by the system importing energy in the form of photons exciting an electron which then can provide the free energy to the endergonic reaction of P+P → P. Furthermore, if one of the trimers acts as a catalyst to the reaction, we can see how an auto-catalytic set is created. Firstly, the photon excites an electron. The excited electron is then used as a free energy source for the endergonic ligation of PP from P+P, catalysed by a DNA trimer. The high energy PP formed during the endergonic ligation then breaks down to its P+P form in an exergonic reaction. The energy from this exergonic reaction is used by the ligation process of the hexamer from the trimers. The hexamer itself catalyzes this reaction. As a result, the hexamer concentration can be driven beyond its equilibrium concentration. The whole process makes an autocatalytic set because the products of the two processes catalyze each other.

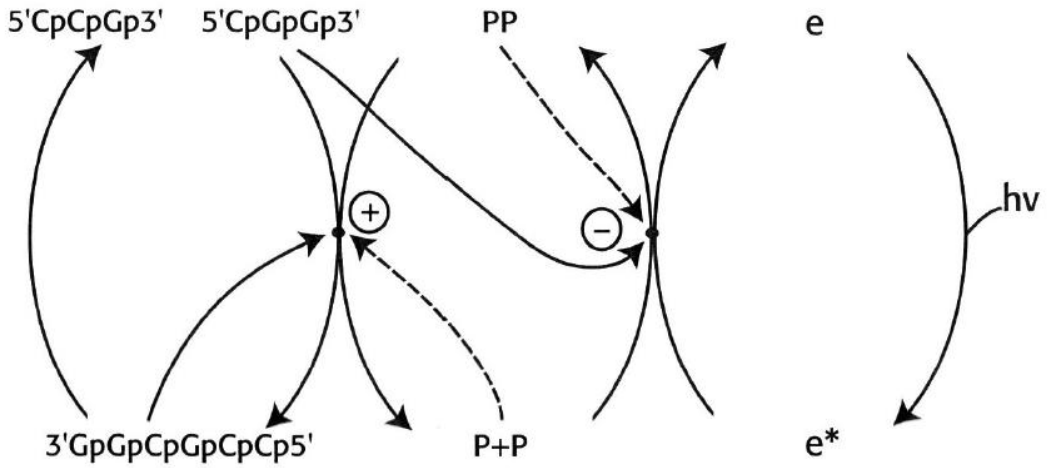

Figure 4.2 – Kauffman's (2000) hypothetical autocatalytic set realising a work-constrain cycle [from Kauffman (2000), p.65].

But Kauffman's conception of the work-constraint cycle, as analogous to Carnot's engine has a few problems. Firstly, although the concept of work-constraint cycle explains the dynamics of a closed system (a system that only exchanges energy with the surrounding), it does not sufficiently explain open systems (that exchange both energy and matter with the surrounding). Biological systems are open to both matter and energy. Therefore, significant part of explaining the biological system is not covered by Kauffman's work-cycle constraint. Secondly, while Kauffman realises the importance of constraints for how useful work is derived, the analogy of the Carnot engine for the autocatalytic system can only take the work-constraint cycle so far. Constraints are essential for work, and they achieve it by directing the thermodynamic flow in a particular way. But while the constraints in the Carnot engine are constructed by an external agency for a specific purpose – that of deriving mechanical work from thermal energy, the same cannot be said for biological systems. Biological systems construct their own conditions for existence. This idea – called self-determination – in the literature [Moreno and Mossio (2015); Montévil and Mossio (2015); Mossio and Moreno (2010)], is the idea that biological organisations determine the conditions for their own survival. What it entails is that biological systems, at a relevant timescale, are capable of taking over (some of) their own boundary conditions, and thus determine their own conditions for survival such that the system can –

a) have a minimal causal agency on their own organisational qualities such that they can dampen out at least some deviations in the environmental conditions, and

b) be self-maintaining such that they can maintain their own organisational qualities over time.

At this point, some clarifications are required. Many qualities have been ascribed to biological systems. These include – self-organising, self-maintaining, and self-determining. But all these terms mean different things.

Self-organising means capability of obtaining spontaneous order from disordered interactions between parts that make up the organisation. A Bénard cell is an example of self-organising systems. The Bénard cell is formed when there is a temperature differential between layers of a liquid, such that the molecules achieve a spontaneous order in the form of their rotation and form a convection cell. The convection cell pattern formed by the molecules is a metastable structure such that it survives as long as the temperature differential is maintained. At the same time, it can have a causal influence on its surrounding such that the surrounding liquid molecules can be captured by the Bénard cell and become part of the cell. But the existence of the Bénard cell depends largely on the external conditions. A disturbance in the external conditions can critically alter the structure of the cells and even its existence.

Self-maintenance is the characteristic of the system to be able to maintain itself over time despite changes in its initial conditions of existence. A biological organisation, for example, is able to maintain itself over time despite changes in its organisational structure. We can have changes that correspond to growth, injury, and general replacement of the components. Despite this, the biological organisation is able to maintain its organisation.

Self-determination, on the other hand, is a more complex idea that involves both self-organisation and self-maintenance, but also considers agency. As already noted, a self-

determining system is said to be able to take over at least some of its boundary conditions and hence the conditions of its survival.

Consider a cell. A cell is composed of various components. One of them is the cell membrane. The cell membrane acts as a barrier between the cell and its environment such that it regulates the inflow and outflow of matter. The cell membrane, in turn, is maintained by the cell itself through various processes. Therefore, the cell is able to regulate its own internal environment through the maintenance of the cell membrane. Self-determination, thus, requires a more complex network of constraints on thermodynamic flow over a more extended period of time, and for a more extensive set of environmental perturbations. The requirement for self-determination is not only that the system is able to maintain a singular state over time, but that it is able to transit from one state to another with a change in external conditions. Self-determination, therefore, requires a more regulated system.

A model that is more suited to understanding constraints and self-determination is Gánti's Chemoton (2003). Gánti aims to provide the basis for the definition of minimal living systems. To this end, Gánti's chemical automaton (Chemoton) serves as a template for a minimal system that defines the necessary condition for what can be called as living. The Chemoton consists of three subsystems (see *figure 4.3*) –

a. A metabolic subsystem – a monomer (A_i) based subsystem that converts nutrients into useful energy that can be used by the other two subsystems,
b. A template replication subsystem that is composed of monomers (V_i) that undergoes template replication to form polymers (pV_n).
c. A membrane subsystem that is composed of components (T_i), which acts as a semi-permeable membrane controlling entry and expulsion of material.

These systems are considered to be stoichiometrically coupled. For example, Gánti theorises the metabolic cycle as autocatalytic such that –

$$A + X \rightarrow 2A + T + Y$$

Therefore, the cycle produces more of its intermediaries (A) along with the membrane construction components (T) while taking in the nutrients (X) and expulsing the waste (Y). Therefore, the metabolic subsystem maintains the membrane. In turn, the membrane acts as a barrier that maintains the composition of the Chemoton, which facilitates the metabolic subsystem.

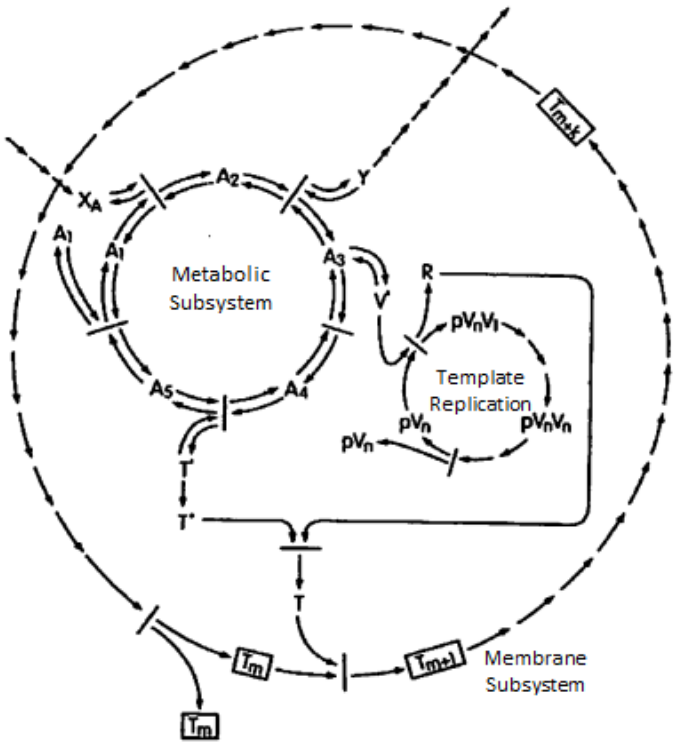

Figure 4.3 –Gánti's Chemoton showing the Metabolic, Membrane and Template replication subsystem and their dependence on each other [adopted, with change, from Griesemer and Szathmáry(2009), p.486].

Since the metabolic subsystem is stoichiometrically[162] dependent on the availability of the right nutrients and outflux of the waste (because otherwise, the chemical reactions will not proceed due to reaching equilibrium), the usefulness of membrane is not only an added feature but a necessity for the correct functioning of the metabolic subsystem. Thus, the two subsystems (metabolic and membrane) form a system that as Gánti notes, expresses a characteristic property of biological systems, which is "separable from the external world and its internal composition differs from that of the environment. It continuously consumes substances that it needs from the environment which are transformed in a regulated chemical manner into its own body constituents"[163].

As the Chemoton grows, it shows the capacity to replicate due to growth of the system due to both metabolism and growth of the membrane, until it reaches a threshold where it divides.

Gánti explains the use of the template replication subsystem. The subsystem takes monomers for the metabolic subsystem and links them together over the cycles of metabolism to create a polymer chain. The polymer chain thus works to keep a count of the metabolic cycle. Although Gánti does not envision it in such a way, the template replication system can work as a control/regulatory subsystem on the Chemoton as the length of the polymer defines the rate of replication of the Chemoton. As Bechtel (2007) notes, without a control/regulatory system, the Chemoton will perform its function as long as the environmental conditions are in, but –

[162] Stoichiometry is the study of quantitative change between reactants and resultants in chemistry bound by law of mass action. Stoichiometric dependency is thus the relationship between how the availability of reactants will result in the quantitative production of the resultants.
[163] Gánti (2003), p.105.

> "Even slight variations in the environment may disrupt such a system. Imagine the environment changed so that a new substance entered the system which would react with existing metabolites, either breaking down structure or building new additional structure. This would disrupt the delicate balance between metabolism and membrane generation that Gánti relies on to enable Chemotons to reproduce."[164]

But a control system, for example, enzymes, can provide specificity to the metabolic system that prevents this from happening. An enzyme catalyzes the reaction by binding to the specific substrate. While Gánti's Chemoton does not use enzymes for its metabolic system, some modern interpretations[165], note that template replication can provide the basis for regulatory control of the Chemoton by providing enzyme like structures.

Gánti's Chemoton utilises two aspects – one is that of dynamic coupling of subsystems such that the systems help produce each other. In the Chemoton case, the metabolic subsystem produces the membrane subsystem, which in turn sustains the metabolic subsystem by maintaining the influx of the nutrients and outflux of the waste. And the other, the regulation of the system through another subsystem (template subsystem).

Gánti's Chemoton shows three aspects of self-determination that have been emphasised–

[164] Bechtel (2007), p.289.

[165] Griesemer and Szathmáry (2009) argue that if we consider the monomers of different types then the polymer chain thus formed will be of a particular sequence such that this sequential order will be free of the underlying stoichiometric processes that formed the monomer (as the rate will not matter, only the sequence of monomers will matter to the polymer chain). This sequence can be used to regulate the system (for example – as enzymes).

1. *Cyclical existence* – where the processes constitute a cyclical process such that each cycle performs some work. Imagine the metabolic subsystem did not show a cyclical existence. In this case, the membrane growth will not be seen as the membrane components are produced by the metabolic cycle. Therefore, the coupling between the membrane subsystem and the metabolic subsystem will not be established. Hence, the condition of self-determination of the system will not be achieved since the system (without the metabolic cycle) will not be able to produce the membrane.
2. The Chemoton organisation achieves a *minimal organisational closure*. Since the metabolic and membrane subsystems constrain each other – the membrane is dependent on, and thus constrained by the metabolic cycle for its production, which is in turn constrained by the membrane which provides the nutrients and expulses waste, thus maintaining the correct concentrations for the metabolic cycle to continue. The characterising of these dependencies as constrains is because the metabolic subsystem's specificity is maintained because of the nutrients (X) available. If, as Bechtel (2007) notes, some other substrates (X') are available, the metabolic cycle may be disrupted. The specificity of the substrate is maintained by the concentration of the specific substrate by the membrane.
3. The Chemoton, at least in its modern understanding, shows the importance of *control*. The control can be divided into two different forms – one is the minimal regulation realised by the membrane and metabolic subsystem themselves due to their stoichiometric coupling. The other regulation is provided by the transcription subsystem which depends on the stoichiometry of the other two

cycles in a limited manner – i.e. it depends on the two cycles for its raw materials and energy, but is decoupled from them in terms of its functioning.

In the modern interpretation of Gánti, we can see the ideas presented and expanded by Moreno and Mossio (2015). Moreno and Mossio (2015) argue that this self-determination of biological organisation is intrinsically linked to two aspects of the organisation –

1. Biological organisations achieve a particular type of closure of the organisation – that of constraints.
2. A regulatory system that acts on these constraints to maintain the organisational closure.

As per (1) in biological organisations, the constraints that direct the thermodynamic flow of the system are maintained by each other. Constraints, thus construed, act upon processes in such a way that they limit the boundary conditions for the processes – i.e. the constraints reduce the possible behaviour that a process can display. Processes, as defined by Moreno and Mossio (2015) are the "whole set of physicochemical changes (including reactions) occurring in biological systems, which involve the alteration, consumption and/or production of relevant entities"[166]. Processes are what realise work by taking part in the transformation of different entities. Constraints, on the other hand, do not take part in work directly. Instead, they impose boundary conditions on the processes such that they themselves are not thermodynamically altered in the process. Consider an enzyme. The enzyme increases the kinetics of the reaction it facilitates but is not altered by the reaction itself. Or consider the vascular system where the blood vessels are not altered by the flow of the blood. Therefore, while there is a temporal

[166] Moreno and Mossio (2015), p.11

asymmetry the entities taking part in the processes as they change during the process, the constraints maintain their symmetry throughout the process.

But over time (at a time scale different than the process), the constraints require maintenance. The crux of the closure of constraints idea is that constraints through their activities, attain a causal closure where they maintain each other, not directly but through the processes they act as constraints upon.

In many ways, the account of constraints presented by Moreno and Mossio (2015) is similar to Robert Rosen's account of closure to efficient causes. Rosen, in his book "Life Itself" (1991) utilises the Aristotelian four causes framework to argue that biological organisations are closed to efficient causes. According to Aristotle, we can provide four kinds of causal explanations pertaining to 'why' question –

a) *Material cause* – about the matter that composes the thing in question. For example – the bronze of a statue, or the wood that a sailboat is made of.
b) *Formal cause* – relating to the form or "what-it-is-to-be". For example – the shape of the statue, or the shape of the sailboat that provides it with the ability to sail.
c) *Efficient cause* – relating to the agency of the change, or why the change exists. Example – the efficient cause of a statue is the sculpture, or the constructor of the sailboat.
d) *Final cause* – the purpose or the sake for which the change exists. For example – the statue exists to depict some imagination in that the sculpture has, or the sailboat exists to sail.

According to Rosen, what differentiates a machine from a biological organisation is that the efficient causes of the former are an agency that lies beyond it. In contrast, the

biological organisation is its own efficient cause. In this conception, every process in the system (as shown in *figure 4.4C*, see figure *4.4* for further details) (A→B, B→*f*, and *f*→φ) has an efficient cause (*f*, φ, and B respectively) produced within the system. But the system is still an open system as the material cause of A lies outside the system. Therefore, the system maintains its own efficient causes by materially producing them. Rosen thus claims "a material system is an organism [a living system] if, and only if, it is closed to efficient causation"[167]. Rosen calls this system Metabolism-Repair (MR) system where the organisation performs metabolism with the help of efficient causes and in turn, repair the efficient causes.

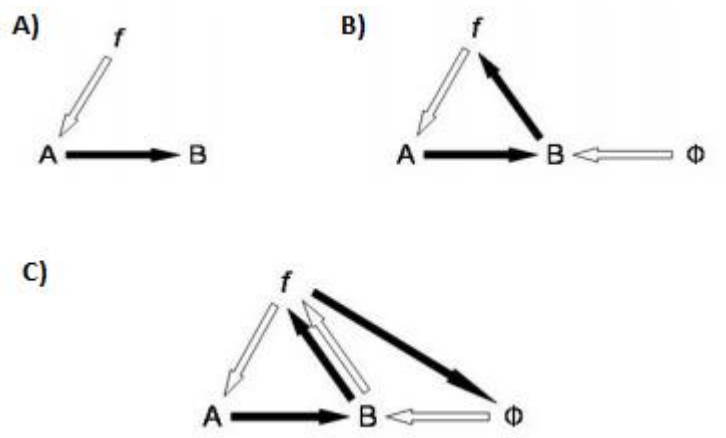

Figure 4.4 – Rosen's M-R system (white arrows show efficient causes; black arrows show material causes). A) B has A as its material cause and f as its efficient cause. B) Adding efficient cause Φ for f leaves Φ without an efficient cause thus leading to infinite regress if we keep adding new causes. The solution to this problem (C) Closure of efficient causes by assigning f as the material cause for φ with B as the efficient cause for the process [from Bechtel (2007), p.291-92].

In what follows, I will discuss the idea of closure of constraints and how it may explain minimal autonomy of biological organisations. Following Moreno and Mossio

[167] Rosen (1991), p244.

(2015) and Montévil and Mossio (2015), I characterise constraints, their closure, and the idea of organisation, and the consequences of accepting the framework of closure of constraints – that of emergent organisation and regulation. Eventually, I discuss the idea of minimal control and regulation to show how they can be used to define some characteristic features of functional organisation using the concept of design explanations.

4.2: Closure of Constraints

The central thesis presented by Mossio, et al is that the configurations have causal powers that are exerted on the surrounding system that they belong to. Some of these causal powers can be exerted as constrains on its surrounding, such that the constraints reduce the Degrees of Freedom of the system that they belong to. So, for example, the cell membrane restricts the movement of molecules across it, thus maintaining the composition of the cell state at that moment. Therefore, the cell membrane acts as a constraint on the molecules by reducing their DOFs (i.e. they can either be inside or outside the cell depending on cell membrane permeability). Or, for another example, enzymes affect the rate of reaction by decreasing the activation energy for the substrates to bind, thus leading the reaction towards equilibrium at a faster pace. Mossio et al. (2013) note that there are two conditions that a configuration (C) must satisfy to be a constraint–

1. "At a time scale τ_i, C is conserved throughout P_{surr}, i.e. there is a set of emerging properties S_1, …, S_n of C which remain unaffected throughout P_{surr};

2. At τ_i, C exerts a causal role on P_{surr}, i.e. there is some observable difference between P_{surr} and P^c_{surr} (P^c_{surr} is P_{surr} under the causal influence of C by virtue of properties $S_1, \ldots, S_n$)."[168]

According to (1), the configuration acting as constraint remains unaffected at the relevant time scale of the process that they act as a constrain on. The enzyme does not get altered during the reaction it facilitates. The cell membrane does not change in its configuration while maintaining a particular state of a cell. As per (2), on the relevant time scale, the function of constraints play is not to extend the possible behaviour of the processes they exert constrains on but to facilitate them (either spatially or temporally). The reactions that enzymes facilitate can take place without the enzyme. The enzyme increases the rate of reaction. Similarly, the processes inside the cell, in theory, do not require the presence of the cell membrane but the cell membrane facilities the processes by maintaining concentrations.

The appeal to configurations works in two ways –

Firstly, it distinguishes the constraints from purely material constraints that any physical process would realise. Any physio-chemical processes can impose constraints on each other. For example, a simple state of occupying a position in space and time can be a constraint on another particle as it cannot occupy the same space in that time. Reactions are also controlled by their stoichiometry – for example, concentrations constrain reactions. While these are essential dimensions of biological causation and have been studied for their role in biological organisations [Nathan (2012)], constraints

[168] Mossio, et al. (2013), p.164. Where P_{surr} are the processes that constraints act upon.

as configurations represent a different causal regime in biology. As noted above, they hold a certain asymmetry that constraints realised by the reactions themselves do not.

Secondly, the configurations construed as constraints capture some basic intuition about biological causation – where lower-level entities form configurations that are causally efficacious. Biological organisations form structures like cells, blood vessels, organs, etc., to which we can ascribe causal roles to, due to their configurational properties. I will discuss this idea in section 4.3 while discussing how configurations realise constraints.

Once Mossio, et al (2013) have established the concept of configurations acting as constraints, they contend that biological organisations achieve a *'Closure of Constraints'* such that the activities of the constraints collectively maintain each other. That is, the constraints realise the conditions for each other's existence. For example, the cell membrane constrains the activities of the cell and in turn, the cell maintains the cell membrane by synthesising the components needed for its maintenance. Similarly, the enzyme helps facilitate a reaction and in turn, is produced by a system that is maintained due to the carrying out of several constrained processes.

To see how closure is achieved, first let us define constraints. As Montévil and Mossio (2015) notes, a constraint on a process can be defined as –

> "Given a process A → B (A becomes B), C is a constraint on A → B, at a specific time scale τ, if and only if the following two conditions are fulfilled:
>
> I. The situations A → B and A_C → B_C (i.e. A → B under the influence of C) are not, as far as B is concerned, symmetric at a time scale τ.

II. A temporal symmetry is associated with all aspects of $C_{A \rightarrow B}$ with respect to the process $A_C \rightarrow B_C$, at time scale τ."[169]

The conditions specify the difference between processes and the constraints that act on them. As per (I) the action of constraints on the process is causally acting on it such that the process in the absence of the constraint ($A \rightarrow B$) is not the same as the process in the presence of constraint C ($A_C \rightarrow B_C$). For example, a circulatory system consisting of blood vessels canalises the flow of oxygen-carrying blood to the cells, while in the absence of the system, the flow of oxygen is purely through diffusion. Therefore, the circulatory system plays a constraining role in the flow of oxygen.

Condition (II) specifies the asymmetry of effects on the constraints while interacting with the process. While constraints have a causal effect on the process, they are not transformed by the process (on the relevant timescale of the process). The blood vessels and the structure of the circulatory system is not altered by the blood flowing in it. Constraints thus act on the processes such that they alter the process but are not altered by the process.

So how do constraints work? Montévil and Mossio (2015) note that constraints work by limiting the thermodynamic possibilities of the process. The circulatory system limits where the blood can go and therefore canalises the flow of oxygen. The cell membrane maintains the concentrations of various molecules inside the cell by limiting their movement to inside the cell. Therefore, the constraints act as a restraint on the thermodynamic flow. They, therefore, act as boundary conditions for the processes that they constrain by limiting the possibilities of the outcome of the process.

[169] Montévil and Mossio (2015), p.182.

Given this conception of constraints, Montévil and Mossio (2015) present their idea of dependency of constraints. A constraint can be said to be dependent on another constraint if its production is dependent on the actions of another constraints – i.e. the generative constraint (the constraint which produces another constraint) is in some way (for example constraining another process that leads to the formation of the dependent constraint) responsible for the creation or maintenance of the dependent constraint. Montévil and Mossio (2015) define the dependency of constraints as –

> "A relationship of dependence between constraints as a situation in which, given two-time scales τ_1 and τ_2 considered jointly, we have:
>
> 1. C_1 is a constraint at scale τ_1,
> 2. There is an object C_2 which at scale τ_2 is a constraint on a process producing aspects of C_1 which are relevant for its role as a constraint at scale τ_1 (i.e. they would not appear without this process).
>
> In this situation, we say that C_1 is dependent on C_2, and that C_2 is generative for C_1."[170]

The dependence of constraint on another constraint becomes apparent only when a sufficient time scale is considered. In case of the dependence of constraints on another, the generative constraint is causally responsible for the creation of the dependent constraint (or some feature of it) not directly, but through constraining the processes that lead to the generation of the dependent constraint. On time scale τ_1, constraint C_1 is conserved while constrains the process $A_1 \rightarrow B_1$, but it needs to be repaired at a different time scale (τ_2), which is where the generative constraint C_2 acts such that it leads to the

[170] Montévil and Mossio (2015), p.184.

production and maintenance of C_1 (see *figure 4.5*). Therefore, the generative constraint maintains the dependent constraint over a time scale different than the dependent constraints active time-period.

Lastly, Montévil and Mossio (2015) define the closure of constraints. The idea of closure is that constraints, construed as above, realise an organisation such that each constraint plays both generative and dependent roles, such that all the constraints are at least dependent on and generative for one other constraint –

> "*A set of constraints C realises overall closure if, for each constraint C_i belonging to C:*
>
> 1. *C_i depends on at least one other constraint belonging to C (C_i is dependent);*
> 2. *There is at least one other constraint C_j belonging to C which depends on C_i (C_i is generative).*
>
> *A set C which realises overall closure also realises strict closure if it meets the following additional condition:*
>
> 3. *C cannot be split into two closed sets.*"[171]

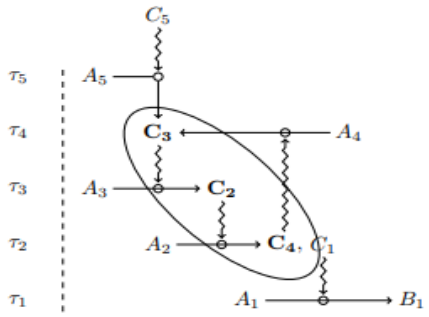

Figure 4.5 – Closure of constraints. The closure takes place between C_3, C_2 and C_4 due to realising interdependencies such that each is the generative constraint for and dependent on another constraint [from Montévil and Mossio (2015), p.186].

[171] Montévil and Mossio (2015), p.186.

There are two points to be noted here. Firstly, the time scale at which the closure is considered is larger than the action of a single constraint. At the time scale of a single constraint, the closure cannot be observed. Secondly, the set of the constraints that is said to bring about closure must not be divisible into subsets that are themselves closed – in which case we will have two closed sets rather than a single closed set.

In *figure 4.5*, C_1, C_2, C_3, C_4, C_5 satisfy the condition to be called constraints functioning at time scales τ_1, τ_2, τ_3, τ_4, τ_5, such that C_5 is generative for C_3, C_3 is generative for C_2, C_2 is generative for C_4 and C_1, and C_4 is generative for C_3. Therefore, the set of constraints C_3, C_2, C_4 form a closed set such that there is a cyclical dependency relation where C_3 is generative for C_2, which is generative for C_4 which in turn is generative for C_4. Therefore, the set C_3, C_2 and C_4 realise a closure such that they are inter-dependent for the maintenance of each other. Furthermore, that C_4 (which is generated at the time τ_2) is generative for C_3 (which is generated at τ_4) only make sense when we consider a cyclical existence where C_4 generated in the previous cycle affects is generative for C_3.

Closure of constraints thus leads to an organisation that is causal for its own maintenance. This is not to say that closure of constraints realises independence from the surrounding. It is still an open system that exchanges matter and energy with its surrounding, and thus is constrained by the environment itself (for example C_5 in the hypothetical system considered above is an external constraint acting upon the closed system of C_3, C_2 and C_4), but the closure is synonymous to what Rosen (1991) called

'closed to efficient causes'[172]. That is to say, the constraints act as a boundary condition on each other through the processes they constraint, and not material causes which do not realise a closure. Therefore, while the system exchanged material and energy with the environment, it is closed to efficient causes for its maintenance.

The closure of constraints idea is thus a refinement and formalisation of ideas previously studied (Gánti's Chemoton and Rosen's closure of efficient causes). Firstly, it distinguishes processes and constraints such that two separate causal regimes are constructed. The first regime is the causal regime that the constrained processes realise. This regime brings about the thermodynamic flow that the system can utilise to realise its essential functions like metabolism. The second regime is the closure that the constraints realise through mutual dependence. Secondly, it defines organisational closure in terms of constraints and not the processes. While processes can minimally constrain each other, the constraints that achieve closure are not thermodynamically coupled, at the relevant time scale, to the processes they constrain. The closure of constraints framework thus differentiates between constraints that are inherent in the processes themselves and control constraints. It is important that the system which is said to be self-determining achieved the closure of the second type of constraints.

The closure of constraints also defines a minimal autonomy of the biological organisation. Ruiz-Mirazo and Moreno (2004) define autonomy of system as -

> "the capacity of a system to manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control:
>
> (i) internal self-constructive processes and

[172] Rosen (1991), p.244

(ii) processes of exchange with the environment. Thus, the system must be able to generate and regenerate all the constraints—including part of its boundary conditions— that define it as such, together with its own particular way of interacting with the environment."[173]

Thus, the idea of minimal autonomy of biological organisations is inherently linked to organisational closure which achieves collective self-maintenance and self-determination while maintaining an organisation open to the thermodynamic flow of both matter and energy. This requires both an organisational closure of constraints and a regulation of this closure.

In the next section, I analyse if constraints require a concept of emergence, and how should the emergent constraints be construed? I show that Moreno and Mossio's (2015) conception of emergence falls short of characterising constraints and must be supplemented with Wilson's (2010) idea of 'Weak Ontological Emergence' through the reduction, restriction, and elimination of Degrees of Freedoms. The motivation behind doing so is to establish a robust ontological picture of configurations as constraints, such that these configurations can be seen as causally efficacious in terms of their influence over the processes. This helps ground functional organisations in biology and thus contributes significantly to debates about non-reductive physicalism in biology. In addition, this provides a conception of constraints that will be useful in modelling robustness of biological systems through second-order control on the first-order processes, thus providing a basis to understand regulation as ubiquitous feature of biological system. This will form the basis for design explanations that explain why certain functions are performed in particular ways in biological systems.

[173] Ruiz-Mirazo and Moreno (2004), p.240.

4.3: On Emergence

Would a closure of constraint account require emergence? Mossio, et al (2013); Moreno and Mossio (2015) believe that it does. Since constraints are configurations that are not reduced to their underlying thermodynamic flow on which they act (at relevant time scale), they require some justification for this. If configurations act as constraints, we must account for how these configurations come about, and if their causal power is reducible to their components. As discussed in the previous section, the constraints are not merely the constraints realised by processes but are present over and above these processes. Therefore, a theoretical justification must be provided for their causal powers. For example, both enzymes and cell membranes are made up of several entities. The enzyme has a chain of amino acids arranged in a specific way. The cell membrane is composed of lipids arranged in a specific way. The question, thus, is that what confers such configurations the causal powers that are irreducible to their individual parts?

Secondly, the existence of constraints is tied to their realising closure such that they contribute to the existence of the organisation. Constraints individually cannot maintain themselves but are instead dependent on other constraints to maintain them such that closure is achieved. Furthermore, this property of achieving closure provides a functional grounding to the constraints. This property is itself dependent on constraints achieving an emergent causal regime such that they are closed despite the processual thermodynamic flow that they act upon. Thus, the emergent property of closure grounds the functionality in the biological domain. Therefore, a concept of emergence that can satisfy both configurations acting as constraints and the constraints realising closure to give rise to an emergent biological organisation can help ground biological function.

Emergence as a philosophical term of art has a long history dating back to Mill (1843) and a subsequent program of more sophisticated studies by the British Emergentist [Broad (1925), Alexander (1920)]. The idea of emergence is simple – the world is composed of layered structures where the fundamental physics and the ontologies defined by it sit at the bottom and the upper layers are formed of organised structures that are composed of the components that lie beneath it. These upper layers form structures that are studied by special sciences (chemistry, biology, sociology, etc.). The upper levels are formed of the lower-level components that are governed by the lower-level laws. Emergentism maintains that the upper-level ontologies have causal powers that are the focus of attention of special sciences (including special science laws). The puzzle of emergence is thus, that while maintaining a commitment to physicalism (that everything in the world is formed of fundamental matter), how can we assign causal powers to upper levels when they are explainable by reducing them to the ontologies and laws of the lower level? The objection is based on the idea of the *Causal Exclusion Principle* – i.e. all physical effects are due to physical causes and nothing over and above that. If we grant that the lowest level (whatever it may be) is causally closed and therefore sufficient to draw causal inferences, it seems that every cause and effect relationship between the upper levels is epiphenomenal. Therefore, it presents a problem for not just emergentism, but for the special science ontologies and laws.

A formal statement of this problem is due to Kim (2006) who notes that emergence requires two conditions – Supervenience and Irreducibility. Supervenience is the relationship between the emergent level and the level it emerges on (base level). An emergent property is said to be supervenient on the properties of the base level such that the property of the emergent whole is determined by the properties of the basal realisers

and their relationships. Therefore, it establishes a necessity for the relationship between the emergent and basal levels such that similar basal conditions (components, relations, and laws) must give rise to similar emergent properties. As Kim (2006) notes –

> "*Supervenience:* If property M emerges from properties N_1, ..., N_n, then M supervenes on N_1, ..., N_n. That is to say, systems that are alike in respect of basal conditions, N_1, ..., N_n must be alike in respect of their emergent properties."[174]

At the same time, as Kim notes, the emergent level is taken to be irreducible to the realiser base. This irreducibility is generally cashed out in terms of functional or causal terms. As Kim (2006) notes –

> "*Irreducibility of emergents:* Property M is emergent from a set of properties, N_1, …, N_n, only if M is not functionally reducible with the set of the N_s as its realiser."[175]

These two conditions, taken together as logical necessities for emergence, lead to a contradiction when we try to explain cause and effects in terms of emergent properties. As Kim (2006) notes, the two necessities lead to epiphenomenal explanations such that the explanation in terms of both emergent properties and basal properties overdetermines the causes of an event. To understand this, consider an emergent property m causes another property m* to instantiate. m has a realiser base p with its p-level property. Similarly, m*, since it is on the same level as m, has a realiser base p*. Therefore, to cause m*, m has to cause p* which realises m*. But p is nomologically sufficient for p*. Therefore, to cite m as the cause for p* overdetermines the causes.

[174] Kim (2006), p.550.
[175] Kim (2006), p.555.

Therefore, m causing m* is epiphenomenal. This is what Kim calls the exclusion argument against emergence.

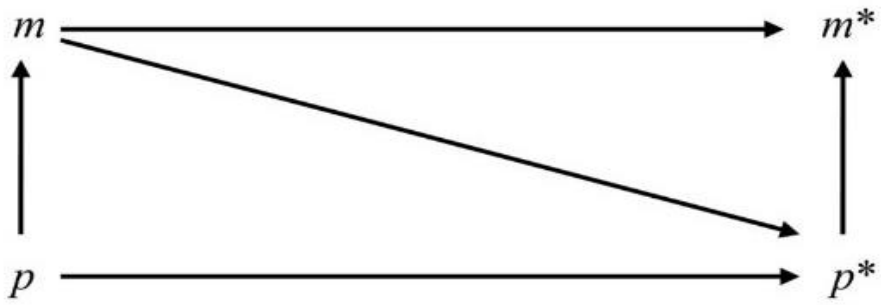

Figure 4.6 – The causal exclusion argument states that property m which emerges on property p, over determines p since p is causally sufficient for p*.*

As Mossio, et al. (2013) correctly note, this has consequences for the special science ontology and causation. If the emergent properties of special sciences ontologies cannot be said to have causal powers that do not turn out to be epiphenomenal, then they must be reduced to their realiser base. This gives rise to a reductive understanding of the world, where the special science ontologies must be reduced all the way down to the fundamental level properties, least they will be epiphenomenal. Kim's argument gives rise to Reductive Physicalism, where the causal analysis of any phenomenon should be logically reducible to the most fundamental level of description and fundamental laws. This makes special science causes to be mere heuristic tools. And therefore, the special science explanations (including biological) are themselves reduced to not tracking actual causes but to a heuristic strategy.

Mossio, et al (2013) present an argument for emergence that according to them, can counter the exclusion argument. In the next section, I analyse this argument and note that it lacks in some regards. Following this, I present another argument for weakly

emergent properties as presented by Wilson (2010) which works better as both a reply to the exclusion argument and as a way to characterise configurations.

4.3.1: A Defence of Emergence

Mossio, et al. (2013)'s approach to defend emergence is based on a two-pronged approach. Firstly, they elucidate the difference between epistemic non-derivability and ontological irreducibility. Emergentists many times talk about two different criteria of how to judge if a property is emergent. On the epistemic front, they argue that the said property is non-deducible (or not predictable) from the basal properties and laws governing them. Bedau (1997), for example, argues that complex systems show macroscopic patterns that can, in principle, be derived from the knowledge of the microstate but only through simulation – i.e. modelling the all systematic interactions leading up to the desired state from its initial state. In Bedau's example of Conway's Game of Life where the state of any cell (in a board) is dictated by its surrounding eight cells. Bedau (1997) argues that the pattern obtained after any given time can be derived from the knowledge of the rule of the game and initial conditions. But that a complex pattern (like a glider gun – involving a gun-like shape that moves across the grid and emits bullets) cannot be predicted unless a complete simulation is run.

Ontological irreducibility, on the other hand, captures the idea that the emergent property is novel with regards to its realiser base properties. This is not to say that they are unpredictable (which they might very well be), but that the emergent property thus obtained is novel and irreducible to the properties of the realiser base. One way to cash

this out is to say that the emergent properties have causal powers irreducible to the basal realisers.

Mossio, et al. (2013)'s aim is to defend a view of ontological irreducibility of biological entities that can work as constraints. Therefore, keeping the issue of epistemic non-derivability aside, let us consider their defence of emergence construed as ontological irreducibility.

Faced with Kim's exclusion argument, Mossio, et al (2013), following both Kim (1998), and Campbell and Bickhard (2011), argue that supervenience must be conceptualised not as atomistic but as relational. The argument hinges on the idea that the supervenience base for the emergent property is not the lower-level entities and their intrinsic properties either taken individually or as a collection, but the configurations of these entities with both their intrinsic properties and relational properties taken together. The crux of the argument is that emergent properties are micro-based macro properties[176], i.e. second-order functional properties that are realised by the first-order properties and their relations. By shifting the supervenience base from atomistic realisers to relational realisers associated with configurations, these very configurations become the locus of emergent powers. Campbell and Bickhard (2011) go further in claiming that the relations of the configurations (or the wholes) are not micro-properties – i.e. they are not available at the lower levels. Thus, the supervenience relation between lower-level components and configurations does not hold. Instead, the supervenience relationship holds between the configurational properties (which include both the intrinsic properties of the components and their relations) and the emergent property. In such a case, Mossio, et al (2013) and Campbell and Bickhard (2011) maintain that

[176] Kim (1998), p.85.

supervenience and irreducibility are not held between the same levels. The emergent base is supervenient and hence synonymous to the configuration and is irreducible to the components taken separately or collectively (without the necessary relations to be in the said configurations). While the emergence relation holds between the configuration (including relational properties) and the micro-level entities that are part of the configurations but are not held in the necessary relational configurations.

Mossio, et al (2013) claim that this appeal to relational mereology supports emergence and avoids epiphenomenalism and causal drainage. Configurations are irreducible to their constituents taken separately but are emergent on their intrinsic properties. Emergent powers of the configurations are nothing over and above the properties of the configurations but are emergent on the properties of the constituents taken separately. They present their own conception of emergence that supports their closure of constraints view as –

1. "The configuration C is not supervenient, yet is emergent on the properties of any proper subset P_{sset} of its constituents."[177]
2. "The configuration C is not supervenient, yet is emergent on its substrate P_{sstr}, i.e. the collection of its constituents taken separately as if they were not constituents."[178]
3. "The configuration C is not supervenient, yet is emergent on its surroundings P_{surr}, i.e. each set of external elements that does not actually constitute C."[179]

These three conditions frame emergent properties as that of configurations emerging on the properties of its constituents (when taken separately or as an unrelated group), on

[177] Mossio, et al. (2013), p.162.
[178] Mossio, et al. (2013), p.162.
[179] Mossio, et al. (2013), p.162.

the subset of the constitution that is not similar to the whole constitution, and on the surroundings of the configuration. But it is worth noting that neither Mossio, et al. (2013), [nor Campbell and Bickhard (2011)] have yet defined what configurations actually are and how do they emerge on their constituents, and by extension – what relations must the constituents be engaged in to be qualified as a genuine emergent configuration? While Mossio, et al. (2013) note that their conception of emergent configurations is very commonplace and includes many 'uninteresting cases', it is not enough to characterise configurations. In theory, the chair I am sitting on, the north star, and the pen kept on my desk make a unique configuration, but it can hardly be called emergent or having causal powers. To be able to define configurations as emergent, one needs a criterion that separates the random cases from those that show genuine emergence. One such way is to emphasise the idea of emergent causal powers and another is to characterise relatedness. Both these are not extensively tended to by Mossio, et al. (2013). Secondly, while (1) and (2) define emergence as the relation between the emergent base and its compositional base, (3) includes the surrounding entities as something the emergent base emerges upon. This seems counter-intuitive. Firstly, the relation between the emergent base and the entities surrounding it is that of cause and effect and not of constitution. Mossio, et al. (2013) use wheel as an example. According to them, the wheel is emergent upon the molecules that form if taken separately (1), a wheel is emergent on half a wheel (2), and the wheel is emergent on its surrounding, specifically the surrounding it has causal influence on (3). But unlike (1) and (2), (3) does not naturally follow. A wheel has causal influence over the ground it rolls on, but is it emergent on the ground?

In what follows, I show how both these objections can be utilised to enrich the concepts provided by Mossio, et al (2013). I use views put forward by Wilson (2010) to argue that while Mossio, et al (2013) are correct in pointing out configurations as the focus of emergence, they must supplement it with stronger ideas of relatedness and causal power to establish an emergent regime that can support their wider picture of closure of constraints.

4.3.2: Degrees of Freedom

How to characterise relational mereological supervenience such that it explains emergent configurations from the basal components? Wilson (2010) puts forward an accounts of emergence that is helpful in answering this question.

Wilson's major focus is on how Non-Reductive Physicalism (NRP) can be defended against reductionism. NRP as a metaphysical thesis applied to sciences holds that (at least) some special science entities are distinct from and "distinctively causally efficacious with respect to any lower-level physically acceptable entities"[180]. To defend NRP against Reductive Physicalism (which denies the distinct existence of special science entities from their physically acceptable realisers) emergence is utilised as a metaphysical thesis to justify the existence of higher-level causally efficacious entities. Therefore, NRP-ists must show that there is something irreducible about the entities whose existence they defend. Wilson (2010) argues that special science entities can be thought of as ontologically emergent if we consider how the emergent entity is brought about by the elimination of degrees of freedom of the basal entities. Degrees of freedom

[180] Wilson (2011), p.122.

are the parameters that are needed to characterise an entity of being in a certain state. As Wilson (2010) writes –

> "*Degree of Freedom (DOF):* For an entity E, characteristic state S, and set of coordinates C, a DOF is any parameter in a minimal set, expressed in coordinates C, needed to characterise E as being in S."[181]

Wilson (2010) notes that DOFs are required to define the functional state of the entity. These states can be defined with various DOFs in different states like –

a. Configurational state (defining the position of the entity in space),
b. Kinematic state (defines the velocities of the entity in the configurational state), and
c. Dynamic state (tracks the energy of the entity like potential and kinetic energies).

In many cases, Wilson (2010) notes, to characterise a certain entity much fewer DOFs will be required than in other cases. A point particle which is constrained to move in a two-dimensional plane, for example, will require only two configurational DOFs rather than three. This can be seen in rigid bodies. Let us say that two entities are equidistant from each other. In such a case, only five configurational DOFs will be required to characterise the system since one DOF of one of the entities can always be derived from the DOFs of the other. And similarly, less kinetic, and dynamic DOFs will be required. This is what Wilson (2010) calls a *reduction in DOFs*, where less DOFs are required to describe the system. Similarly, we can have *restrictions in DOFs* – where the DOFs are not reduced but are constrained to remain within a specific limit. For example, a point

[181] Wilson (2010), p.282.

particle which cannot move beyond a specific limit defined by two planes has restricted configurational DOFs which cannot take values on either side of the planes but have values only between the values of the limits defined by the planes. Lastly, *elimination of DOFs* happens when the certain DOFs are eliminated from the system such that they are no longer needed to characterise the system. For example, the electric field associated with spherical conductors requires only the DOFs of the particles at the boundary of the sphere. The DOFs of other particles do not matter and hence can be eliminated from the description of the system.

Wilson (2010) notes that only elimination of DOFs leads to weak ontological emergence (as a sufficient condition). But weak ontological emergence so characterised, will exclude many interesting cases of emergence. For example, atoms and molecules under such consideration will not be considered emergent. This is because atoms and molecules only require restriction and reduction in DOFs. An atom is characterised by reduction of DOF of an electron in a particular orbit, and restriction in DOFs for protons and neutrons which form the nucleus. Similarly, molecules will require only reduction in DOFs. While Wilson (2010) is interested in more complex cases, I believe that reduction and restrictions in DOFs can also characterise emergence. While Wilson might be correct in pointing out that not all cases of restriction and reduction of DOF lead to interesting cases (a tabletop made of wood does reduce the degrees of freedom of the molecules in it but does not form an interesting case of emergence). But by limiting to elimination, Wilson (2010) loses some interesting cases that do rely on reduction and restriction alone. Furthermore, restrictions and reduction might lead to near elimination conditions. If say, the reduction of multiple DOFs (unlike the point particle example) or stringent restriction are placed on an entity, on epistemic

considerations, they can be considered as eliminated. It seems then that elimination can be seen as a higher degree of restriction and reduction. What will be such interesting cases of reduction and restriction is, of course, an epistemic endeavour. But to remove such interesting cases seems unjustified. Since the DOF based emergence is a sufficiency condition, we can still have a different condition that can specify what separates the interesting cases from the uninteresting. Some authors have focused on non-linearity [Bedau (1997)][182], and complexity of the relations to justify these differences. But I will not get into that debate any further. For now, the idea that reduction, restriction and elimination of DOFs can give rise to weakly emergent properties, is sufficient to explain Mossio, et al (2013) characteristic configurations.

Configurations, as mentioned by Mossio, et al. (2013), can be seen as reduction, restrictions, or elimination of DOFs. Configurations, by the virtue of constraining their components in such a way that the DOFs required to characterise the configuration are restricted, reduced or eliminated as compared to the DOFs required to characterise their components when taken separately or in collection without the relations that are characteristic of the configuration, are weakly ontologically emergent on the lower-level components. For example, an atom is emergent on the electrons, protons, and neutrons that form it by restricting and reducing the DOFs. An electron is always present in an orbit which is equidistant from the nucleus such that the DOFs required to represent it are reduced. Similarly, the DOFs required for characterising the protons and neutrons are restricted to be inside the nucleus. Hence, when defining a state of the atom, we require fewer DOFs to characterise it as compared to when they are not a part

[182] Badeu (1997) argues that non-linear systems are not merely additive in nature – i.e. their resultant property is not merely the sum of some aspects of their individual properties. In such cases, Badeu (1997) maintains that the properties are candidates for emergence.

of the atom. Furthermore, the relations can be explained using physical theory – like strong nuclear forces. Thus, the description of an atom is physically acceptable, but at the same time, the atom is weakly emergent because of the restrictions and reductions in DOF.

Similarly, biological entities (entities treated by the domain of biological theories) are emergent because they are composed of configurations of entities composed in such a way that they restrict, reduce, or eliminate the DOFs of the entities that they are formed of. A cell membrane is made of lipid molecules arranged in such a way that they form a barrier, in the process. The cell itself is made up of various organelles which are made up of molecules arranged in a specific way. Thus, the entities are weakly ontologically emergent.

Some clarifications about emergence construed in this way are in order –

1. The concept of emergence so construed, is a *realised-realiser* relation, i.e. the emergent base is realised by the lower-level base. While there is some debate about what realisation entails [Craver and Wilson (2006), Baysan (2015)], the basic gist is that the realised base having a property or activity Q, is realised by the realiser base having a property or activity P, such that the relation is asymmetric (i.e. having P entails having Q, but not vice versa). Therefore, the realisation relationship holds a dependency relation where having P is sufficient but not necessary for having Q.
2. Because of the realisation relationship, the causal powers of the realised level are nothing over and above the causal powers of the realiser base. That is to say, that there are no novel causal powers that may be conferred to the higher-level. Instead, powers of the realised level are merely realised on the higher-

level. Wilson (2011), for example, states that the realiser-realised relation is a subset realisation where the realised level has a proper subset of the token powers of the realiser level –

> *"Subset Condition on Causal Powers (SCCP):* The token powers of a realised [mental] state M on a given occasion are a non-empty proper subset of the token powers of the (lower-level relational, physically acceptable) base state P realising M on that occasion."[183]

By adopting a token-identical view of powers for the realised base, Wilson (2011) argues that the realised level has a distinctive causal profile which is a proper subset of the realiser level power profile. Since the profile is distinctive from the realiser level, it makes the realised level causally autonomous.

3. The SCCP condition, as Wilson (2011) notes, can account for functional realisation. Functional realisations identify the causal role of the realised state with the function it performs. The functionally realised states have causal powers that are context dependent. That is, the function is defined in the wider context of their existence. For example, the functional property of having pain is realised by the C-fibre stimulation only exists in the wider context. A C-fibre in a petri-dish cannot be said to realise the pain. Craver and Wilson (2006) note that the appeal to properties of the realised state needs a wider background condition or 'a world beyond it'.

   Secondly, the functions can be multiply realised. As Wilson (2011) notes, when talking about functional realisation with multiple realizability, such that the lower states are composed of different components from each other, the

[183] Wilson (2011), p.128

> realised states can still have a similar functional profile because the realising types each will have the powers associated with the realiser type (along with other different powers). The functionally realised states will be of similar type because they will have token causal powers of the realisers which are similar in at least holding the similar powers that form the set of causal powers of the realised state, even if their whole power profile is different.

Conceptualising emergence of configurations in terms of DOFs, thus brings forth some interesting properties –

Firstly, the configurations are realised by their constituents such that the realisation shows the qualities of having real causal powers which are only a subset of the powers of the lower-level realisers. This can be read as actual powers versus possible powers. The lower-level realisers can realise many more possible configurations/orientations (both real and hypothetical) than the configuration in question. But what provides the grounding to the configuration in question is that it is actually realised and has actual powers.

Secondly, the realised base can be multiply realisable – i.e. realised by multiple realiser bases. Therefore, configurations can be defined as a type of emergent base if they have the same properties (manifested in causal powers) irrespective of their composition. In case of configurations as constraints, the configurations have the causal powers of exerting causal influence on the surrounding processes. Thus, they can be defined as constraints irrespective of their different composition.

Lastly, the idea of self-determination and autonomy itself is an emergent property of an organisation of constraints achieving closure. Since the organisation is nothing over and above the constraints organised in such a way that they are constrained by other

constraints (hence reducing, restricting or eliminating their DOFs), we have a conception of self-determination based on the idea of the organisation realising a specific type of configuration.

Mossio, et al. (2013) conceptualise their idea of emergence in terms of configurations. But they offer very little in terms of what configurations are. As I have argued, the reduction, restriction, elimination of DOFs offers a better picture of what configurations are. Combined with the SCCP conceptualisation of realiser-realised relation and functional realisation, their idea of emergence becomes more robust. Firstly, the emergent property is emergent on the properties of the unrestricted properties of the lower-level components when taken separately. This is because the unrestricted components need all the DOFs to characterise them. Secondly, the emergent property is emergent upon any subset of the configuration because not all constraints that define the emergent base are in place. Lastly, although Mossio, et al. (2013) present it as emergent relation, the emergent base has causal powers over its surrounding because of it having a functional realisation that has a context-dependent causal role which can be exerted on its surrounding. The context-dependency in case of the autonomous biological system is provided by the condition of closure.

Thus, adopting Wilson's idea of emergence as reduction, restriction and elimination of DOF provides a conceptual tool for explaining configurations as constraints, the functional realisation of these constraints, and the organisation as a whole.

4.4: On the Explanatory Import of Constraints and Control

In this chapter, I have argued that the closure of constraints idea offers a way to think about minimal autonomy of biological systems such that they are able to constrain their own processes and realise self-maintenance of the constraints themselves. In this section I argue that that the constraints-control dynamics, when looked from the perspective of the minimal autonomy account argues for understanding control as a ubiquitous phenomenon in biological systems that is realised in service to preserving a steady state dynamics grounded in thermodynamics of the system and can be explained by appeal to design principles [Wouters (2007), (2013)].

The structure of the section is as follows. First, I introduce the idea of control exerted by constraints in different ways in biological organisation – stoichiometric, catalytic, and regulatory. I further show that these constraints can be acted upon by other constraints such that it leads to the preservation of minimal autonomy of the system and, in some cases, novel adaptive behaviour. Finally, I show how this idea of control can be grounded in the thermodynamics of the system which explains why certain controls will be exerted in a certain way to increase the adaptive response of the system to the environment. This is achieved not by offering a causal-mechanism explanation of the control itself, but by comparing them to hypothetical or real systems that fail to show the same behaviour in their absence. Finally, I take the example of glycemia regulation to show how all these concepts fit together.

4.4.1: Control – Stoichiometric, Catalytic, and Regulatory

The closure of constraint idea of minimal autonomy introduced in this chapter lies at the heart of minimal organisational stability of the biological organisation through the establishment of dependence between (at least a set of) constraints within the organisation. As has been noted, the minimal organisation so realised shows stability by controlling its own processes through stoichiometric coordination and realising constraints on processes. Control can be executed in various different ways in the basic constitutive regime of the organisation through kinetic, spatial and template control [Bich (2016)] –

Kinetic control is control on the kinetic processes such that they can be directed towards a specific state. Enzymes, for example, exerts control on a reaction by pushing it towards the forward direction.

Spatial control is exerted by defining spatial limitations on processes. For example the cell membrane constrains the motion of molecules in and out of the cell and thus helps maintain a specific concentration that is needed for the functioning of the cell (the specific concentration is maintained by selective influx and outflux of material and is not a behaviour but a dynamic one).

Template control is exerted by specifying the sequence of molecules that are required by the system at a specific time. For example, the DNA, RNA and related mechanisms that code for and produce amino acid sequences through a specific coding-decoding scheme act as template controls.

These minimal control mechanisms help the system to maintain itself by exerting constrains on processes and thus directing energy to processes in an organised manner.

All these controls, realised by themselves, lead to a stoichiometric dependency between constrained processes such that they achieve a minimal self-maintaining organisation. But the minimal organisation thus established is open to being disrupted when perturbed by the environment and failure of internal dynamics. For example, the Ganti's chemoton presents all the above-mentioned controls between subsystems. But the chemoton is open to disruption by perturbations in both internal and external environment. Failure of the membrane subsystem can lead to the failure of metabolic subsystem by disrupting the delicate stoichiometric balance that is required for it to function. And as a result, put the whole organisation is jeopardy. To counter these kinds of perturbations, biological systems show strategies that can be called control over the minimal organisation.

As argued in the last section, constraints are emergent configurations that exert causal influence on their surroundings (processes) due to reduction, restriction and elimination of DOFs of their own components. Therefore, constraints can be altered by bringing about a change in their configurations such that their emergent causal profile changes. An example of this would be allosteric enzymes that change their configurations due to binding of substrates. This further leads to change in their behaviour. The change in constraints, therefore, is able to change how the processes that they affect, such that regulation is possible. Regulation thus forms a keyway of control in biological organisations. This can further enable biological systems to self-maintain in the face of changing external conditions through the exertion of the right constraints at the right time, thus providing them with robustness.

Control as defined within the closure of constraints view, is the way in which the system alters its own constraints to realise a minimal autonomy such that the closure can sustain itself over time and in face of perturbations.

To understand this, let us consider the example of glycolytic pathway [Winning and Bechtel (2018)]. The glycolytic pathway of glucose metabolism is a system consisting of various processes that leads to the conversion of glucose into pyruvate and in the process phosphorylates ADP to ATP. ATP in turn, acts as an important power source for cellular functions, including production of the very catalysts that are used in the glycolytic pathways. The first level of control in the system is purely stoichiometric. ATP, as a product of the glucose metabolism is important for the production of the enzymes used by the metabolic pathway itself. Thus, establishing a cyclical dependency. The first level of control can be exerted by the system through stoichiometric control – i.e. changing the concentrations of enzymes to alter the stoichiometry of the metabolic process. The second level of control, though, is exerted by control on the enzymes acting as constraints on the metabolic process itself, not through stoichiometric variation of the presence or absence of enzymes, but through the dedicated subsystems acting on the catalytic constraints themselves.

To understand this, consider the first three steps of the glycolytic pathway –

1. Phosphorylation of glucose into Glucose-6-phosphate which is facilitated by the enzyme Hexokinase. Hexokinase is inhibited by Glucose-6-phosphate.
2. Conversion of Glucose-6-phosphate into Fructose-6-phosphate.
3. Phosphorylation of Fuctose-6-phosphate into Fructose-1,6-biphosphate through the action of Phosphofructosekinase (PFK).

The PFK works as an allosteric enzyme such that its configuration can change in accordance with substrates that bind to it. When the concentration of ATP in the cell increases, the phosphorylation of PFK inhibits its catalytic action, thus stopping the phosphorylation of fructose-6-phosphate to fructose-1,6-biphosphate, thus stopping the

downstream pathway of conversion of glucose to pyruvate and stopping the production of more ATP. When ATP is low, the cell can initiate a conversion of ADP to ATP through the enzyme adenylate kinase according to the reaction –

$$ADP + ADP \rightarrow ATP + AMP$$

The AMP produced in this reaction leads to activation of PFK thus leading it to facilitation of phosphorylation of fructose-6-phosphate to fructose-1,6-biphosphate. As a result, the glycolytic metabolism can continue leading to generation of ATP.

The inhibition of PFK leads to more inhibitory activities. Since the Glucose-6-phosphate and Fructose-6-phosphate are in equilibrium, an increase in the latter leads to an increase in the former. As the concentration of glucose-6-phosphate increases, the hexokinase is inhibited and the phosphorylation of glucose to glucose-6-phosphate is stopped. Therefore, the conversion of glucose to glucose-6-phosphate is inhibited (see *figure 4.7*) This prevents the cell from utilising the glucose at all. Thus, preventing wasteful use of glucose, thus preserving glucose for other cells that might need it.

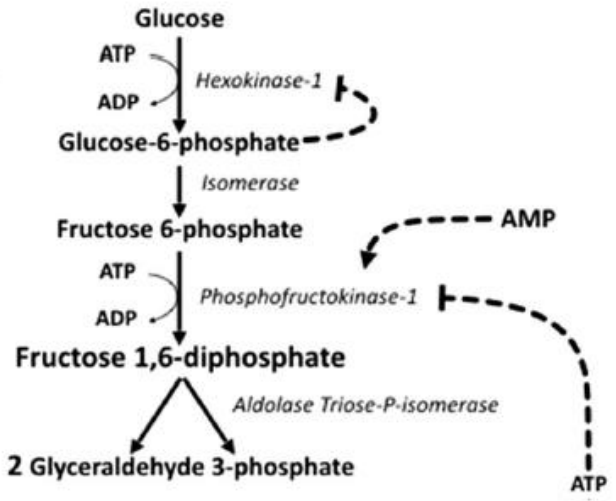

Figure 4.7 – Control of glycolytic pathway. ATP inhibits PFK while AMP activates PFK. Hexokinase is inhibited by access concentration of Glucose-6-phosphate. Therefore, when ATP is in access, the

glycolytic pathway is inhibited by ATP's action on PFK, when ATP is deficit, the pathway is activated. Adapted (with changes) from Winning and Bechtel (2018), p.301.

While the glycolytic pathway is under the control of various other enzymes, which are produced by the cell and thus can constrain the pathway at different by their availability or inhibition/activation, the PFK control is of high importance. As Berg, Tymoczko, Stryer (2002) argue in their introductory book to biochemistry, the phosphorylation of fructose-6-phosphate to fructose-1,6-phosphate represents the first irreversible step in the glycolytic pathway such that the reaction commits the cell to the glycolytic pathway, it represents an all or nothing effect such that the fructose-1,6-phosphate so produced has no other use but to be acted upon by other enzymes to pyruvate (as the glucose-6-phosphate can be used up in glycogen production[184], the production of glucose-6-phosphate is not necessarily a commitment to the metabolic pathway). It becomes important to stop this reaction to stop the glycolytic pathway at this step to prevent the metabolism from carrying out. Thus, preventing wastage of cellular resources and frivolous production of ATP.

Glucose metabolism thus shows the features of control that can be exerted by constraints through mutual dependence on the thermodynamic flow of the system.

Firstly, the inhibition/activation of PFK is a consequence of change in concentrations of various downstream products of the same closely related processes such that increase

[184] In certain cells (like liver and muscle) of multicellular organisms, the inhibition of phosphorylation of glucose to glucose-6-phosphate by hexokinase leads to another effect. These cells have an enzyme glucokinase which is not inhibited by high concentrations of glucose-6-phosphate like hexokinase. In the absence of hexokinase's catalysing action due to non-requirement of ATP as dictated by PFK, the glucokinase continues to phosphorylate glucose to glucose-6-phosphate. But, instead of entering the glycolytic pathway the glucose-6-phosphate instead is directed towards production of glycogen (a storage form of glucose), therefore leading to storage of excessive glucose as glycogen for future use.

in the concentration leads to an inhibitory effect. The PFK thus, represents a constraint on the glucose metabolism pathway such that it is maintained and controlled by other processes in the system that are themselves a downstream product of the constraining action of PFK.

Secondly, it shows the system's capacity to shift between processes aided by alteration in constraints that realise them in order to maintain a homeostasis through dynamic stability such that the constitutive regime can be maintained.

Thirdly, it explains minimal autonomy of the system to alter its internal dynamics in face of internal changes and needs of the system.

But there are limits to this. A huge variation, either internal or external, may lead one or more constraints to fail such that closure is not realised. In such a case, the organisation will seize to exist. To compensate for this, Moreno and Mossio (2015) note that in addition to 'local robustness' due to organisational closure, biological organisations also possess a 'global response' which is brought about by a dedicated regulatory system maintained by regulatory constraints that act on the constitutive constraints. Regulatory constraints have two important characteristics –

1. Regulatory constraints form a class of constraints which are not constitutive. They are what Moreno and Mossio (2015); Bich (2018) call *second-order constraints* – i.e. constraints that act causally on the constitutive constraints. In this way, they regulate the function of the constitutive constraints when required. During normal functioning, the regulatory constraints are not active. They become active only when the first order constitutive constraints are

disturbed by perturbations, such that they can help the system shift from one constitutive organisation to the other (see *figure 4.8*).

2. The regulatory constraints are still maintained by the first-order processes and constraints on a relevant time scale, but not in the same way as constitutive constraints are. Regulatory constraints realise a second-order closure such that they bring about (many times reversible) changes in the first-order regime, and thus realise a closure with the separate regimes rather than localised processes of a singular regime (*figure 4.8*). The closure is realised between the first-order causal regimes C_1, C_2… C_n and C_1', C_2'… C_j and the regulatory constraints that bring about the transition from the first to the second.

Regulatory systems, in this view, are not part of the constitutive regime that realises closure of constraints, but additional constraints that act on this regime while being decoupled from the stoichiometry of the constitutive regime itself. The maintenance of these constraints still depends on the realisation of the constitutive regime, but they are not directly dependent on the state of the constitutive regime. In such way, the Regulatory subsystem shows an independence in its functioning to the constitutive regime.

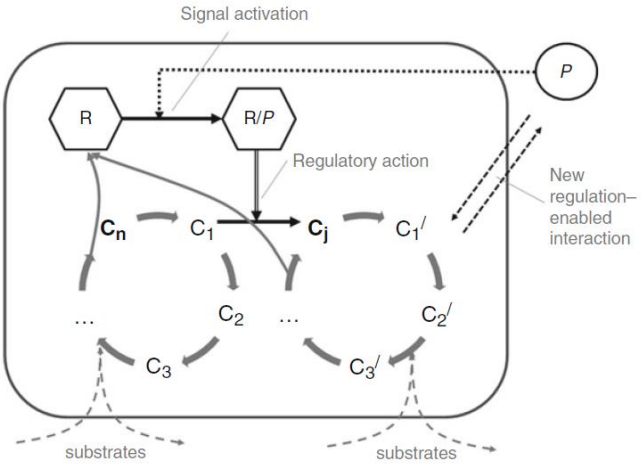

Figure 4.8 – Regulatory systems – R represent the regulatory constraint on the constitutive regime – C_1, C_2 … C_n during normal operation. In presence of perturbation (P), the regulatory constraint (R/P) brings about a new constitutive regime – C_1, C_2', …. C_j. [from Moreno and Mossio (2015), p.35].

4.4.2: Regulation, Robustness, and Complex Behaviours

Basic regulatory subsystems are shown by cellular mechanisms. The regulation on operons is a standard example. Operons (a collection of genes) is regulated by mechanisms that inhibit or activate them according to the requirement of the cell. For example, in *E. coli* energy is usually derived by the metabolism of glucose. But when glucose is not available, the bacteria switch to breaking down lactose into glucose and galactose to derive energy.[185] The process requires the transcription of a region of the genome called lac operon consisting of three encoding regions (genes) –

a. *lacZ* – which transcribes the enzyme LacZ (β-galactosidase) which cleavages the lactose molecule into glucose and galactose,

[185] Griffiths, Wessler, Carroll, and Doebley (2015), p.400-12.

b. *lacY* – which encodes LacY (Beta-galactoside permease) which is a membrane protein that embeds into the cell membrane and facilitates the intake of lactose, and

c. *lacA* – which transcribes LacA (β-galactoside transacetylase)

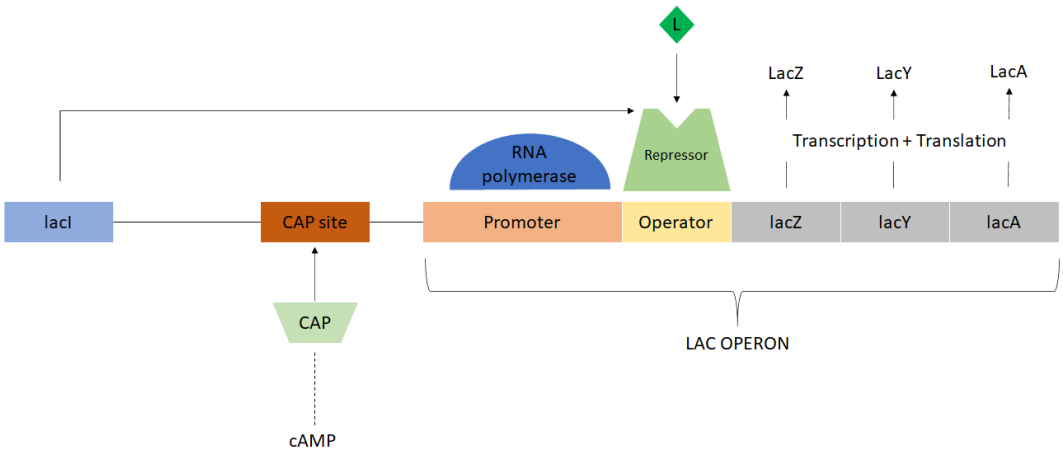

Figure 4.9 – Schematic diagram of lac operon.

The lac operon is generally transcribed at a very low rate. This is because another region of the genome – lac inducible (*lacI*) produces the lac repressor that binds to the operator region of the lac operon, thus preventing the RNA polymerase which binds to the promoter region, from transcribing the *lacZ*, *lacY*, and *lacA* genes, which helps convert lactose to glucose. But when lactose is present in the system, a lactose metabolite – allolactose binds to the repressor. The allolactose brings about a configurational change in the repressor, which in turn unbind from the operator. Thus, in the presence of lactose, the RNA polymerase is not constrained by the repressor and can transcribe lacZYA, thus facilitating the catalysis of lactose to glucose. Taken in isolation, the lac operon functions to catalyze lactose into glucose, thus providing glucose for the cell to derive ATP from it. But this is only a part of the picture. The lac operon transcription mechanism works as a regulatory mechanism for metabolism in bacteria.

To understand this, we must understand how lac operon is itself regulated (see *figure 4.9*) in the bacterial cell. In normal circumstances, both lactose and glucose are present in the environment. Therefore, the lactose molecule will bind with the oppressor and the oppressor will unbind from the operator, thus activating the transcription of the lac operon by the RNA polymerase at a higher rate. This will increase the glucose concentration further, even though glucose is already present. This is a wasteful process as the cell does not need the lactose to be broken into glucose to derive energy. Instead, the energy of the cell is wastefully being utilised to break the lactose when it is not needed. Therefore, the bacterial cell has another mechanism to regulate the lac operon transcription. This regulatory mechanism works by increasing the rate of transcription by RNA polymerase in very specific conditions that are only brought about when the glucose concentration in the cell is low. This is facilitated by cAMP – which is activated when glucose is in low concentration. As a result, it activates a catabolite gene activator protein (CAP). The CAP protein binds with the CAP site and increases the rate of transcription of the lacZYA genes. Therefore, helping in the catabolism of lactose into glucose only when the glucose is in low concentration.

The function of the lac operon regulation is thus not just about the catalysis of lactose when it is present in the cell. It is instead a regulatory mechanism for glucose concentration in the cell. When the glucose concentration in the cell is low, the cells catalyzes lactose to maintain glucose metabolism. In this way, the lac operon regulation works as a regulatory constraint on the metabolic organisation of the cell (which is maintained by constitutive constraints). The cAMP-CAP regulates the shift of the cell from a purely metabolic glucose-based regime to a regime that supplements the metabolic regime with the catabolic regime. This helps maintain the closure of the

metabolic breakdown of glucose when perturbations in the environment lead to instability of the metabolic cycle. The cAMP-CAP mechanism thus acts as *second-order constraints* on glucose metabolism and maintain the system by indirectly regulating the glucose concentration through the regulation of the lac operon.

The lac operon regulation model shows some features of regulation. Firstly, it does not give rise to new mechanisms but only regulates the switching of mechanisms that are already present in the cell. To do this, the control on the operon is achieved through both an inhibition/activation mechanism which is sensitive to allolactose and a positive feedforward mechanism that is sensitive to the glucose metabolism process. The negative control works to inhibit the mechanism when it is not needed and activating it when it is needed. The positive control helps regulate the lac genes to be translated at a higher rate depending on the state of the underlying mechanism through a signalling system involving cAMP as the signal. These processes realise a new causal regime in which the constitutive regime is changed in various ways. Firstly, the cell now shifts from a purely glucose based metabolic regime to a lactose based catabolic regime that maintains the energy requirement for other processes in the cell. Secondly, the cell actively imports lactose through the action of Beta-galactoside permease on cell membrane which allows the transportation of lactose inside the cell. Through both these processes, the regulatory subsystem is able to bring about a change in the dynamics of the cell and enable it to maintain its stability through changing its constitutive regime.

Another aspect of regulation is that regulation opens the opportunity for the constitutive organisation to realise *new adaptive behaviours* that can be exploited to realise existence not only through wider variations but also exploitation of these wider variations in the external environment for a more complex interaction with the

environment. In the case of *E. coli*, the presence of regulation of lac operon shows how the bacterium cell can exist both in glucose and lactose rich environment, thus providing an adaptive response to change in concentrations of glucose by shifting to another source for energy. Regulation thus, forms an important aspect of biological organisations, such that the minimal organisation that achieves closure, is maintained through perturbations.

In addition, regulatory systems can bring about robustness by shifting to different constitutive regimes [Bich (2018)]. Robustness is a ubiquitous property of biological organisations. It is generally defined as "a property that allows a system to maintain its functions despite external and internal perturbations"[186]. While the definition provides a general framework for understanding robustness, it does not capture the diversity of the mechanism by which robustness can be achieved. Secondly, defining robustness thusly, gives a false impression that robustness is synonymous with stability. While stability is an important feature of a robust system, such that effects of perturbations can be minimised through dynamic response from the system (dynamic stability/homeostasis) it is not the only way biological systems can show robustness. As Kitano (2004) 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. In

[186] Kitano (2004), p.826.

other words, robustness allows changes in the structure and components of the system owing to to perturbations, but specific functions are maintained"[187]

Dynamic stability and homeostatic properties of a system are essential for the system's functioning and are maintained by the organisation. What homeostasis and dynamic stability represent is an affinity for the system to stay in a stable state space (attractor state) such that any disruption is balanced out by change in dynamics of the system till the system returns to its stable state. This includes strategies like redundancy and degeneration. In such cases, the dynamic stability of the system may be disrupted such that it gets displaced from its stable basin of attractor such that it is incapable of returning back to it through its constitutive causal regime.

To counter these kinds of effects, the system can employ various strategies. Firstly, the system can show degeneracy such that various molecules or even pathways can show similar functionality such that they can perform the same task while being produced by and involving different processes. Secondly, they can show redundancy such that various copies of the same molecules (like enzymes) such that the failure of one does not disrupt the processes. Lastly, they can show diversity such that they do not depend on singular sources of materials from the environment.

For example, in the lac operon example, the glucose metabolism and lactose catabolism are degenerate as they both are used by the bacteria to derive energy. The degeneracy of these two paths is utilised by the bacteria to derive energy when in different conditions, thus maintaining itself even when one of them fails, thus showing diversity.

[187] Kitano (2004), p.827.

Robust realisation of the system can also lead to adaptive behaviour. For example, the robustness of the bacteria to derive energy from both glucose and lactose distinct strategies realised by the presence of lac genes shows how the bacteria can adapt to both glucose rich and lactose rich environments.

In what follows, I explore the relationship between robustness, regulation, and adaptive behaviour. My aim is to show that in many cases regulation plays an important role in achieving robustness and realising adaptive behaviour. To do this, I use the concept of design explanations as introduced by Wouters (2007) to define functional dependencies in physiology.

4.5: Design Explanations

Wouters (2007) introduces the idea of design explanation as a type of comparative explanation in physiology. Design explanations explain why a certain trait, rather than another, is present in an organism. Wouters (2007) argues that design explanations offer an explanation to the above question by showing comparatively how the trait contributes to the overall sustenance of the system and provides it an advantage as compared to another trait performing same or identical function. Design explanations are present looking affair, i.e. the consideration of how the trait evolved is not the part of the explanation. The aim of the explanation is thus to show why an organism has a certain trait rather than an conceivable alternative. This is done by showing that the organism would be worse-off in the absence of the trait. The explanation thus offered, explains by appealing to two aspects –

1. How the functional trait fits with the overall organisation, including other traits. Thus, showing a functional dependency.
2. How this trait enables the organisation to survive in the environment and increase its life chances.

For example, Wouters' considers the question – why organisms on land breathe through a lung based respiratory system, while organisms in aquatic environment breathe through gills/skin? The answer, according to Wouters is that of the oxygen concentration and diffusion rate. In water, the concentration of oxygen is much less than that of oxygen in the air. Furthermore, the diffusion rate is much lower, and the viscosity of water is much higher. Due to these considerations, the aquatic organism has to ventilate as much water as possible. The problem is solved by having gills that have an immensely larger surface in contact with water as compared to lungs with air and has much thinner membrane to facilitate diffusion. In comparison, lungs offer advantages to land dwellers by providing a counteraction to gravity – a problem much less pronounced in water dwellers due to the buoyancy that water provides. The breathing mechanism of lungs is much more powerful than gills and therefore is able to push the air up and out as a counter to the gravitational force. Thus, if we imagine a land dweller with gills, we can understand why it will not be able to draw oxygen like a fish can in water. The comparative system that we thus create is that of fiction, but it gives us some understanding about the real system we want to study. By doing so, we can therefore figure out some functional dependencies that constrain how a certain function can be performed given the internal organisation, external conditions, and some advantage the way of doing the function can provide under such constraints. In this way, Wouters' design explanations offer two kinds of functional explanations – that of the function of

the trait as biological role and how it enables the functioning of the organisation as a whole, and the function of the trait as the advantage it provides to the life chances.

But as Wouters (2007) notes, design explanations explain more than just functional attributes. Design explanations point out what advantage performing a function in a specific way is related to internal and external conditions – thus showing functional dependencies. These function dependencies are dependencies realised by the internal organisation, external circumstances, and the laws of nature that constrain the process. In case of respiration, Fick's law of diffusion and the concentration of oxygen in the water/air, determine why fish need to have a larger surface area with thinner membrane, while the comparatively larger effect of gravity (along with Fick's law) explains why land dwellers need lung like structure. Therefore, the presence of lungs/gills is explained by the functional capabilities they provide in a particular scenario.

The explanatory import from design principles is thus explaining how the realisation of a trait in a particular way is better than the conceivable (real or hypothetical) alternatives given the constraints of the system, those imposed by the environment, and the laws of nature. Applying this to the question of why biological systems employ regulatory subsystems to achieve robustness and adaptive behaviour leads to some interesting observations when the internal organisation of closure of constraints and the laws of thermodynamics and biochemistry are considered.

Let us consider the glycolytic pathway again. The presence of control explains a feature of the system – that of how the system controls how much glucose is used to derive energy. The importance of the control system in this case is inherently linked to the comparative advantages it provides to an uncontrolled system. In the latter, the lack of control will lead to glycolytic pathway constantly being active. This will lead to two

consequences – firstly, the glycolytic pathway is itself an energy intensive process. It requires ATP to phosphorylate many of its components and needs catalysts that are produced by other systems that need energy for their processes. Secondly, the storage of the ATP in cells is not easy as it is highly reactive and undergoes dephosphorylation, thus it gets readily depleted by the processes. Therefore, in such a system wasteful production and usage of ATP will hamper with proper functioning of the system itself. In contrast, the controlled glycolytic cycle presents the solution to both these problems. It prevents wasteful glycolysis, and instead facilitates the storage of glucose in the less reactive glycogen form which can be stored in the cells. It further explains a functional trait of the organism of storing excess source of energy within itself. This is an adaptive feature of how the organism interacts with the environment. This functional trait thus increases the life chances of the organism.

Similarly, the lac operon regulation helps the bacterium to shift between two different constitutive regimes to obtain energy. But one could ask the question – why does the bacterium not derive energy from both glucose and lactose at the same time? In theory, we can conceive of a bacterium which possesses both the lactose catalytic capability and glucose metabolic capability at the same time. But this would lead to the bacterium performing lactose catalysis when a less work intensive source of energy is present and sufficiently exploitable for its energy requirements. The regulatory mechanism instead, functions to reduce the energy wastage by the bacterium during glucose availability by inhibiting the lactose catabolism, thus maintaining a constitutive regime that requires lesser work to be maintained. Furthermore, by comparing it to systems where both lactose and glucose are metabolised at the same time, we can understand the limitations for the bacterium to do the same. Humans for example, can

metabolise both lactose and glucose at the same time. But, to be able to do that, they need – a) a storage capacity for glucose that is much higher than the bacterium cell can possess, and b) a regulatory system that can affect this storage over and above the simple control system of glycolytic pathway (see example of glycemia regulation below). This highlights the function of regulation as a trade-off between the organisation's total capacity and the energy required to maintain them.

Therefore, design explanations coupled with the idea of control and regulation based on the theoretical concept of closure of constraints explains why certain functions are performed in certain ways in biological systems. By comparing the system to a hypothetical or real system, we can ask questions about why a certain function is performed in a certain way and not the other. The answer depends on the internal organisation, external conditions, and the applicability of some general laws. In the case of control and regulation, the application of design explanations shows that given the requirement of maintenance of the minimal closure of constraints and to channel the thermodynamic flow in an optimal way within the limited capabilities afforded by the internal organisation, the strategies of control and regulation are better suited than other strategies (like having an uncontrolled processes, or having multiple systems work at same time).

As a final example, and to make this clearer, I take the example blood glycemia regulation in humans to show how all these concepts fit together.

Example – Glycemia Regulation

Consider the glycemia (blood glucose concentration) regulation in humans [Bich (2018)] – The glucose levels in the blood is contained within a range by the actions of various processes in the body. There is glucose uptake by cells for their energy requirements. The liver, which works both as a storage for glucose in the form of glycogen (Glycogenesis) and converts the stored glycogen and fats to glucose (Glycogenolysis and Gluconeogenesis) when glucose need in the system increases, works as a main constraint on these processes.

Other constraints include the cardio-vascular system that helps in glucose transportation to the cells in the body, gut cells which are responsible for glucose absorption, and cells which are responsible for glucose metabolism. All these constraints and processes together form a constitutive regime such that the homeostatic steady state glycemia can be maintained.

During fasting (when no new glucose is introduced into the body through food uptake), and when no high energy processes are taking place in the body, the glycemia levels can be maintained in a steady state by an intricate balance between the processes of glycogenesis and glycogenolysis and gluconeogenesis. In such a case, the activity of insulin and glucagon (which work as regulatory constraints on the process) on the process is minimal as insulin and glucagon (hormones produced by intestines) are present in basal values.

But when a sudden uptake of glucose (feeding) or sudden utilisation of glucose (energy extensive activities) by the system take place, this homeostasis maintained at basal conditions can be disrupted. As a result, the regulatory constraints of insulin and

glucagon (respectively) need to re-establish the homeostasis. Insulin and glucagon are produced in pancreatic alpha and beta cells, respectively, and stored in vesicles. Both the insulin and glucagon storing vesicles are sensitive to signals that are results of high and low glycemia, respectively. When glycemia decreases beyond the homeostatic range (hypoglycaemia), glucagon in released by the glucagon storing vesicles. Glucagon works to increase the rate of glycogenolysis and gluconeogenesis in the liver. Thus, it increases the release of glucose from the liver and restores the glycemia to its homeostatic level.

On the other hand, when the glycemia increases (hyperglycaemia), the insulin storing vesicles release the insulin in the blood stream. Insulin works by performing three functions –

a. it facilitates the uptake of glucose by muscle tissues,
b. it increases the rate of glycogenesis in the liver, and
c. it restricts the release of glucagon.

Note here, that the insulin and glucagon do not directly act upon the glucose, *but on the constraints like liver and muscle cells to drive processes towards a certain outcome*. Therefore, they act as second-order constraints on first-order constitutive regime and their constraints. Insulin in addition, acts upon pancreas by inhibiting the release of glucagon. Insulin, in addition, also works as a third order regulatory constraint upon the organisation by acting on the second order constraint of glucagon regulation.

The glycemia regulation thus, shows how maintaining homeostasis requires multiple levels of control on organisation –

1. At the first level, the basic constitutive regime is capable of homeostasis at a level that can provide stability to variations during the basal functioning of the system. But it is not sufficient for large perturbations that the system might encounter due to both internal and external disruptions.
2. For this, a second order regulatory organisation is required that works to re-establish homeostasis after a large perturbation.
3. Furthermore, this regulatory organisation might be under higher order regulation itself (insulin regulated glucagon).

Thus, the organisation shows hierarchical ordering such that regulatory mechanisms work to maintain the lower order processes [Bich (2018)]

But in addition, the regulatory processes also help the system realise complex interactions with the environment. In case of glycemia regulation, the perturbations that cause disruptions in the homeostatic behaviour of the basal conditions are processes like feeding and energy intensive work like exercise. In such cases, the constitutive regime is incapable of maintaining homeostasis. Therefore, the regulatory subsystem helps the system get back to its homeostatic state. But this also means that a system that is regulated in such a way can support behaviours like consuming quantities of food that cannot be supported if the constitutive regime alone was responsible for maintaining homeostasis. Therefore, the regulatory system assists the response in such a way that new behaviours can be produced. This adaptive new behaviour is assisted by the ways in which the regulatory system can affect the underlying constitutive regime. Therefore, the new behaviour that describes a further adaptive response (example being able to consume huge amounts of food in one go), is grounded in the capacities of the

underlying thermodynamics of the system and its maintenance through a complex web of regulation.

Applying design explanations to the case of glycemia regulation, we can imagine a system which does not have the insulin regulation, and therefore the maintenance of this homeostasis can be easily disrupted. Therefore, the system will not be able to perform some behaviours that an insulin enabled system can. One of this, as discussed above, is consuming large quantities of food without the system being displaced out of homeostasis to a dangerous level. In comparison, a similar constitutive regime without the regulatory intervention of insulin can be driven out of homeostasis by the same behaviour. This is exactly the case in type 2 diabetes patients where the general precautionary prevention for the disruption to this homeostasis is to consume smaller portions of food and at more regular intervals, such that the constitutive regime can maintain homeostasis without the assistance of insulin. The presence of regulatory subsystem that act upon the constraints in the constitutive regime thus provides a way to explain the adaptive behaviour derived from the presence of the regulatory subsystem. Therefore, the presence of the insulin/glucagon regulatory system offers a design explanation about why the regulatory arrangement is important for both maintenance of the glycemia and for adaptive behaviour that the system shows.

4.6: Conclusion

In this chapter, I have argued that closure of constraint idea, clubbed with presence of control explains the autonomous nature of biological organisation. This autonomy is exercised as self-determination through self-maintenance and is embedded in the idea

of constraints achieving closure. I further argued that constraints should be understood as weakly emergent configurations that derive their causal powers from restriction, reduction, and elimination of DOFs of the components of the configuration. This provides us with a way to understand constraints as active agents in bringing about the said constitutive regime. Furthermore, the organisational complexity in biological organisations is realised by exerting control on these constraints, thus bringing about maintenance of the constitutive regime during both internal and external perturbations. The analysis of the idea of control and regulation also provides us with an explanation of why certain features of biological systems – like robustness, adaptive response, and homeostasis are realised in a certain way through the idea of design explanations grounded in thermodynamic possibilities of the system.

The chapter thus contributes to the debate about explanations in biology in two ways. On an ontological level, it provides a foundation to think about biological systems as minimally autonomous. It, thus, is a reflection on the question about how biological systems are different from non-biological system. This was done by analysing the ideas of closure of constraints by philosophers like Mossio, Montévil, Bich, et al., and understanding the nature of constraints. Significantly, I add to the debate by providing a coherent idea of emergence that can explain the nature of constraints such that they can have causal powers that are irreducible to their physical realisation, and thus, provide a construal of constraints as configurations that explains why these configurations have causal effect over their surroundings and thus able to enable the system to maintain itself. By doing so, I have, in my opinion, substantially added to the view presented within the closure of constraints framework. In addition, the idea of emergence presented to make the closure of constraints theory more intelligible also has

consequences for objects posited by biological theories in general. The idea of reduction, restriction, and elimination of DOFs [Wilson (2010)] along with the SSCP [Wilson (2011)] helps construe biological objects as irreducible to their realisers with respect to the causal powers they exert. Thus, establishing a non-reductive physicalism fit for biology where the actual realisation of configuration posited as biological entities have real causal powers not available at to their realisers.

Secondly, I have conceptualised a model for explanations in biology based on the idea of closure of constraints and its implications – that of control, regulation, and robustness – and have shown how the ubiquity of these features answer design questions – why a certain function is performed in a certain way and not in any other way. I have, thus, combined the second-order regulatory constraints idea within the closure of constraints theory with Wouters' Design explanations [Wouters (2013)] to construct a way of modelling systems that explains the prevalence of certain ways of performing a function in biological systems as opposed to other possible ways. The crux of the argument hinges on why regulation is important in biological systems and thus limits the possible ways of performing a function due to the constraints exerted on the processes, and thus give us general organisational qualities that can answer design questions that we can pose. These design explanations are thus based on the self-maintaining nature of biological systems and how they enable the system to increase its life chances. I believe that this fills a gap in the literature, especially with regard to design explanations. While design explanations are posed as ways of performing certain functions in a certain way to increase life chances, the ubiquity of regulation as a design in biological systems is not explicitly answered within the design explanations framework. I believe that by combining the two – design explanations and closure of

constraints – not only we can answer the design questions, but also of the ubiquity of regulation by showing how regulated systems perform better at maintaining the system in a steady state far from equilibrium dynamics.

To further elucidate on this, I take the example of bacterial chemotaxis in the next chapter. I hope to achieve two things through this example. Firstly, I show that no one model can answer all the possible questions we can pose about the chemotactic system. And secondly, I show how the response of the chemotactic system is not possible by a single module/mechanism but requires systems working in tandem with each other.

CHAPTER FIVE
Explaining Bacterial Chemotaxis

Up till now, I have explored how various models in biology explain a system's function. In chapter 2, I argued that mechanistic models explain by elucidating how entities and activities bring about the phenomenon to be explained. In chapter 3, I argued that the mathematical models of systems biology explain by constructing minimal models of the system to explain certain formal dependencies. In chapter 4, I argued that the system theoretic concept of minimal organisation explains by showing how certain features – like closure – are important to explain the constraints on being alive, and how the thermodynamic grounding of how control is exerted explains why certain ways of performing a function are more desirable than others.

In this chapter, I take the example of Bacterial chemotaxis to show how various modelling exercises must be considered to explain the behaviour of the Robust Perfect Adaptation that *E. coli* realises.

5.1: Bacterial Chemotaxis

One of the most studied examples with regards to biological robustness in the recent years has been bacterial chemotaxis. Bacteria, like *E. coli*, move with the help of their external flagella. The movement of the flagella provide motility to the *E. coli* by executing two kinds of movements – 'run' and 'tumble'. During 'run', the rotation of the flagellum 'motor' is counterclockwise (CCW). The CCW movement aligns all the

flagella in a bundle and allows the bacteria to move in a specific direction. During 'tumble', the rotation of flagellum motor is clockwise (CW). The CW movement breaks up the bundle and the flagella rotate in an asynchronized manner, thus rotating the bacteria in a stationary place. After each tumble, the direction of bacteria is randomly changed, and the run is executed in this direction.

When the bacterium cell is in a uniform environment (such that the concentrations of nutrients or toxins is uniform through the surrounding), the bacteria executes a random walk, where it shifts between run and tumble. But when the concentration shows a gradient, the bacterium executes a directionality through a biased random walk with its movement towards the higher concentration of the nutrients (and/or lower concentrations of toxins). In the biased random walk, the runs are executed for a longer time with fewer tumbles in between (see *figure 5.1*).

What is remarkable about the bacterial chemotaxis behaviour is that it is perfectly adaptive. What this means is that after a period of decreased tumbling frequency during the biased random walk, the frequency of the tumble (f) increases till it reaches its initial frequency (f_0) that it had before the detection of the chemical gradient.

The biomolecular network of the perfectly adaptive chemotaxis is well understood (see *figure 5.2*). The bacterial receptor that detects the environmental attractants is called the Methyl Accepting Chemotaxis Protein (MCP). The MCP has two proteins – CheA and CheW. CheA is a sensor kinase that auto-phosphorylates when MCP is activated by an attractant. CheW assists with the interaction between CheA and MCP. CheA phosphorylates a response regulator protein CheY to CheYp. The CheYp diffuses with the flagella motor and brings about clockwise rotation, thus facilitating tumbling.

CheZ dephosphorylates CheYp, converting it back to CheY that can be phosphorylated by CheA again.

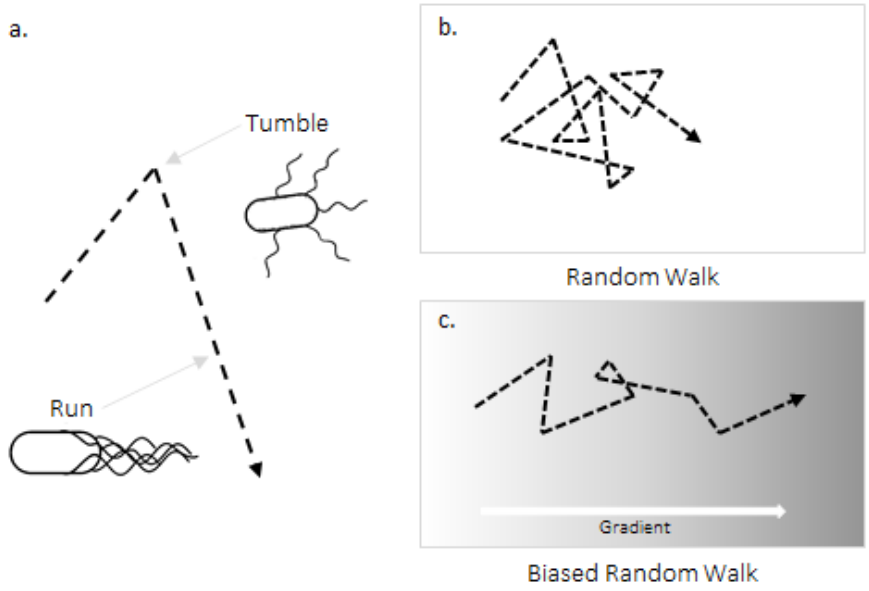

Figure 5.1 – Bacterial Chemotaxis. a. Organisation of flagella during Tumble and Run. b. Random walk during uniform concentration. c. Biased Random walk during a gradient of attractants (in this case nutrients).

In addition, CheB and CheR act as regulators in this process. CheR methylates MCP while a phosphorylated CheB (which obtains its phosphorus group from CheA) demythelates MCP. When the attractant is not bound to the MCP, CheA auto-phosphorylates itself using ATP. It then donates its phosphate group to CheY or CheB. At the same time, CheR slowly methylates the MCP while CheBp demethylates the MCP at a rate faster than the methylation by CheR. Therefore, the steady state random walk is maintained by the methylation-demethylation dynamics of CheR and CheB.

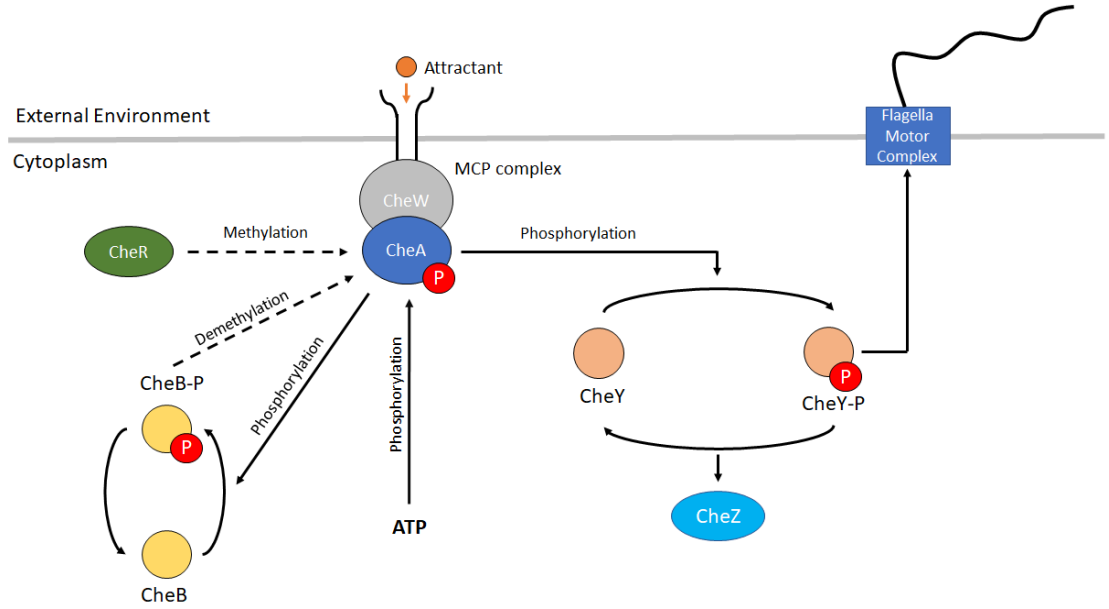

Figure 5.2 – A schema of chemotaxis showing the entities and activities that form the chemotactic mechanism. See text for details.

When an attractant is attached to the MCP, the activity of the CheW-CheA complex decreases. As a result, CheA's rate of phosphorylation decreases. This leads to less phosphorylation of CheY and CheB. As a result, the CheYp production falls and the flagella maintains its counter-clockwise rotation leading to 'run'. At the same time, the CheR methylates the MCP, but because the CheB phosphorylation rate is low, the demethylation rate of MCP is lower than the methylation rate. But a higher methylation of MCP leads to more activity of CheW-CheA complex, thus increasing the production of CheYp, thus leading to clockwise rotation and reinstating of 'tumble'.

As noted above, the chemotaxis in *E. coli* is perfectly adaptive as due to the presence of attractants, the frequency of tumble decreases, and the bacteria can perform longer runs. But after some time, the frequency of tumble reaches its original steady-state frequency (which is the same as without the attractant attached levels). But it has been found that the mechanism also shows robustness along some parameter variation. The

perfect adaptive response of the bacteria to environmental attractants shows robustness with respect to CheR concentration. That is, the bacteria tumbling frequency returns to its steady state value irrespective of the CheR concentration over a wide range of values. This has been called as robust perfect adaptation (RPA) of bacterial chemotaxis in the literature [Stelling, et al (2004), Alon (2007a), Barkai and Leibler (1997), Yi, et al (2001)].

The modelling of RPA has been a point of contention in the mechanist – mathematical/computational model debate. There have been attempts to explain the RPA through various models – some more successful than the others. Alon (2007a) considers two such models to show how RPA in chemotaxis is achieved – first is a *fine-tuned model* based on the mechanistic details that have been derived from experimental data, and the second is a *mathematical model that identifies the network motif* that is realised by the chemotactic system. Let us consider them in turn.

5.1.1: Fine-Tuned Model of Bacterial Chemotaxis

The first model considered by Alon (2007a), a fine-tuned model by Knox, et al. (1986), tries to explain the mechanisms by showing the effect of CheR and CheB concentrations on methylated and unmethylated receptor complex (MCP-CheW-CheA) – represented as X_m and X respectively. The CheR methylates X at saturation rate, while CheB demethylates X_m with Michaelis-Menten kinetics. The methylated receptor (X_m) phosphorylates CheY which leads to tumble.

When an attractor binds, the activity of X_m decreases and hence less CheY is phosphorylated leading to less tumbling and longer runs, but the demethylation activity

of CheB also decreases, until concentration X_m increases due to activity of CheR. Thus, returning the tumbling frequency to pre-attractant levels. Hence achieving perfect adaptation due to fine tuning of decrease in CheB activity and reduction in activity per methylated receptor (X_m) after attractant binding (see *figure 5.3*).

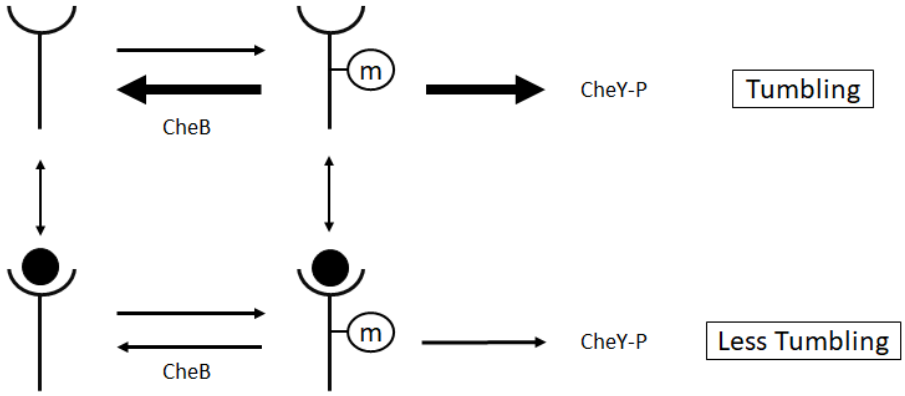

Figure 5.3 – Knox, et al (1986) fine-tuned model showing the relation between tumbling frequencies in case of no attractant binding and attractant binding. From Alon (2007a), p.143.

But, as Alon (2007a) shows, the adaptation is not robust to CheR concentrations. The relationship of perfect adaptation can be noted as –

$$A_0 = A_2 \tag{1}$$

Where A_0 is the activity of receptors in steady state before attractant binding and A_2 is the activity of the receptors in steady state with the attractant bound after an initial dip. Furthermore, A_0 can be defined as

$$A_0 = a_0 X_m \tag{2}$$

Where a_0 is the activity per methylated receptor and X_m is number of methylated receptors.

Similarly, A_2 is equal to –

$$A_2 = a_1 X'_m \tag{3}$$

Where a_1 is the activity per methylated receptor after attractor binding such that $a_1 << a_0$ (due to a drop in the activity after attractor binding) and X_m'is the number of methylated receptors at the new steady state A_2.

Furthermore, solving the equation for X_m and X_m', as a function of CheR and CheB activity and concentrations, we get –

$$X_m = \frac{KV_R R}{V_B B - V_R R} \tag{4}$$

$$X'_m = \frac{KV_R R}{V'_B B - V_R R} \tag{5}$$

Where V_R is the rate of methylation by CheR, R is CheR concentration, V_B is rate of demethylation by CheB, and B is CheB concentration. While V_B' is the change in demethylation rate by CheB after attractor binding.

Substituting (2), (3), (4), and (5) in (1); we get the condition for perfect adaptation –

$$a_0 \frac{KV_R R}{V_B B - V_R R} = a_1 \frac{KV_R R}{V'_B B - V_R R} \tag{6}$$

The system shows adaptive behaviour. Alon (2007a) shows this by taking a parameter set. Let us assume, that $a_0 = 10$, since $a_1 << a_0$, let us assume $a_1 = 1$. Let us also set K=1, $V_R R = 1$, and $V_B B = 2$. In this case A_0 can be solved with equation (2), such that –

$$A_0 = \frac{a_0 KV_R R}{V_B B - V_R R} = \frac{10}{2-1} = 10$$

And since $A_0 = A_2$, solving the equation (6), we get the value of $V'_B B$ as 1.1. Therefore, A2 becomes –

$$A_2 = a_1 \frac{KV_R R}{V'_B B - V_R R} = \frac{1}{1.1-1} = 10$$

Therefore, we can see that $A_0 = A_2$. The system returns to its original steady state after a decrease in activity (see *figure 5.4a*). But is the system robust against CheR concentration? Alon (2007a) considers this by setting the value of V_RR to 0.8, thus reducing it by 20%. Keeping all the other parameters the same, we get the values of A_0 and A_2 as –

$$A_0 = \frac{a_0 K V_R R}{V_B B - V_R R} = \frac{8}{2 - 0.8} = 6.66$$

$$A_2 = a_1 \frac{K V_R R}{V_B' B - V_R R} = \frac{0.8}{1.1 - 0.8} = 2.33$$

Therefore, the system shows variation when the value of V_RR changes and fails to show perfect adaptation. It does not return to the original frequency after the initial change in frequency (see *figure 5.4b*).

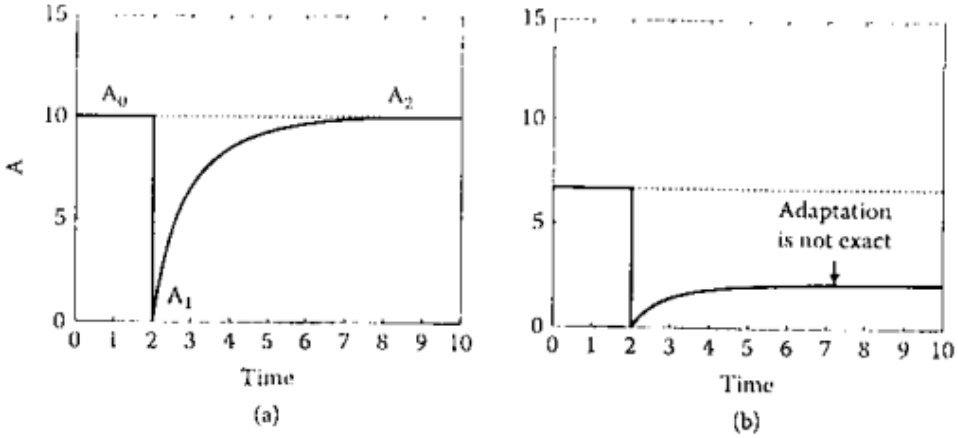

Figure 5.4 – Response of the fine-tuned model to a. when CheR activity is kept constant – perfect adaptation is observed. b. when CheR activity is reduced by 20%, perfect adaptation is not observed. Form Alon (2007a), p.144.

What this shows is that although a fine-tuned model, showing dependencies between the frequency of tumble and the activity of CheB achieves an adaptive response, it fails

to be adaptive when CheR concentration is changed. Thus, the system is not robust with respect to CheR concentration, contradictory to what has been observed in experimental data.

Let us now consider the mathematical model of bacterial chemotaxis as first presented by Barkai and Leibler (1997) and see if it fairs any better at providing an explanation for the RPA behaviour.

5.1.2: Barkai-Leibler Model of Bacterial Chemotaxis

The second model considered by Alon (2007a), a mathematical-computational model by Barkai and Leibler (1997), models the system a bit differently. Although some details are kept the same – like CheR acts at saturation rate and CheB with Michaelis-Menten kinetics, some additional assumptions are added. Instead of assuming the system to occupy two steady states – that of with and without attractant attached to the receptor complex, Barkai and Leibler (1997) assume the system has three attractor states – the unmethylated receptor (X), a methylated but inactive receptor (X_m) and a methylated and active receptor (X_m*). Only X_m* phosphorylates CheY. The CheR methylates X and converts it to X_m. X_m then transit rapidly between active (X_m*) and inactive (X_m) state. The phosphorylation of CheY only happens by active methylated receptors (X_m*). CheB only acts on active methylated receptors (X_m*) and demethylates them to X (see figure 5.5). The model consists of several ODEs describing the situation[188].

[188] For detail see Barkai and Leibler (1997) and Alon (2007a).

In the case of no attractor binding with X_m, the X_m* reaches a constant concentration as the demethylation exactly balances the methylation influx.

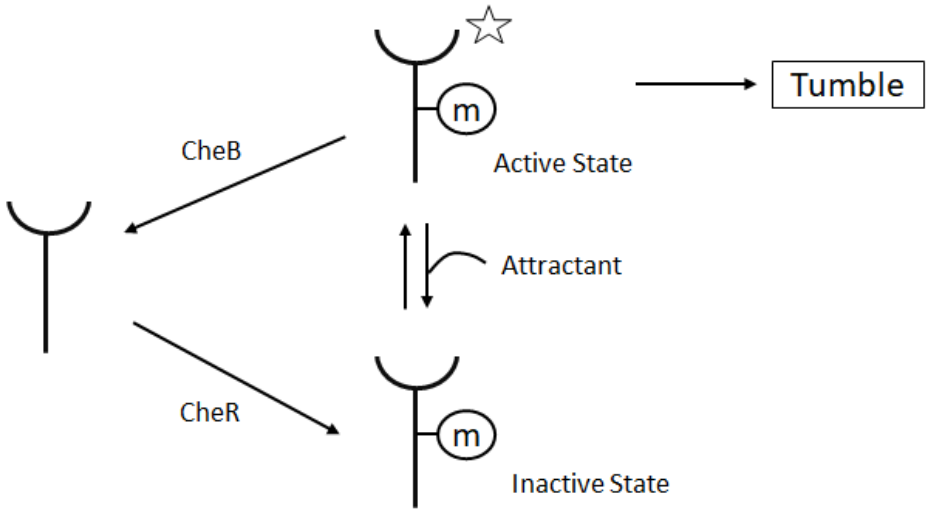

Figure 5.5 – Berkai-Leibler model for chemotaxis. CheR constantly methylates the receptors. CheB only demethylates active receptors. From Alon (2007a), p.146.

Therefore, the system achieves a constant frequency of tumbling. When the attractant binds to the receptor X_m, it decreases the probability of active state X_m*. As a result, the phosphorylation of CheY decreases leading to decrease in tumble frequency. Since CheB can only act on X_m*, the demethylation rate decreases. CheR, on the other hand, methylates X to X_m constantly. Therefore, the number of X_m gradually increases. As the total number of X_m increases, the active X_m* also gradually increases, which increases the demethylation rate by CheB. Eventually the demethylation rate and methylation rate are balanced such that a steady state is achieved again. Therefore, the system achieves an adaptive response by adjusting the concentration of active X_m* against the methylation and demethylation rates. Furthermore, the response time of the model is

faster than the rate of unbinding of the attractant, therefore, the system is capable of adapting to the bound attractant's presence and return to its original switching frequency regardless. This model also shows robustness to parameter change. Alon, et al (1999) tested the model against genetically altered *E. coli* strains that produced higher CheR concentrations, and subsequently for altered concentrations of other chemotactic proteins and found that the adaptive response showed robustness for the changed concentrations.

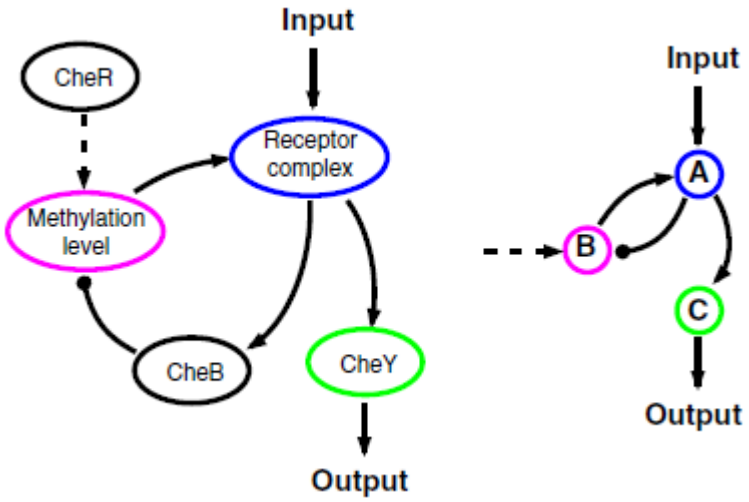

Figure 5.6 – NFBLB motif as realised by Barkai-Leibler model of methylated receptors. From Ma, et al (2009), p.771.

Ma, et al (2009), further show that the Barkai-Leibler model realises a *n*egative *f*eed*b*ack *l*oop with a *b*uffering node (NFBLB) topology where a three-node motif is realised in which one node acts as the buffer. As shown in *figure 5.6,* this motif can realise RPA through buffering the methylation level when the kinetics of its activation (by CheR) are at saturation level. The model, thus, by optimising the RPA response shows that there is a certain necessary condition for RPA that the mechanistic model does not capture – that of CheB demethylating only active receptors.

Yi et al. (2000) analyse the Barkai-Leibler model and argue that the model realises an Integral Feedback Control (IFC) system for the concentration of the methylated receptors. An IFC is realised when the system maintains some variable with reference to a desired value. A thermostat is an example of a IFC such that it maintains the temperature within a desired range set by the user. Analogously, as Yi, et al. (2000) argue, the bacterial system is able to maintain the concentration of methylated receptors and thus facilitate the return to steady state tumbling frequency after an initial shift away from it. To do this, Yi, et al (2000) analysed the ODEs of the Barkai and Leibler (1997) model and show that a solution to the ODEs offers a characteristic IFC equation when certain assumptions, that are also included in the Barkai-Leibler model, are met. These assumptions include –

a. CheB only demethylates active receptors,
b. the activity of unmethylated receptors is negligible, and
c. the rate of action of CheB and CheR and the concentration of CheR is independent of ligand presence.

Relaxing any of these assumptions shows a deviation from the adaptive response.

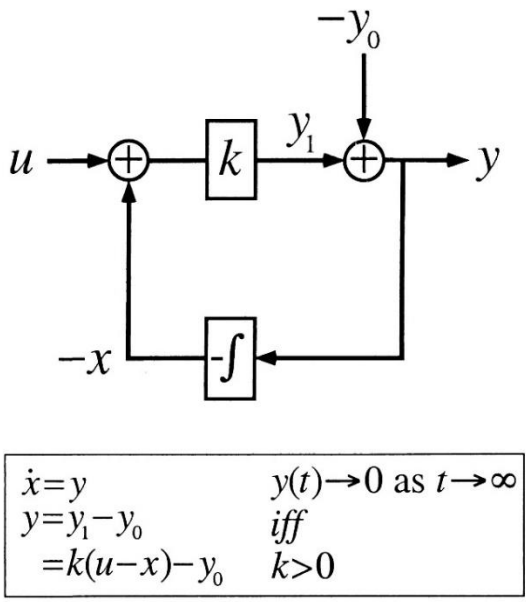

Figure 5.7 – Diagram for IFC principle. The IFC ensures that a response to input u, provides a reliable output y through calibration of the actual output y_1 and desired output y_0 through a feedback of the integral of the error x. From Yi, et al (2000), p.4650.

Yi, et al (2000) argue that the IFC is a necessary and sufficient condition for the RPA seen in bacterial chemotactic network. The formulation of IFC makes the study of the complexity of the system relatively easy by providing an easier way to track dependencies of variables. As discussed in Chapter 3, Yi, et al's argument for chemotaxis realising an IFC would qualify as a minimal model, where the formal constraints elucidated by the mathematical derivation of the system shows how the system belongs to a particular universality class. The IFC is realised by both chemotactic network and thermostats and thus is realised by a myriad of systems with differing composition, and causal relationships.

What makes such systems similar are the dependency relationships that they show. As noted in Chapter 3, mathematical models do not necessarily track mechanistic details while explaining the phenomenon, instead they may idealize and abstract away from

mechanistic details and use optimisation assumptions to model the system. Issad and Malaterre (2015) claim that such models instead explain through the mathematical derivation. While a causal interpretation of the model in terms of mechanism might be possible, the model's explanatory force is not drawn from these causal details but from the derivation itself.

As I have shown, the BL model is able to model the chemotactic response in *E.coli* by introducing some optimising assumptions during mathematical derivation. As a result, it is able to identify an organisational feature of the chemotactic network in form of the NFBLB architecture. In recent literature, there has been a debate about what the Barkai-Leibler model, and the IFC realisation actually explain. This debate is set within the larger debate about the relationship between mechanistic models and mathematical models. In what follows, I analyse what the BL model and its IFC realisation actually explains. And how it can lead to better understanding of the mechanistic model, but to assert that this is all it does is wrong. On the other hand, I show that the claims about the explanatory powers of the mathematical model can be overstated and that this should be avoided.

5.2: Mathematical Models and Mechanistic Models

Recently, Matthiessen (2017) has argued that the IFC must be construed as a mechanistic schema. Matthiessen (2017) states that the models like IFC and Barkai-Leibler model (BL henceforth), represent a form of explanation that presents the mechanism in abstract through mechanistic schemas that can be filled in. Schemas thus,

stand in a representational relationship with the actual mechanism that they explain[189]. Matthiessen (2017) argues that the abstract representation of the mechanism by the mathematical model helps us understand the organisational structure of the mechanism while leaving the finer details to be filled in as per the epistemic and pragmatic interests [Bechtel and Levy (2013), see Chapter 3]. The assertion by Matthiessen is part of a larger debate about the relationship between mathematical modelling and mechanistic explanations, and therefore has bearing on how mathematical models in biology must be construed [Levy and Bechtel (2013), Brigandt (2015), Piccinini and Craver (2011)].

In what follows, I analyse Matthiessen's arguments and show that while one role of mechanistic models is to supplement mechanistic models, it would be wrong to assume that mathematical models only form mechanistic schemas. Mathematical models have an explanatory role independent of elucidating the functioning of the mechanism, which I have explored in Chapter 3, in terms of minimal models.

Matthiessen (2017) makes three arguments –

1. The mathematical models explain in virtue of constructing 'how possible' models. This is evident in how BL model specifies the three conditions that the model must follow to function in the desired way. Thus, according to Matthiessen, mathematical models must be construed as a constraint-based reasoning that works as a heuristic device for mechanistic discovery. In case of chemotaxis in *E. coli*, the BL model and the IFC model, as per this view, elucidate what the underlying mechanisms can possibly bring about the

[189] This is consistent with the (OC2') conception of ontic constraints on mechanistic models as argued in Chapter 2.

phenomenon. The three conditions, thus, place a constraint on what possible mechanisms are necessary for the realisation of chemotactic response.

2. A specific criticism of Braillard's (2010) appeal to design principles as explanatory. As noted in previous section, design principles [Wouters (2007)] explain by drawing comparisons between different systems to show why a certain function must be performed in a certain way. Braillard (2010) takes the basic idea of design explanation and claims that the IFC model is a design principle which is necessary and sufficient for chemotactic response. As Matthiessen points out, this claim cannot be substantiated, especially for all chemotactic networks including other bacterial species like *B. subtilis*.
3. The mathematical model does not answer all the questions regarding the bacterial chemotactic network – including why different bacteria (like *B. subtilis*) achieve robustness through different methods.

In what follows, I argue that Matthiessen is wrong on (1) but correct on (2) and (3). For (1), I argue that while Matthiessen is correct in pointing out that mathematical models can function as a heuristic for constructing 'how possible' mechanistic models, this does not exhaust the utility of mathematical models which explain through construction of minimal models that help generalize some features of the system which can have explanatory role on their own.

5.2.1: Explanation: Beyond Mechanisms

While there is merit to Matthiessen's argument (1), I believe it stems from a limited understanding of mathematical models. Mathematical models in systems biology can

have two roles in biological explanation – one is derived from the heuristic role it plays in mechanism discovery and explanations, and the other is of constructing minimal models. The former can be utilised as a way of assessing, detailing, and modelling mechanisms. The latter is about providing general explanatory principles through the construction of minimal models.

In the case of bacterial chemotaxis, the models play both roles – the first being explicitly provided by the BL model, and the second being the derivation of IFC from the BL model. The BL model, proceeds by modelling the chemotaxis system by constructing differential equations and then solving them to show a solution to the observed system behaviour. The solution is reached by employing optimising assumptions that can have biological meaning. Some of these have been found to hold true – such as CheR working at saturation rate. While others still need to be established – such as CheB only acting on active receptors (more on this later). In any case, the optimising assumptions act as conditions on how the underlying mechanisms must be organised. For example, the three conditions in the BL model point to certain conditions that the underlying mechanisms must together satisfy to realise the BL model.

The other role though, is to provide an abstract, idealized, and optimised minimal model of the system such that we can explore some minimal dependency relationships between different parameters and variables. By constructing this minimal model, the representation becomes far removed from the details of the specific model and moves towards a general representation that can be applied to more than one target system. For example, the olfactory receptors in mammals show an RPA like structure that results in the receptors becoming insensitive to the olfactory ligands and returning to their original

state. The modelling of the system has been found to be very similar to the BL model[190], with assumptions that correspond to BL model, albeit realised by different proteins taking part in different mechanisms. Thus, the steps in the construction of the model provide a generality to the model that can be exploited to explain more than one model with similar architecture. Therefore, the model does not explain specific features of the causal mechanism but only abstract generic organisational features, such that how the mechanistic details realise them can be excluded from the explanation.

This is consistent with the trade-off between generality and abstraction as highlighted by Cartwright (1983). Cartwright (1983) argues that empirical adequacy and explanatory power trade-off each other. In her view, generality of a model/theory has less explanatory power in particular cases. Therefore, any model that accounts for more and more factors specific to the case, the more explanatory it will be for the particular case, but less general. The models of systems biology as discussed, provide us with generality but are less explanatory for specific cases. As discusses in Chapter 3, this is because of the methods of model building and how abstraction, idealization, and optimising assumptions are used, which lead to construction of minimal models that are stripped away from all the irrelevant features – including many causal features of the mechanistic models.

This is also consistent with what Levy and Bechtel (2013) have in mind when they argue that abstraction explains by moving away from the causal details and construing the mechanism in abstract terms such that the organisational structure is apparent instead of the fine-grained causal details of how the components bring about the

[190]As Gorur-Shandilya, et al. (2017) note while modelling the Olfactory receptor "a minimal two-state model of the olfactory receptor-olfactory co-receptor (Or-Orco) complex with an adaptation architecture similar to that of the bacterial chemotaxis system".

phenomenon. In their example of the persistence detector, the FFL realises a particular organisation that explains how the system deals with noise in the signal. We could, in principle, construe the persistence detectors as a mechanistic schema that stands in for the more detailed description of the system in which it is realised. But as Levy and Bechtel (2013) themselves note, the study of abstract organisations is a study of patterns of connectivity which, while can be utilised for mechanistic purposes, is focused at studying the dynamics of the system. The architecture of the persistence detector represents a class of systems that shows certain dynamics due to its pattern of organisation. As they note (while discussing the arabinose system and persistence detector) –

> "Note an important feature of this analysis: it does not depend on the details of the arabinose system. Persistence detection is a consequence of the abstract organization of the system. Following Alon, we have given specific details about the parts and operations in the arabinose circuit. But these do not serve to elucidate why the mechanism operates as a persistence detector. They only show that it is present in the arabinose context and suggest why it is useful there. Altering the details of the components as long as they meet the minimum conditions for fulfilling the role in the organizational schema does not change the behavior, whereas altering the organization changing what is connected to what does"[191]

The case of IFC being realised by chemotaxis network is based on even more abstraction. Yi, et al (2000) note that if the optimising assumptions of BL model are met, then the methylated receptors achieve a steady state feedback control that can be characterised as an IFC. The IFC is much more relational than the BL model that it is

[191] Levy and Bechtel (2013), p.253.

based on. What it does is to track the change of variable over time under the constraints placed by the BL model rather than providing any representational characteristic. An analogy might help. A transistor in engineering circuits is a three-port device that can work as a switch or an amplifier. A transistor can have many constructions – like Bipolar Junction Transistor (BJT) and Field Effect Transistor (FET) – which can be further classified into Junction Field Effect Transistor (J-FET) and Metal Oxide Semiconductor Field Effect Transistor (MOSFET). Furthermore, the construction of the transistor can be classified according to its sourcing – either as Positive-Negative-Positive (PNP) or Negative-Positive-Negative (NPN). The construction of the transistor as switch or amplifier in each combination is different depending on its sourcing and construction. Furthermore, the effect of the biasing (how the voltage is applied) is different in each construction. The PNP and NPN react differently to the differing Collector-Emitter/Drain-Source biasing. The characteristic response though, is a relational property between the biasing applied and the type of transistor used. The transistor thus, can have various constructions and biasing, with different token realisations, but the relational properties between these various parameters define its function as a switch/amplifier. As a result, when the right relational properties are realised, irrespective of the specific characteristic of the construction of the transistor, it will show a characteristic response. Therefore, the characteristic response is of a transistor type rather than of a token transistor. The response, thus, defines characteristic dependencies between variable and the parameters. As can be seen from the graph (*figure 5.8*) – the variables can be different. The BJT has variable response of the Channel current (I_C) to different values of Base current (I_B) represented by different slopes in the graph. The FET has variable Drain current response (I_D), which is analogous to I_C, to the Gate-Source Voltage (V_{GS}). In the first case, the characteristic

response is a relationship between current and current at different terminals, in the later, it is between the current at a terminal and the voltage at two terminals. The implementation of the switch/amplifier by each type of transistor is different – it requires specific configurations dependent on the construction which must be complemented by the bias. But nonetheless, the realisation of the switch/amplifier by all these configurations share some important characteristics that can be ascribed to the transistor function type. The IFC similarly, defines the relational properties between the parameters and the variables such that it can be instantiated by many different configurations. Yi, et al (2001) show that when the assumptions of BL model are considered, the configuration shows an IFC type response. This kind of generality has been a cornerstone for understanding general organisational principles throughout various networks in biomolecular systems[192].

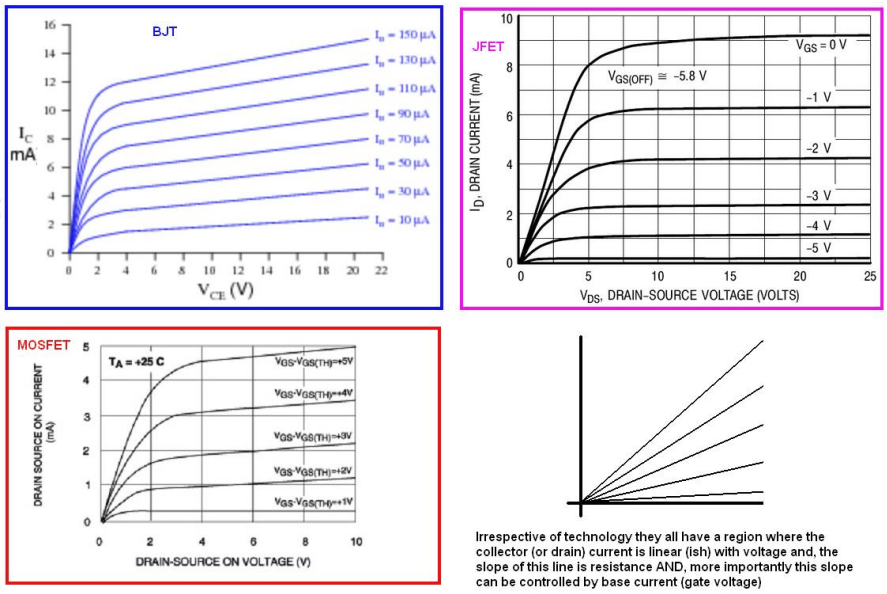

Figure 5.8 – Characteristic response of different constructions of Transistors.

[192] von Bertallanfy (1968) for example shows how exponential law can be applied to both growth and decay in very disparate systems – like radioactive decay, antibiotic action on bacteria, population growth, and autocatalytic rate. Also, see Green and Wolkenhauer (2013).

While studying specific systems, like the bacterial chemotaxis or olfactory receptors, we can certainly use these models as schemas for detailed mechanistic models, but it does not exhaust the utility of mathematical models. Firstly, there are other ways in which mathematical models can help construct mechanistic picture of the system. As Baetu (2015b) notes, a mathematical model can be used to determine the completeness of the mechanistic explanation by analysing parameter sufficiency through simulation. If, say, the dynamic simulation of the mathematical model or the landscape does not match what behaviour of the mechanism is observed to be, then it means that the mechanistic model is not complete. Baetu (2015b) writes –

> "If model simulations match experimental data, it can be argued that a more complex model including additional parameters is not needed. In as much as all the parameters have a clear physical interpretation… a close match between simulation and experimental measurements of the phenomenon is taken as evidence supporting the claim that a more complex mechanism including additional components and the physiological context of other mechanisms is not needed to produce the phenomenon."[193]

The heuristic roles that can be played by mathematical models of systems biology in mechanistic discovery and analysis are thus, diverse – it can be used as constraint based reasoning for identifying the constraints that are acting on the mechanism or it can be used to study if the mechanistic model is adequate to explain the phenomenon in question. To construe them as mere mechanistic schemas would therefore be a mistake.

[193] Baetu (2015b), p.782

At the same time the generality offered by mathematical models show that they are far richer in their explanatory role.

But what are the explanatory virtues of mathematical models? I have argued in chapter 3 that mathematical models can be used to construct minimal models that show how certain causal features are irrelevant to the phenomenon displayed. I have argued that in deriving equilibrium conditions, the initial concentration is irrelevant. The equilibrium concentration depends only on the rate of production and the degeneration constant. This does not mean that the initial condition is irrelevant in charting the trajectory of the system towards equilibrium, but that equilibrium is obtained depends on the structural features of the system – that production and degeneration take place. A similar case can be made for the feedback control. Araujo and Liotta (2018) study the topological requirements for RPA and concluded that there are two basic topologies that realise the RPA. One is with an opposer module (which can consist of more than one opposer nodes), and another with a balancer module. They showed that these topologies show RPA properties at many nodes (see *figure 5.9*). While a detailed discussion of their work is out of the scope of my thesis, they sum up their findings as follows –

> "These findings represent a significant advance in our understanding of the basic structures underlying the complex and evolving networks occurring in nature. In many biological contexts… the underlying signaling networks are so complex and high-dimensional, so prone to change over time, and so extraordinarily variable from one realization to another (even from one cell to… cell), that the networks themselves are virtually impossible to define concretely at any useful level of detail. Although most investigators view this variability as a source of intractable complexity… our work reveals that these networks may now be considered from

the point of their unexpected simplicity—that is, as decompositions into well-defined basis modules."[194]

This shows that the property of having RPA in can be studied as a property of the system without analysing its underlying details. Therefore, the realisation of RPA in a particular system can be functionally explained by studying the topology. In this way, the topological analysis shows a generality that of how a universality class of systems show some properties.

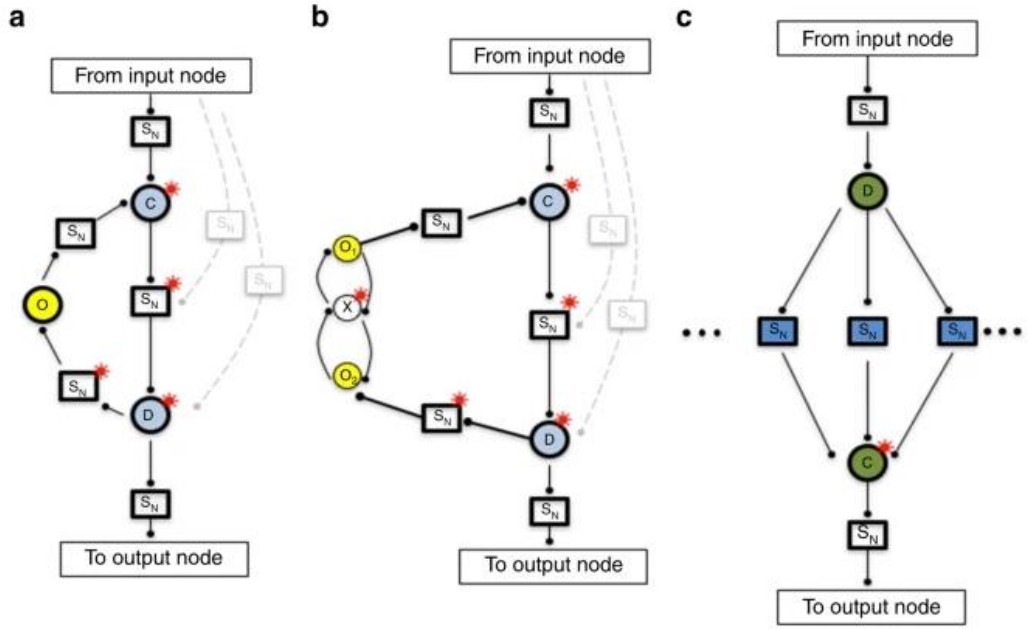

Figure 5.9 – Two modules that realise RPA – Opposer module (a) and (b), and Balancer module (c). All nodes that exhibit RPA are marked with red asterisk. S_N represents any motif or sub-network that can be embedded in the system. The nodes show RPA irrespective of the architecture of S_N. All interactions may be activating or inhibitory. The difference between (a) one opposer module and (b) two opposer module is that in (b) both opposer module work in concert. The Balancer module (c) consists of at least two parallel route segments. Form Araujo and Liotta (2018), p7.

[194] Araujo and Liotta (2018), p.8-9.

Furthermore, it has been observed that mathematical models can be analysed with respect to changes in parameter combinations. Daniels, et al (2008) show that multi-parameter networks show parameter combinations that are 'stiff' such that any change in these parameters will bring about a change in the systems overall functioning. On the other hand, other parameters form what can be called 'sloppy' parameter spectra. The system shows a characteristic feature of being sensitive to changes in the stiff parameters while insensitivity to change in sloppy parameters. This means that a system's behaviour will change when intervened on certain parameters but not when intervened on the sloppy ones. Daniels, et al (2008) while discussing the sloppy parameters note–

> "The model is robust… because of the mathematical behavior of chemical reaction networks, which are naturally weakly dependent on all but a few combinations of reaction parameters."[195]

As the quote suggests, the study of model as a whole with parameter combinations can help understand how the underlying network topology may have parameters that show sloppiness. In such a case, an intervention on some parameter combinations will show no or much less effect on the functioning of the network as an intervention on the stiff parameters will.

What Daniels, et al (2008) imply, thus, is that the characteristic of robustness towards interventions on certain parameter combinations might be a result of the characteristic organisation that is encoded in the formal dependencies that the mathematical model captures. Therefore, mathematical models can explain certain features of robustness by

[195] Daniels et al. (2008), p.393

the virtue of their very realisation irrespective of the underlying mechanistic realities. This is solidified by Yi, et al (2000) argument that once we have the mathematical model of chemotaxis *in E. coli*, we can see how the relationships described by it realise the IFC, which will show certain functional capacities.

Models like the IFC, thus do not only explain the whole behaviour of one system, but explain how the robust response to the perturbations, such that the system is able to return back to its base variable value, as a result of this general organisation. The IFC is thus a functional realisation that has multiple realisability. The role of mathematical model here is more than constructing the mechanistic schema. They explain by the virtue of their derivation. Robust control of the chemotaxis is thus explicable by the displaying that the system realises an IFC type of organisation, and then explaining what realising such organisation entails.

Mathematical models can thus play various roles in biological modelling. They can work as mechanistic schemas by constructing 'how possible' models as Matthiessen (2017) argues but they can also provide general principles. The construal of what they entail depends on the epistemic interests. As Green and Wolkenhauer (2013) argue, following von Bertalanffy's early general systems theory, that we can construe the use of mathematical models in systems biology of different kinds of explanations.

Figure 5.10 shows the different kinds of mathematical models and their aims and corresponding abstractness. At the representational level, the model describes the behaviour of a specific system through formalising the interactions. On the Functioning level, the model describes the dynamic behaviour of the system. At this level, the correspondence between the model and the specific system is established through experimentation and model fitting. As Green and Wolkenhauer (2013) argue, network

motifs can be seen as a good candidate for this level as they can represent the interactions of the system through signs that can be interpreted as representation for more concrete specific cases. But, as they note, while they can be used as templates for more specific models for study of individual systems, their epistemic value cannot be reduced to merely acting as templates. This is because the model has a generality to it and any specific interpretation needs to be established through experimentation. The BL model fits neatly in the functional level, as it defines the dynamic behaviour of a specific system (chemotaxis) but is still generalizable enough to be used in other systems (olfactory receptors). The Organisation level, in contrast, helps establish general principles of which a specific case is an instantiation of. At this level, the generality becomes an epistemic virtue because it explains a type rather than a token instantiation. The IFC fits with the description at the organisation level as it is a general organisational principle of which the BL model is an instantiation, but the realisation of an IFC type can have multiple realisability and multiple token instantiations.

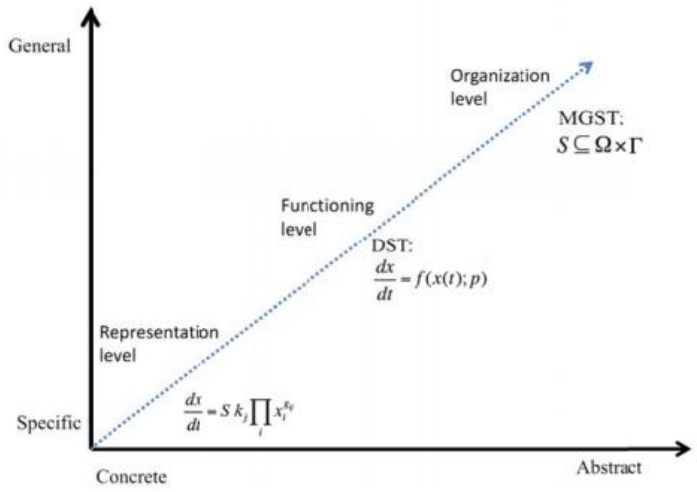

Figure 5.10 – Different roles of mathematical model. From Green and Wolkenhauer (2013)..

Therefore, while I agree with Matthiessen (2017) that the BL/IFC model of bacterial chemotaxis can provide the basis of forming a mechanistic schema and that the utilisation of this aspect of the mathematical model is a fruitful step towards mechanistic discovery, I disagree that the BL/IFC should be only considered as mechanistic schema. This shows an ignorance towards how mathematical models are constructed and how and what they explain.

5.2.2: IFC and Design Explanation

I have argued in the last section that Matthiessen (2017) is wrong in arguing that the BL or IFC model should be construed only as mechanistic schema. In this section, I analyse his second argument – that of IFC being construed as design explanation [a charge specifically at Braillard (2010)] and show why he is correct. To say that the chemotactic network in *E. coli* realises IFC is explanatory is correct. But at the same time, it is important to note that the stronger claim that IFC is necessary and sufficient for the explanation of RPA *tout court* is wrong.

Braillard (2010) argues that the IFC realisation in bacterial chemotaxis shows why IFC is necessary and sufficient for RPA. Braillard (2010) bases his argument on the claim by Yi, et al (2000) that IFC is necessary and sufficient for RPA. Furthermore, he envisions this as what has been called Design explanations, a term that he borrows from Wouters (2007). As discussed in chapter 4, Wouters' Design explanation is a comparative functional explanation which answers questions like – "why do fish respire through gills/skin, while land animals respire through lungs?" The answer, according to

Wouters is that of the oxygen concentration and diffusion rate that differ widely in water and in air (see Chapter 4 for details). By pointing out how these factors constrain the ways in which the function can be performed, Wouters shows that gills are not feasible as a breathing apparatus on land and a land dwelling gilled being will not be able to survive. In this way, Wouters' design explanations offer two kinds of functional explanations – that of the function of the trait as biological role and how it enables the functioning of the organisation as a whole, and the function of the trait as the advantage it provides to the life chances.

Braillard's (2010) argument in support of the claim that IFC is necessary and sufficient for RPA is thus derived from the application of design principle to bacterial chemotaxis of *E. coli*. First, the comparative designs that are taken by Braillard (2010) are the fine-tuned model and the BL model. He, therefore, shows how one design fits the explanation better than the other in case of RPA. Second, based on Yi, et al's claim that IFC is necessary and sufficient for RPA, Braillard hypothesizes that there is a functional dependency between RPA and IFC. He thus states that "[d]esign explanations… show why a given structure or design is necessary or highly preferable in order to perform function or to have important property like robustness"[196].

But RPA as a property can be realised through other topological organisations. Ma, et al (2009), for example, studied three-node motifs to see which motifs are capable of performing RPA. They find negative feedback loop with a buffering node (NFBLB) and incoherent feedforward loop with a proportioner node (IFFLP). In addition, Ma, et al (2009) also specify the conditions under which these topologies can provide perfect or near perfect adaptation. As Ma, et al (2009) note, the NFBLB is realised in the BL model

[196] Braillard (2010), p.55.

of bacterial chemotaxis. For, IFFLP on the other hand, they could not find any biological example. But it is nonetheless a possibility that IFFLP can be used to achieve RPA. That the IFFLP achieves IFC type behaviour needs to be established by mathematical derivation as Yi, et al (2000) did for BL model. Therefore, Braillard's claim that IFC is necessary for RPA is disputable.

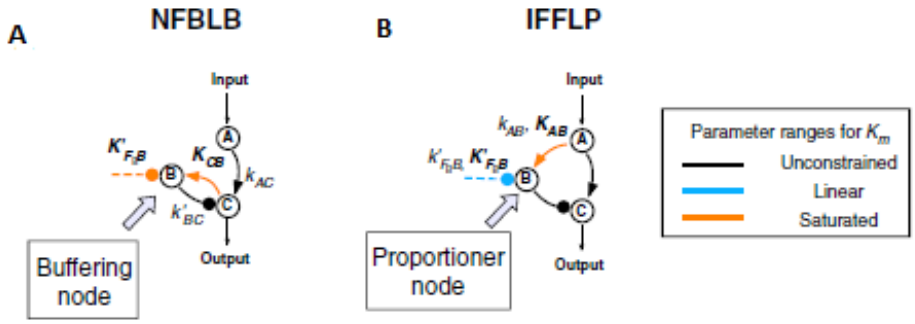

Figure 5.11 – Three node motifs that can realise RPA. NFBLB is consists of a feedback with a buffer node, while IFFLP is a feedforward with a proportioner node. The coloured arrows show the conditions on the kinetics for achieving RPA. Ma, et al (2009), p.765.

The realisation-realiser relation between IFC and its property of showing RPA is asymmetrical. While IFC type dependency relations, when realised, will show RPA like properties, it cannot be established that IFC is necessary for RPA, as other structures can show RPA. As an example, consider the transistor working as a switch. The dependency relations shown by transistor explains why it can work as a switch when certain parameter values are achieved. Nonetheless, it doesn't mean that there can be no other structure for switches. Neither does it mean that the transistor acts as a switch for all its possible parameter combination. Similarly, feedback loops, when they realise a particular parameter combination, such that they feed back an integral of the difference between desired and actual state realise RPA such that they provide robustness against

variation in some parameter (or parameter combination) values, can achieve an adaptive response towards a specific state. But it does not establish that this is the only way RPA can be achieved or that feedback loops always realise RPA. Therefore, the conclusion for necessity and sufficiency of one design cannot be established without extensive empirical evidence.

Green, Levy and Bechtel (2015) have utilised what they call 'a thin notion of design' to explain the relation between motifs and functions by pointing out to how the particular organisational features of the motif can explain their function in the system as a 'present-looking affair'. This thin notion of design is derived by studying the motif and how it can perform certain functions (like FFL as persistence detector – see Chapter 3). Their argument significantly differs from Braillard's argument as they base their notion of design on optimisation. Therefore, they show – given the constraints on the function and its realisation, a particular design is optimal. Thus –

a) they do not imply that a particular design is the only possible design for a function,
b) their idea of design is bound by the constraints placed by the system, and
c) the design is optimal in the sense that it might not be realised by the system in its optimised form but reflects that in ideal optimising conditions the design performs the function[197].

[197] In this respect, a better analogy for design explanation of the sort expressed in Green, Levy and Bechtel (2015) might be that of hub-and-spoke [Godfrey-Smith (2009a)]. The motif or the topology in question is an optimised model that forms the hub such that specific individual cases form the spoke where the optimised model is applicable. The application of the hub model to the spoke cases requires both the explanatory imports from the hub model and the ad-hoc information of the spoke case. This I think better represents the view of design principles as used in systems biology.

As I will show in the next section, the third point relates to bacterial chemotaxis and therefore the sufficiency of IFC for chemotaxis can be challenged.

5.2.3: Is RPA Sufficient to Explain Bacterial Chemotaxis?

In the previous section, I argued that mathematical models serve as more than mechanistic schemas. They show how the whole dynamically produces the phenomenon in question [Fagan (2015)] albeit in abstract and idealized form [Levy and Bechtel (2013)] by employing optimising assumption. I, therefore, argued that Matthiessen's position that mathematical models of systems biology work as mechanistic schemas alone is flawed. Nonetheless, the specifics of the system matter. As Matthiessen points out, the variability in the switching frequency in bacteria is huge, such that different bacterium, even of the same species, show widely varying switching frequencies. Secondly, it can be contested that the IFC module realisation at the receptor level is sufficient to explain the robustness of the whole chemotaxis network. It has been found that in *E. coli* the CheY-CheZ-CheA form a cascade of feedback loops that explains robustness of concentration of CheYp and how in the absence of changes in receptor binding activity, or small changes in receptor activity, is able to maintain the CheYp concentration. Almogy, et al (2001) show that the stability of CheYp concentration is not only dependent on the dynamic stability of the methylated receptor concentration, but requires another regulatory system downstream at the motor complex that involves feedback loops which include CheZ and free CheAs (*figure 5.13*). Similarly, in *B. subtilis* it has been found that the regulation of CheYp requires three different modules that show robustness to perturbations such that removal of any two of them severely affects the regulatory system. As a result, the robustness of the whole

chemotaxis network is much more than that of *E. coli* [Tindall et al. (2008), Rao, et al (2004)].

Almogy, et al (2001) study shows that in *E. coli* the CheYp concentration at motor complex is not determined by the concentration of active receptors alone. In the case of small changes in receptor activity (when ligand concentration is low), the negative feedback loops formed by CheZ and CheZ-CheA complex (which forms what they call the Methylation independent control) are sufficient to maintain a steady concentration of CheYp at the motor complex, thus bringing about random walks. In case of larger changes though, the Methylation Independent Control is insufficient for maintaining the steady concentration and needs the regulatory control from the Methylation Dependent Control module (the Receptor-CheB-CheR complex). The two-stage regulation of the CheYp is responsible for dynamic stability of the system. Furthermore, they note that the Methylation-Independent control can restore steady state frequency at a faster rate than the Methylation Dependent control. The relationship between Methylation Dependent Control module and Methylation Independent Control module is not exactly a combinational effect. As Almogy, et al (2001) note –

> "CheZ and CheAs-CheZ may serve in two parallel, negative feedback loops that down-regulate the level of CheYp... Our modelling study predicts that the methylation- independent mechanism by itself provides an almost exact adaptation for small ligand concentrations, an adaptation that becomes increasingly impaired for larger changes in receptor activity."[198]

[198] Almogy, et al (2001), p.3024.

What the extended study of chemotaxis beyond the Methylation Dependent module (whose functioning can be captured by the BL and IFC models) shows is that the robustness of the network is not localised to one robust module but is distributed throughout the network. It is not only important that the methylated receptor concentration is maintained in response to perturbations in ligand concentration, but also that the downstream component that connects the methylated receptor dynamics is translated to a dependable activity at the motor complex. Furthermore, the Methylation Independent module, due to being robust to some extent in its functioning, prevents smaller variations in the network from affecting the bacteria's tumbling frequency when minor changes in ligand concentrations take place. Studying this system in this way presents a different picture of robustness as compared to studying Methylation Dependent control alone. Both the systems assist each other in weeding out the effect of noise in the environment, and to restore dynamic equilibrium after larger perturbations. Therefore, the system's robustness is a distributed feature.

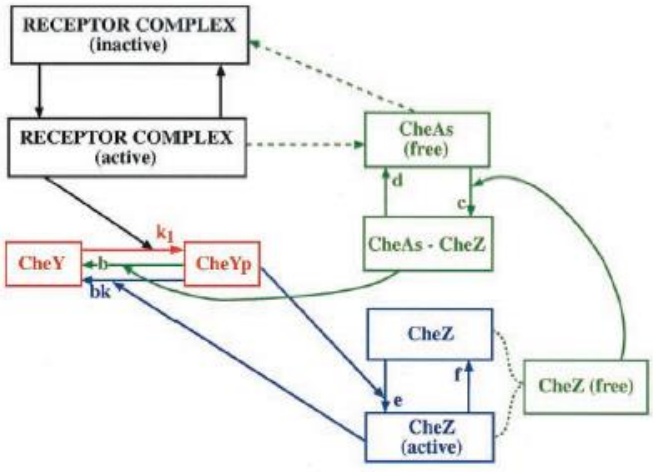

Figure 5.12 – The Methylation Independent module for control in E. coli. Almogy, et al (2001), p.3018.

In the terminology introduced in Chapter 4, the Methylation Independent module forms the first level of control, such that during minor perturbations, the bacterium is able to maintain itself in steady state equilibrium due to the action of the phosphorylating/dephosphorylating feedback loops involving CheZ and CheZ-CheA. This is achieved by a stoichiometric control over the availability of CheYp to the motor complex and control over free CheAs which are available to bind to the MCP complex. The intricate control thus leads to random walks before it returns to tumbling therefore maintaining a consistent tumbling frequency. The constraints on the system realised by limiting the availability of CheA to bind at the MCP cite by coupling it with CheYp concentration. But this minimal control is not sufficient to maintain the system during higher concentration of ligands. As the ligand concentration increases, the balance between active and inactive methylated receptors is disturbed – with attractants pushing more methylated receptors to inactive state, while repellents push more receptors to active state. In such cases, the bacterial system must rely on the Methylation Dependent module which varies the number of active MCP sites through methylation and demethylation. The IFC in this scheme of things, describes how the Methylation dependent regulatory system returns the system back to the original concentration of methylated receptors after an initial dip or surge due to high binding of attractant/repellent molecules, thus reinstating the Methylation Independent control.

In this regard, the application of the regulatory control framework explains why Methylation Dependent module must realise an IFC like property when faced with perturbations that are beyond the capacity of the Methylation Independent control to deal with. Since the aim of the regulatory system is to reinstate the original regime of

the *E. coli* system, the IFC like property ensures that the initial regime is reinstated even when the concentration of the attractant/repellent is high. This enables the bacteria to reinstate its normal functioning. In the absence of IFC like response, the bacteria will be stuck in the run mode for a long time. The IFC realisation, provides the system with the property of returning to its original tumbling frequency and continue with its normal functioning. But this is based on understanding the minimal control provided by Methylation independent module and why it needs to be regulated upon. Therefore, while we can say with certainty that *E. coli* needs an IFC type response to reinstate its original regime, we cannot say that for other kinds of RPA response in other systems (like *B. subtilis*) without such analysis.

In this respect, Matthiessen is correct. Firstly, he points out that different species of bacteria use different systems to attain robustness. *E. coli* uses two feedback loops in its Methylation Independent control while *B. subtilis* uses three different modules in combination [Rao et al (2008)]. Therefore, we cannot say for certainty that IFC type response is realised in *B. subtilis* without a similar kind of analysis. Secondly, the variation in switching frequency from one bacteria cell to the other shows that there is much more complex mechanisms and variability than the model portrays. These are important questions that must be answered.

While Yi, et al (2000) hold that IFC is necessary and sufficient for understanding RPA, this might not be true, especially when it comes to the robustness of the whole system. The IFC as per BL model is realised in the Methylation Dependent module, but it might not itself realise perfect RPA. This is because all the conditions of the optimising assumptions might not be met. As already stated, these conditions must be tested for their validity experimentally. While most of the conditions have been verified

experimentally (Alon 2007a), some conditions have not been. For example, that CheB demethylates only active receptors has no direct experimental proof, although some indirect experimental proof has been found – for instance in case of sudden increase in attractant, the demethylation rate has been seen to drop steeply [di Bernardo, et al. (2005), Mello and Tu (2003), see Cosentino and Bates (2011)]. But, as Cosentino and Bates (2011) argue, even if CheB demethylates inactive receptors such that the rate of demethylation of inactive receptors is much less than rate of demethylation of active receptors, the mathematical model shows that an imperfect RPA network will still be realised with the IFC realisation, such that the lesser the demethylation of inactive receptors, the closer the model will behave like an ideal IFC. Thus, mathematical models can explain some deviations from the ideal.

But, if this is the case, then how is RPA achieved? As argued above, the Methylation Dependent module which realises an IFC through the optimising assumptions of BL model is a regulatory system acting in congruence with Methylation Independent module which is capable of achieving homeostasis for CheYp through feedback loops between CheZ-CheY and CheZ-CheA when the concentration of ligands is low. Therefore, while the Methylation Dependent control might not be able to realise perfect RPA response, if it realises a response close enough, then the Methylation Independent module can fine tune the response to realise a perfect RPA.

The mathematical model of bacterial chemotaxis explains the functioning of an optimised RPA module. In reality, though, the optimising assumptions of such an idealized model might not be reliably met by the system. As I have argued before, while discussing Fagan's constitutive mechanistic explanations in section 5.2, the bearings of idealization and abstraction must be considered while considering how the model can

be reliably applied to the system. This further solidifies the idea that mathematical models in systems biology describe a general class of systems rather than a particular system. This happens because of the idealizing assumptions that are necessary for construction of such models, and the optimising assumptions which describe the functioning of the system at an optimal level. While a non-ideal response can be derived from the mathematical model, the question then shifts to how the other aligned systems compensate for it.

5.3: Conclusion

In this chapter, I have provided an example of how different models and perspectives are needed to successfully explain chemotaxis in *E. coli*. The mechanistic model helps identify the entities and activities of the network. But a fine-tuned quantitative model based on the mechanistic details alone is insufficient to account for the phenomenon displayed as a whole. In comparison, a dynamic model with the study of the topological requirements and the optimising assumptions yielded better results. A further abstract organisational account shows how the RPA response in *E. coli* is achieved through the implementation of IFC principle. Contra Matthiessen (2017), I argued that while the mathematical model can be understood as a mechanistic schema such that the optimising assumptions provide a constraint on the underlying mechanisms, this characterisation of mathematical model is incomplete. I argued that mathematical models provide a generality due to the idealization and abstraction techniques of the modelling exercise. This generality means that the model explains because of dependency relations that it captures, and thus is applicable to a wider variety of systems with differing mechanistic details. I have also cautioned, contra Braillard (2010), that

this does not mean that the realisation of RPA always requires implementation of IFC. That the IFC is implemented in a system showing RPA needs to be established and cannot be taken as a given. Therefore, IFC is not necessary for RPA. Thus, the idea of design explanation cannot be reliably applied in this case.

Finally, I argued that given the complexity of the system, we cannot state that IFC is sufficient for RPA. It might be the case that only a near perfect adaptive character is realised by the actual chemotaxis system. Therefore, I argued that it is not sufficient to study just the IFC to explain robustness of the chemotaxis. I argued that this is because the idealizing and optimising assumptions explain the system in an ideal manner which might not correspond to biological reality.

Instead, we must study how the whole system is integrated such that the robustness is distributed and can only be explained by studying and combining multiple mechanistic and mathematical models. In this regard, the explanatory import from Mossio, et al's view on biological autonomy and regulation provides a useful framework for understanding how control is exerted on the minimal constitutive regime such that different hierarchical levels of control are needed to explain the functioning of the system. The RPA module can be construed as a regulative control over a constitutive control exerted by the underlying constitutive regime such that the latter can explain adaptive behaviour to a particularly smaller perturbation, while the former is required to bring the system back to its normal functioning after a larger perturbation. In combination, this schema provides a response in which the integrity of constitutive regime is maintained in the face of large environmental perturbations, while providing a way to understand the adaptive response to such perturbations. What is important is

to understand that none of the systems might be capable of realising the perfect RPA response on their own.

The different explanations offered by different models thus shows that a plurality of explanatory frameworks and models is needed to explain the realisation of RPA by *E. coli* chemotactic system.

CONCLUSIONS

My aim in this thesis has been to understand how explanatory strategies in biology are employed to explain and model biological systems in different ways. In this regard, I paid attention to three different explanatory frameworks and clarified their explanatory powers.

Different models use different theoretical frameworks, techniques, and cater to different pragmatic and theoretical interests. Mechanistic models are characterised by their strategies of decomposition and localisation to enable an explanation based on components that bring about the phenomenon and their constitutive and causal relevance. Mathematical models in system biology explain by constructing optimised models that show a wider applicability and explain by showing why certain causal features are irrelevant when it comes to explaining the system's behaviour. Design explanations explain by constructing false hypothetical models that show by comparison to real cases, why certain dependencies are important for the realisation of a function.

An important feature of providing explanations in biology, though, is that models can have consequences for explanatory frameworks that they are not specifically built for. Mathematical models can be construed as mechanistic schemas for particular system application. Models play heuristic role for explanations other than the explanatory framework they have been constructed within, but it should not be assumed that this is the only role the model plays. Caution must also be paid to not overestimating the explanatory power of a model. This is particularly true for models that imply a kind

of generality (like mathematical models). I have shown that the generality does not necessarily show that the model, as it is, is applicable to a particular behaviour *tout court*, and the particulars of the system must be considered.

I have also argued in favour of a specific view of what separates the systems in biological domain from the non-living. While complexity is a feature that is found in both biological and non-biological systems, biological complex systems can achieve an autonomy that ensures their maintenance over a wide range of conditions. This is achieved by ensuring a closure of emergent constraints. This defines a minimal ontology of biological organisations. I have argued that this also has explanatory import for understanding why certain functions are performed in a certain way in biological systems and why regulation should be seen as a ubiquitous biological need. While I have not explored the whole theoretic import of this approach, a few can be stated here.

Firstly, the closure of constraints idea provides a general framework for understanding constraints as a primary construction in biological organisations. Constraints thus construed are real, emergent entities that need a wider biological and philosophical analysis especially because, unlike physical systems, these constraints are actively acted upon by the organisation through closure. This can have significance for further explanations and modelling of biological systems. One way to exploit the idea of constraints is to understand them within the framework of complexity. While a systematic analysis of constraints and their role in complex dynamic systems is not yet available, Hooker (2011) provides an introductory analysis with certain problem statements for the import of constraints in the description of complex dynamic system especially how different models can be combined.

Secondly, it must be analysed what constraints mean for other explanatory frameworks. Recently, Winning and Bechtel (2018) have argued that mechanistic explanations are resource poor for analysing control hierarchies in biological and neurological systems. Their solution is to argue for understanding and incorporation of time-dependent, variable constraints to ground the causal powers of mechanisms. This, they argue, will also require a capacity-based metaphysics of mechanisms [see Illari and Williamson (2011)]. While I will not analyse this idea here, it certainly has some merits to explain why mechanisms are the way they are (as they are constrained by actions of other mechanisms). Both these aspects are important to analyse and I hope to pursue them in other places.

A final aspect that becomes apparent from my analysis of models and explanations in biological systems is that of explanatory pluralism. Biology is a model-based science. Biologists routinely construct and analyse models at various levels and to answer a variety of questions. The modelling exercise in biology is diverse and can analyse a system in various ways. A plurality of models operate in different disciplines in biology and are bound by their theoretical framework and heuristics. There are thus, a plurality of models and plurality of representations of a given system.

This is also because accounting for the complexity of biological systems is difficult and needs various approaches are needed. The explanatory pluralism is thus a consequence of this complexity. As a result, explanatory pluralism is both grounded in pragmatic considerations of a model-based science and an ontological complexity that manifests in different ways at different levels. Still, a further analysis to differentiate the two is warranted.

Explanatory pluralism in biology can be identified with three distinct domains –

Firstly, our current scientific practices in biology have a *plurality of models*. Therefore, a pragmatic consideration of a model-based science that deals with a multitude of questions should be to take an explanatory pluralism. As I have argued, different models have different explanatory goals in mind. This does not mean that they cannot be explanatory for different fields. Instead the plurality of models is explanatorily relevant for other explanatory purposes. Models can act as tools for discovery and constraint-based reasoning for various explanatory exercises and therefore serve a multitude of roles including understanding, prediction, and control.

Secondly, the biological domain is that of a *plurality of questions* – Biologists often ask different questions about the system. A basic framework of this division is the distinction between 'how' and 'why' questions, but a better framework might be Tinbergen's four question framework. Biologists thus can ask causal questions about how a mechanism underlies the phenomenon/behaviour, they can ask why a certain behaviour is present – both through an analysis of its evolution and through analysis of the survival value it provides. They can also ask questions about ontogeny and the development of a behaviour. All these questions require different forms of explanations – answering the question about why a behaviour is more suitable with an answer to how it is realised is not explanatorily relevant to the question. As I have shown, the multitude of questions can help explain features of the system in different ways. In bacterial chemotaxis, for example, the question of why the chemotactic network achieves RPA cannot be answered simply by pointing to the composition but by showing how certain organisational features are required (along with their optimisation assumptions). Similarly, the disruption of immune system response by HIV is not only dependent on

the mechanism of its action but the topology of the T-cells (see chapter 3). The import of these explanations is closely linked to the questions asked.

Lastly, there are *different types (or modes) of explanations*. This is a more controversial form of pluralism and is deeply dependent on what explanation means. Do explanations in biology have to be inherently causal? Are good explanations mechanistic? If so, how do we accommodate the sense of pluralism in our analysis of what a good explanation is? Furthermore, are their ontic norms that define what a good explanation should look like? If so, what should they be, and how do we analyse models based on these norms? I do not have a clear answer to this question. But I ascribe to the idea that the adequacy of an explanatory model should be first and foremost determined by the epistemic criteria. This will include pragmatic aspects, as well as an analysis of heuristics and the limitations inherent in the modelling exercise. The adequacy of such models, therefore, must be judged according to the questions they answer as determined by the modeller and the wider scientific community.

But what about integration? In the first chapter, following Mitchell (2003), I introduced the idea of integrative pluralism. Integrative pluralism essentially states that the plurality of descriptions/models/explanations neither leads to a plurality of ontology nor means that these descriptions/models/explanations are in competition with each other. Instead, the presence of this plurality is pragmatically motivated, and therefore, we can, in principle, integrate these models to provide a fuller picture of the phenomenon. To this end, I had argued that the first step must be to show that no model/description/representation can capture all the features relevant to a full understanding of the phenomenon. This, I believe, I have elucidated successfully with my example of bacterial chemotaxis in the last chapter. The second step, then, is to show

that it is possible to integrate these models. In what follows, I show how this might be possible for the models considered within this thesis.

C.1 Towards Integrative Pluralism

Any pragmatic theory of models must account for the diversity of models. Since models serve different outcomes, and are built within different frameworks, different models can represent the given phenomenon in various different ways. Integrative Pluralism holds that this plurality seen in our representations/models/descriptions does not necessarily imply a plurality of ontologies. As Mitchell (2003) notes, it is in principle possible to integrate different explanatory models to give a fuller picture of the phenomenon to be explained.

While an answer about the more general answer about the possibility of integration in biology as a whole is beyond the scope of this thesis, I wish to explore a more specific issue of integration between the different models considered in this thesis.

To start with, let us examine the relationship between mechanistic and mathematical models. As stated in chapter 3, the information from the mathematical model cannot just be read off to provide a mechanistic model of the same phenomenon. Instead, mathematical models are constructed with their own assumptions, idealizations, and goals. Nonetheless, both these models can be complementary to each other in two ways. Firstly, mathematical models of mechanisms can explain the dynamic behaviour of the mechanism by situating the mechanism within a class of systems that show a certain kind of behaviour. By doing so, the mathematical models can explore the structural features of a proposed mechanism.

On the other hand, the same structural features associated with a class of systems can help discover mechanisms by limiting the possible mechanistic realisations. Green and Jones (2016) have called this constraint-based reasoning. Constraint-based reasoning proceeds by showing how different systems, despite their different make, up show certain structural-functional dependencies which can be captured by formal constraints on the system. These formal constraints can thus be used to understand the necessary conditions for mechanistic realisations by limiting the allowable space for the realisation. Thus, mathematical models can facilitate the discovery of mechanistic models. For example, the chemotaxis mechanism in *E. coli* must show an RPA type response, which the mathematical model shows is achieved by an IFC type architecture. Thus, we know that any proposed mechanism must be able to show an IFC type response to be a possible mechanism for the chemotactic network. Thus, mathematical models can be used both to explore the structural-functional dependencies of a mechanism and as a result, help in mechanism discovery.

The closure of constraints model as design explanations in biology, as explored in chapter 4, has different ramification for mechanisms. As Winning and Bechtel (2018) have pointed out, we cannot think of mechanisms as working in isolation. Instead, mechanisms must be seen within the larger framework of interacting with other mechanisms. As noted in chapter 4 and further in chapter 5, regulation plays a crucial role in biological systems. Within the closure of constraints view, regulatory constraints form second-order constraints acting on first-order constraints. This regulation provides the system with stability and robustness beyond the stability and robustness of individual modules. This feature of robustness can be explained by design explanations – as to why certain functions are performed a certain way. So, while mechanistic models

and mathematical models explain how certain entities and activities are instrumental in bringing about the phenomenon, and why certain dependencies are needed to realise it, respectively, the design explanations explore why the phenomenon occurs in a certain way in first place. This can be done by pointing out the constraints on the system that must be closed to causes. For example, while the mechanistic model of chemotaxis elucidates the role of entities such as CheA, CheB, CheW, etc and activities such as methylation, demethylation etc, and the mathematical model explains why the system must realise an IFC like architecture, the design explanations explain why the RPA type module is instrumental in maintaining the functionality of the whole bacterial system in the first place. This is done by elucidating two basic ideas. First, that if RPA is not achieved, the bacterium might end up in a run phase forever when encountering a gradient, thus exhausting its energy during the run phase and dying. Thus, the property to become insensitive to a particular level of the stimulant helps the bacterium cell adapt to changing conditions, thus providing robustness to the system. And secondly, by pointing out how the constraints of one module on another (methylation dependent module, and methylation independent module) realise a regulatory regime that helps maintain the stability of the system in different conditions. Thus, as Winning and Bechtel (2018) note, it is not sufficient to understand how isolated mechanisms work, but to understand how they constrain and maintain each other to bring about and sustain the system. Design explanations, clubbed with the idea of closure of constraints, can provide an explanatory framework to explore this idea further.

To sum up, integration of models that have been explored in this thesis, bring forth the idea of constraints in two ways – formal and material. Firstly, the constraint-based reasoning of mathematical models can help explain both the dependencies of the system

through formal constraints and help discover possible underlying mechanisms. Secondly, material constraints on systems can help explain why a system performs certain functions in specific ways and thus explain why the dependencies elucidated by formal constraints are present in the first place. These constraints are essentially biological because they form a causally closed regime. By doing so these constraints are able to maintain the stability of the system by exerting control on the processes and maintaining each other. It also explains why there is a ubiquity of regulation in biological systems, and why we must consider how these constraints act, are realised, and are maintained. Without this insight, we will miss two essential questions that we can ask about the realising mechanism. Firstly, why the mechanism is the way it is, and secondly, how do mechanisms work together to form a biological organisation that is stable and is able to maintain itself. While studying of mechanism in isolation does offer insight to how a phenomenon is brought about, unless they are considered as a part of a very specific kind of biological organisation, they fail to answer these 'why' questions. For the mathematical models of systems biology, the framework provides a motivation to understand generic principles elucidated by the formal constraints. What has been called 'Design Principles' by authors like Green and Jones (2016); Levy and Bechtel (2013); Green, Bechtel and Levy (2015) beg the question of why only certain designs captured by formal constraints are found in biological systems. While a decisive answer is difficult to provide here, there might be a relation between the very nature of biological organisation and regulation that dictate the formal constraints captured by these designs. Certainly, the need of maintaining the stability of a system by material constraints would formally constraint the possible designs. For example, retaking the example of chemotactic network, the sensitivity to the chemical receptor must have a variability so as to prevent the bacterium being stuck in the run phase. This points to the

need of regulation such that when sufficient time has passed in the run phase, the bacterium can reinstate its tumble phase despite the chemical signal remaining the same strength. This would further constraint the possibilities of how the underlying mechanism must act. These constraints, therefore, capture the necessary conditions for a viable mechanism and can be mathematically captured as structural-functional dependencies that are required. As it happens, in this case, the dependencies are captured by the design that we can call IFC-like. Therefore, the ubiquity of regulation, the need for stability, and the presence of robustness can dictate the formal constraints on the system. Thus, these models complement each other and enrich our understanding of biological systems when integrated. Therefore, imploring us to be dedicated to Integrative Pluralism.

BIBLIOGRAPHY

- Adami, C. (2002) – "What is complexity?", *BioEssays* 24 (12): 1085–1094.
- Alexander, S. (1920) – *Space, Time, and Deity: The Gifford Lectures at Glasgow 1916–1918*, 2 volumes, London: Macmillan.
- Alon, U., Surette, M. G., Barkai, N. and Leibler, S. (1999) - 'Robustness in Bacterial Chemotaxis', *Nature*, 397: 168–71
- Alon, U. (2007a) – *Introduction to systems biology*, Boca Raton: Chapman and Hall/CRC.
- Alon, U. (2007b) – "Network motifs: theory and experimental approaches*", Nat Rev Genet* 8, pp450–461
- Almogy, G.; Stone, L. and Ben-Tal, N. (2001) – "Multi-stage regulation, a key to reliable adaptive biochemical pathways", *Biophys. J.* 81: 3016–3028.
- Araujo, R.P., Liotta, L.A. (2018) – 'The topological requirements for robust perfect adaptation in networks of any size', *Nat Commun* 9, 1757 (2018).
- Artzy-Randrup, Y.; Fleishman, S.J.; Ben-Tal, N. & Stone, L. (2004) – "Comment on "Network motifs: simple building blocks of complex networks" and "Superfamilies of evolved and designed networks"", *Science* 305(5687), 1107.
- Badii, R. and Politi, A. (1997) – *Complexity: Hierarchical structures and scaling in physics*. Cambridge: Cambridge University Press.
- Baetu, T. M. (2015a) – "From Mechanisms to Mathematical Models and Back to Mechanisms: Quantitative Mechanistic Explanations," *in Braillard, P.A. and Malaterre, C. (eds.) (2015) – Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, Dordrecht: Springer: 345-63.
- Baetu, T. M. (2015b) – 'The Completeness of Mechanistic Explanations', *Philosophy of Science*, 82(5): 775–786.
- Barkai, N. and Leibler, S. (1997) - 'Robustness in Simple Biomedical Networks', *Nature*, 387: 913–7.
- Batterman, R. and Rice, C. (2014) – "Minimal Model Explanations", *Philosophy of Science*, 81(3): 349-376.
- Baysan, U. (2015) – "Realization Relations in Metaphysics", *Minds & Machines* 25, 247–260
- Bechtel, W. (2007) – "Biological Mechanisms: Organized to Maintain Autonomy", in *F. Boogerd, J. H. Bruggeman, H. V. Hofmeyr, and H. Westerhoff (Eds.), Systems biology. Philosophical foundations.* Dordrecht: Elsevier: 269-302
- Bechtel, W. (2008) – *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*, New York: Lawrence Erlbaum Associates.

- Bechtel, W. (2011) – "Mechanism and Biological Explanation", *Philosophy of Science* 2011 78:4: 533-557
- Bechtel, W. and Abrahamsen, A. (2005) – "Explanation: A Mechanistic Alternative", *Studies in History and Philosophy of the Biological and Biomedical Sciences,* 36: 421–441.
- Bechtel, W. and Abrahamsen, W. (2009) – "Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations", In *Leitgeb and Hieke (eds.), Reduction: Between the Mind and the Brain.* Ontos: 177-190.
- Bechtel W and, Abrahamsen A. (2010) – "Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science", *Stud Hist Philos Sci.* 2010;41(3): 321-333.
- Bechtel, W. and Abrahamsen, A. (2013) – "Thinking Dynamically About Biological Mechanisms: Networks of Coupled Oscillators". *Found Sci* 18: 707–723.
- Bechtel, W. and Richardson, R.C. (1993) (reprint 2010) – *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Cambridge, MA: MIT Press/Bradford Books.
- Becskei, A. and Serrano, L. (2000) – "Engineering stability in gene networks by autoregulation" *Nature* 405: 590-593.
- Bedau, M.A. (1997) – "Weak Emergence", *Philosophical Perspectives,* 11: 375–399.
- Berg J.M., Tymoczko J.L., Stryer L. (2002) - *Biochemistry*. 5th edition. New York: WH Freeman
- Bich, L. (2016) – 'Systems and Organizations Theoretical tools, conceptual distinctions and epistemological implications', in *G. Minati, M, Ambram, E. Pessa (eds.) (2016) - Towards a Post-Bertalanffy Systemics*, New York : Springer: 203-209
- Bich, L. (2018) – 'Robustness and autonomy in biological systems: how regulatory mechanisms enable functional integration, complexity and minimal cognition through the action of second-order control constraints', in *M. Bertolaso, S. Caianiello and E. Serrelli (eds.) (2018), 'Biological Robustness. Emerging Perspectives from within the Life Sciences'*, New York: Springer: 123-147
- Bishop, R. (2011) – "Metaphysical and Epistemological Issues in Complex Systems", in *Hooker, C. (ed) (2011) – Philosophy of Complex Systems,* Elsevier: North Holland : 105-136.
- Bokulich, A. (2016) – "Fiction as a Vehicle for Truth: Moving Beyond the Ontic Conception", *The Monist* 99 (3): 260-279
- Boogerd, F; Bruggeman, J.H.; Hofmeyr, H.V. and Westerhoff, H (eds.) (2007) – *Systems biology. Philosophical foundations*. Dordrecht: Elsevier:

- Boogerd, F.C.; Bruggeman, F.J. and Richardson, R.C. (2013) – "Mechanistic Explanations and Models in Molecular Systems Biology", *Found Sci* 18**:** 725-744.
- Braillard, P. (2010) - 'Systems Biology and the Mechanistic Framework', *History and Philosophy of Life Sciences,* 32: 43–62.
- Braillard, P.A. and Malaterre, C. (eds.) (2015) – *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, Dordrecht: Springer.
- Brigandt, I. (2015) – "Evolutionary Developmental Biology and the Limits of Philosophical Accounts of Mechanistic Explanation" in *Braillard, P.A. and Malaterre, C. (eds) Explanation in biology : an enquiry into the diversity of explanatory patterns in the life sciences*, Dordrecht : Springer, [2015]: 135-74.
- Brigandt, I.; Green, S. and O'Malley, M. (2017) – "Systems Biology and Mechanistic Explanation", in *Glennan, S. and Illari, P. (eds) - The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Taylor & Francis: 362-374.
- Broad, C.D. (1925) – *The Mind and Its Place in Nature*, London: Routledge & Kegan Paul.
- Campbell, R.J. and Bickhard, M.H. (2011) – "Physicalism, emergence and downward causation", *Axiomathes,* 21(1): 33–56.
- Cartwright, N., 1983, *How the Laws of Physics Lie*, Oxford: Clarendon Press.
- Colombo, M.; Hartmann, S. and van Iersel, R. (2015) – "Models, Mechanisms, and Coherence", *The British Journal for the Philosophy of Science,* Volume 66, Issue 1, March 2015: 181–212.
- Conesa, A. and Mortazavi, A (2014) – "The common ground of genomics and systems biology", BMC Syst Biol 8, S1 (2014). https://doi.org/10.1186/1752-0509-8-S2-S1.
- Cosentino, C. and Bates, D. (2011) - *An Introduction to Feedback Control in Systems Biology*, CRC Press (2011): Boca Raton
- Craver, C.F. (2007) – *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press.
- Craver, C.F. (2014) – "The Ontic Account of Scientific Explanation", In *Kaiser, Scholz, Plenge, and Hüttemann (eds.) (2014) - Explanation in the Special Sciences: The Case of Biology and History*. Springer: Verlag: 27-52.
- Craver, C.F and Bechtel, W. (2007) – "Top-down Causation Without Top-down Causes", *Biology & Philosophy,* September 2007, Volume 22, Issue 4: 547–563.
- Craver C.F. and Wilson, R. (2006) – "Realization", In *Thagard, P.; Gabbay, D.M. and Woods, J. (eds) (2006) Philosophy of Psychology and Cognitive Science*, Amsterdam: North Holland: 81–104.

- 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*, Volume 19, Issue 4, August 2008*:* 389-395.
- Darden, L. (2006) – *Reasoning in Biological Discoveries: Mechanism, Interfield Relations, and Anomaly Resolution*, New York: Cambridge University Press.
- de Regt, H.W. (2009) – "Understanding and Scientific Explanations", in *de Regt, H.W., Leonelli, S. and Eigner, K. (2009) - Scientific Understanding: Philosophical Perspectives*, University of Pittsburgh Press (2009): 21-42.
- de Regt, H.W.; Leonelli, S. and Eigner, K. (2009a) - *Scientific Understanding: Philosophical Perspectives*, University of Pittsburgh Press (2009)
- de Regt, H.W.; Leonelli, S. and Eigner, K. (2009b) – "Focussing on Understanding", in *de Regt, H.W., Leonelli, S. and Eigner, K. (2009) - Scientific Understanding: Philosophical Perspectives*, University of Pittsburgh Press (2009): 1-17.
- Dekel, E.; Mangan, S.; Alon, U. (2005) – "Environmental selection of the feed-forward loop circuit in gene-regulation networks", *Phys Biol.* 2005;2(2): 81-88.
- di Bernardo, D.; Thompson, M.J.; Gardner, T.S.; Chobot, S.E.; Eastwood, E.L.; Wojtovich, A.P.; Elliott, S.J.; Schaus, S.E.; Collins, J.J. (2005) – 'Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks', *Nat Biotechnol*, 2005 Mar;23(3): pp377-83.
- Diéguez, A. (2013) – "When Do Models Provide Genuine Understanding, and Why Does It Matter?", *History and Philosophy of the Life Sciences*, Vol. 35, No. 4, 2013: 599–620.
- Eldar, A.; Dorfman, R.; Weiss, D.; Ashe, H.; Shilo, B.Z.; Barkai, N. (2002) – "Robustness of the BMP morphogen gradient in Drosophila embryonic patterning", *Nature* 419(6904): 304-308.
- Emonet T. and Cluzel P. (2008) – 'Relationship between cellular response and behavioral variability in bacterial chemotaxis', *Proceedings of the National Academy of Sciences of the United States of America*, 2008 Mar;105(9): 3304-3309.
- Enver, T.; Pera, M.; Peterson, C.; Andrews, P.W. (2009) – "Stem cell states, fates, and the rules of attraction", *Cell Stem Cell.* 2009;4(5): 387-397.
- Fagan, M.B. (2015) – "Collaborative explanation and biological mechanisms", *Studies in History and Philosophy of Science* Part A, Vol 52: 67-78.
- Fisher, R.A. (1930) – *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford
- Friedman, M. (1974) – "Explanation and Scientific Understanding", *Journal of Philosophy*, 71: 5–19.
- Gánti, T. (2003) – *The principles of life*, New York: Oxford.

- Gimlich, R.L. and Gerhart, J.C. (1984) – "Early cellular interactions promote embryonic axis formation in Xenopus laevis", *Developmental Biology* 1984 Jul;104(1): 117-30.
- Glennan, S.S. (1996) – "Mechanisms and The Nature of Causation", *Erkenntnis,* 44: 49–71.
- Glennan, S.S. (2002) – "Rethinking Mechanistic Explanation", *Philosophy of Science,* 69*:* S342–S353.
- Glennan, S. and Illari, P. (eds) (2017) – *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Taylor & Francis.
- Godfrey-Smith, P. (2009a) – "Models and fictions in science". *Philos Stud* 143: 101–116.
- Godfrey-Smith, P. (2009b) – "Abstractions, Idealizations, and Evolutionary Biology", In *A. Barberousse, M. Morange, and T. Pradeu (Eds.), Mapping the future of biology: Evolving concepts and theories*, Berlin: Springer: p47.
- Gorur-Shandilya, S.; Demir, M.; Long, J.; Clark, D. A. and Emonet, T. (2017) – 'Olfactory receptor neurons use gain control and complementary kinetics to encode intermittent odorant stimuli', *eLife, 6,* e27670. https://doi.org/10.7554/eLife.27670
- Green, S. (ed.) (2017) - *Philosophy of Systems Biology: Perspectives from Scientists and Philosophers*, Springer International Publishing. doi:10.1007/978-3-319-47000-9
- Green, S., and Jones, N. (2016) – "Constraint-Based Reasoning for Search and Explanation: Strategies for Understanding Variation and Patterns in Biology", *Dialectica*, 70: 343–374.
- Green, S.; Levy, A. and Bechtel, W. (2015) – "Design sans adaptation", *European Journal for Philosophy of Science* 5 (2015): 15-29.
- Green, S.; Şerban, M.; Scholl, R.; Jones, N.; Brigandt, I and Bechtel, W. (2018) – "Network analyses in systems biology: new strategies for dealing with biological complexity", *Synthese* 195:1751–1777.
- Green, S. and Wolkenhauer, O. (2013) – "Tracing Organizing Principles: Learning from the History of Systems Biology", *History and Philosophy of the Life Sciences,* 35: 553-576.
- Griesemer, J. and Szathmáry, E. (2009) – "Gánti's Chemoton Model and Life Criteria," in *Rasmussen, Chen, Packard, Bedau, Deamer, Krakauer, and Stadler (eds*.) (2009) - *Protocells: Bridging Nonliving and Living Matter,* MIT Press: 481-512.
- Griffiths, P. and Stotz, K. (2013) – *Genetics and Philosophy: An Introduction*, Cambridge University Press: Cambridge, UK.
- Griffiths A.J.F.; Wessler S.R.; Carroll S.B. and Doebley J.F. (2015) – *Introduction to genetic analysis (11th Edition)*, New York, NY: W.H. Freeman & Company, [2015]

- Gross, F. (2015) – "The Relevance of Irrelevance: Explanation in Systems Biology". In *Braillard, P.A. and Malaterre, C. (eds.), Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Dordrecht: Springer: 175-198
- Gross, F. (2016) – "Heuristic Strategies in Systems Biology", *Journal of Philosophical Studies*, 9(30), 1-18.
- Gunawardena, J. (2014) – "Models in biology: 'accurate descriptions of our pathetic thinking'", *BMC Biol* 12, 29. https://doi.org/10.1186/1741-7007-12-29
- Güttinger, S. and Dupré, J. (2016) – "Genomics and Postgenomics", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/win2016/entries/genomics/>.
- Heinrich, R. and Schuster, S. (1998) – "The modelling of metabolic systems. Structure, control and optimality", *Biosystems.* 1998 Jun-Jul;47(1-2): 61-77.
- Hempel, C. and Oppenheim, P. (1948) – "Studies in the Logic of Explanation", *Philosophy of Science,* 15: 135–175.
- Hempel, C. (1965) – *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: Free Press.
- Hochstein, E. (2016) – "One mechanism, many models: a distributed theory of mechanistic explanation", *Synthese* 193 (5): 1387-1407.
- Hodgkin, A.L. and Huxley, A.F. (1952) – "A quantitative description of membrane current and its application to conduction and excitation in nerve", *The Journal of Physiology*. 117 (4): 500–44
- Holmes, E. E.; Lewis, M. A.; Banks, J. E.; Veit, R. R. (1994) – "Partial Differential Equations in Ecology: Spatial Interactions and Population Dynamics", *Ecology, Wiley,* 75 (1): 17–29.
- Hooker, C (ed) (2011a) – *Philosophy of Complex Systems, Volume 10 in Handbook of the Philosophy of Science*, Elsevier: North Holland.
- Hooker, C (2011b) – "Introduction to Philosophy of Complex Systems", in Hooker, C (ed) (2011) – *Philosophy of Complex Systems, Volume 10 in Handbook of the Philosophy of Science*, Elsevier: North Holland: 841-909.
- Hooker, C. (2013) – "On the Import of Constraints in Complex Dynamical Systems", *Foundations of Science,* 18: 757–780
- Hormuth, D.A., 2nd; Eldridge, S.L.; Weis, J.A.; Miga, M.I. and Yankeelov, T.E. (2018) – "Mechanically Coupled Reaction-Diffusion Model to Predict Glioma Growth: Methodological Details", *Methods in molecular biology (Clifton, N.J.), 1711*: 225–241
- Huang, S. (2004) – "Back to the biology in systems biology: What can we learn from biomolecular networks?", *Briefings in Functional Genomics,* Volume 2, Issue 4, February 2004: 279–297.

- Huang, S.; Eichler, G.; Bar-Yam, Y. and Ingber, D.B. (2005) – "Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network", *Physical Review Letters* 94(12): 1–4.
- Huang, S.; Guo, Y.P.; May, G. and Enver, T. (2007) – "Bifurcation dynamics in lineage-commitment in bipotent progenitor cells", *Developmental biology* 305 (2): 695–713.
- Huneman, P. (2010) – "Topological explanations and robustness in biological sciences", *Synthese* 177: 213–245.
- Huneman, P. (2018) – "Outlines of a theory of structural explanations", *Philosophical Studies,* 175 (3): 665-702.
- Ideker, T.; Galitski, T. and Hood, L. (2001) – "A New Approach to Decoding Life: Systems Biology", *Annual Review of Genomics and Human Genetics* 2001 2:1: 343-372.
- Illari, P. (2013) – "Mechanistic Explanation: Integrating the Ontic and Epistemic", *Erkenntnis.* 78: 237-255.
- Illari, P.M.; Russo, F. and Williamson, J. (2011) – *Causality in Science*, Oxford University Press.
- Illari, P.M. and Williamson, J. (2011) – "Mechanisms are real and local", In *Illari, Russo, and Williamson (eds), Causality in the Sciences*, Oxford: Oxford University Press: 818-44.
- Ilsley, G.R.; Luscombe, N.M.; Apweiler, R. (2009) – "Know your limits: assumptions, constraints and interpretation in systems biology", *Biochimica et Biophysica Acta,* 2009 Sep;1794(9): 1280-1287.
- Isaac S.C. and Jernvall, J. (2002) – "A gene network model accounting for development and evolution of mammalian teeth", *Proceedings of the National Academy of Sciences* Jun 2002, 99 (12): 8116-8120.
- Issad, T. and Malaterre, C. (2015) – "Are Dynamic Mechanistic Explanations Still Mechanistic?" In: *Braillard, P.A. and Malaterre, C. (Eds.). Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences,* Dordrecht: Springer: 265-292.
- Isalan, M.; Lemerle, C.; Michalodimitrakis, K.; et al. (2008) – "Evolvability and hierarchy in rewired bacterial gene networks", *Nature,* 2008;452(7189): 840-845.
- Itzkovitz, S.; Blat, I.C.; Jacks, T.; Clevers, H.; van Oudenaarden, A. (2012) – "Optimality in the development of intestinal crypts*" Cell,* 2012;148(3): 608-619.
- Jones, M. (2005) – "Idealization and abstraction: A framework" In M. R. Jones and N. Cartwright (Eds.), *Idealization XII: Correcting the model. Idealization and abstraction in the sciences* (Vol. 86) Amsterdam: Rodopi: 173–217.
- Jones, N. (2014) – "Bowtie Structures, Pathway Diagrams, and Topological Explanation", *Erkenntnis,* 79: 1135–1155.

- Kaiser, M.I.; Scholz, O.R.; Plenge, D. and Hüttemann, A. (eds.) (2014) – *Explanation in the Special Sciences: The Case of Biology and History,* Springer: Dordrecht
- Kaplan, D.M. and Craver, C. (2011) – "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective", *Philosophy of Science,* 78(4): 601-627.
- Karr; Sanghvi; Macklin; Gutschow; Jacobs; Bolival Jr.; Assad-Garcia; Glass; Covert (2012) – "A Whole-Cell Computational Model Predicts Phenotype from Genotype", *Theory,* Volume 150, Issue 2: 389-401.
- Kauffman, S.A. (1970) – "Articulation of Parts Explanation in Biology and the Rational Search for Them" In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1970: 257–272.
- Kauffman, S.A. (2000) – *Investigations*, Oxford University Press: New York
- Kim, J. (1998) – *Mind in a physical world*: *An Essay on the Mind-Body Problem and Mental Causation*, Cambridge: MIT Press.
- Kim, J. (2006) – "Emergence: Core Ideas and Issues", *Synthese,* 151(3): 547–559
- Kitano, H. (2004) – 'Biological robustness', *Nat Rev Genet* 5: pp826–837
- Kitano, H. and Oda, K. (2006) – "Robustness trade-offs and host–microbial symbiosis in the immune system". *Mol Syst Biol* 2: 1-10
- Kitcher, P. (1989) – "Explanatory Unification and the Causal Structure of the World", in *Kitcher, P. and Salmon, W. (1989) – Scientific Explanation*, Minneapolis: University of Minnesota Press: 410–505.
- Kuo, P.; Dwight, W.; Banzhaf Leier, A. (2006) – "Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence", *BioSystems* 85 (3): 177–200.
- Ladyman, J.; Lambert, J. and Wiesner, K. (2013) – "What is a complex system?", *Euro Jnl Phil Sci* 3: 33–67.
- Leonelli, S. (2007) – 'What is in a Model? Combining Theoretical and Material Models to Develop Intelligible Theories', in Laubichler, M.D. and Müller, G.B. (eds) *Modeling Biology Structures, Behaviors, Evolution, Vienna Series in Theoretical Biology,* MIT Press (2007).
- Leonelli, S. (2009) – "Understanding in Biology: The Impure Nature of Biological Knowledge", in de Regt, H.W., Leonelli, S. and Eigner, K. (2009) - *Scientific Understanding: Philosophical Perspectives*, University of Pittsburgh Press: 189-209.
- Leonelli, S. (2014) – "What Difference Does Quantity Make? On the Epistemology of Big Data in Biology", *Big Data & Society,* 1(1), April 2014
- Leonelli S. (2019) – "The challenges of big data biology" *Elife.* 2019;8*:* e47381. Published 2019 Apr 5.
- Leuridan, B. (2010) – "Can Mechanisms Really Replace Laws of Nature?", *Philosophy of Science*, 77(3): 317-340

- Levins, R. (1966) - "The strategy of model building in population biology", *American Scientist,* 54, no. 4 (1966): 421-31.
- Levins, R. and Lewontin, R. (1985) – *The Dialectical Biologist*. Harvard University Press (1985)
- Levy, A. (2013) – "Three kinds of new mechanism", *Biology and Philosophy,* 28 (1): 99-114
- Levy, A. (2018) – "Idealization and abstraction: refining the distinction", *Synthese* (2018)
- Levy, A. and Bechtel, W. (2013) - 'Abstraction and the Organization of Mechanisms', *Philosophy of Science,* 80: 241–61.
- Longino, H.E. (2005) – 'Complexity and Diversity all the way', *Metascience (2005) 14: 185–194*
- Love, A.C. and Nathan, M.J. (2015) – "The Idealization of Causation in Mechanistic Explanation", *Philosophy of Science,* 82(5): 761-774.
- Ma, W.; Ala, T.; Hana, E.S.; Lim, W.A. and Tang, C. (2009) – 'Defining Network Topologies That Can Achieve Biochemical Adaptation', *Cell* 138 (4): 760–73.
- Machamer, P.K. (2004) – "Activities and Causation: The Metaphysics and Epistemology of Mechanisms", *International Studies in the Philosophy of Science,* 18: 27–39.
- Machamer, P.K.; Darden, L.; and Craver, C.F. (2000) - "Thinking about Mechanisms", *Philosophy of Science,* 67: 1–25.
- Mangan, S. and Alon, U. (2003) – "Structure and function of the feed-forward loop network motif", *Proceedings of the National Academy of Sciences* Oct 2003, 100 (21): 11980-11985.
- Mangan, S.; Zaslaver, A. and Alon, U. (2003) – "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks". *J Mol Biol.* 2003;334(2): 197-204.
- Matthiessen D. (2017) – 'Mechanistic Explanation in Systems Biology: Cellular Networks', *The British Journal for the Philosophy of Science* 68 (1): 1–25.
- Mayr E. (1961) – "Cause and effect in biology", *Science,* 1961;134(3489): 1501–1506.
- McShea, D.W. (1996) – "Perspective: Metazoan complexity and evolution: Is there a trend?", *Evolution,* 50: 477–492.
- McShea, D.W. (2000) – "Functional Complexity in Organisms: Parts as Proxies", *Biology and Philosophy,* 15, 641-68.
- Mc Manus, F. (2012) – "Development and mechanistic explanation", *Studies in history and philosophy of biological and biomedical sciences,* 43: 532-41.
- Mello, B.A. and Tu, Y. (2003) – 'Perfect and near-perfect adaptation in a model of bacterial chemotaxis', *Biophysical journal,* 84(5): 2943–2956.

- Mill, J.S. (1843) – *System of Logic*, London: Longmans, Green, Reader, and Dyer. [8th ed., 1872.]
- Mitchell, S. (2002) – 'Integrative Pluralism', *Biology and Philosophy 17(2002): 55–70.*
- Mitchell, S. (2003) – *Biological Complexity and Integrative Pluralism*, Cambridge, 2003.
- Montévil, M. and Mossio, M. (2015) – "Biological organisation as closure of constraints", *Journal of Theoretical Biology*, Elsevier, 2015, 372: 179-91.
- Moreno, A. (2007) – "A systemic approach to the origin of biological organization", in *F. Boogerd, J. H. Bruggeman, H. V. Hofmeyr, and H. Westerhoff (Eds.), Systems biology. Philosophical foundations* Dordrecht: Elsevier: 243–268.
- Moreno, A. and Mossio M. (2015) – *Biological autonomy. A Philosophical and Theoretical Enquiry*, Dordrecht: Springer.
- Morgan, M.S. and Morrison, M. (eds.) (1999a) – *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.
- Morgan, M.S. and Morrison, M. (1999b) – "Models as Mediating Instruments", in *Morgan, M.S. and Morrison, M. (eds.), Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press: 10-37.
- Morrison, M. (1999) – Models as Autonomous Agents, in *Morgan, M.S. and Morrison, M. (eds.) (1999a), Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press: 38-65.
- Mossio, M., Saborido, C., and Moreno, A. (2009) – "An organizational account of biological functions", *British Journal for the Philosophy of Science*, 60: 813–841.
- Mossio, M., and Moreno, A. (2010) – "Organizational closure in biological organisms", *History and Philosophy of the Life Sciences*, 32(2–3): 269–288.
- Mossio, M.; Bich, L., and Moreno, A. (2013) – "Emergence, closure and inter-level causation in biological systems", *Erkenntnis*, 78(2): 153–178.
- Nathan, M.J. (2014) – "Causation by Concentration", *The British Journal for the Philosophy of Science,* Volume 65, Issue 2, June 2014: 191–212.
- Needham, C.J.; Bradford, J.R.; Bulpitt, A.J. and Westhead, D.R. (2007) – "A primer on learning in Bayesian networks for computational biology", *PLoS Comput Biol.* 2007 Aug;3(8): e129.
- Nicholson, D.J. (2012) – "The Concept of Mechanism in Biology", *Studies in History and Philosophy of Science* Part C 43 (1): 152-163.
- O'Malley, M.A. and Dupré, J. (2005) – "Fundamental issues in systems biology", *Bioessays*. 2005;27(12): 1270-1276.
- Piccinini, G. and Craver, C.F. (2011) – "Integrating psychology and neuroscience: Functional analyses as mechanism sketches". *Synthese,* 183: 283–311.

- Powell, A. and Dupré, J. (2009) – "From molecules to systems: the importance of looking both ways*", Studies in history and philosophy of biological and biomedical sciences* 40 (1): 54–64.
- Rao, C.V.; Kirby, J.R. and Arkin, A.P. (2004) - 'Design and Diversity in Bacterial Chemotaxis: A Comparative Study in *Escherichia coli* and *Bacillus subtilis'*, *PLoS Biology*, 2: 0239–52.
- Rao C.V.; Glekas G.D. and Ordal G.W. (2008) – 'The three adaptation systems of *Bacillus subtilis* chemotaxis', *Trends Microbiol.* 2008;16(10): 480-487.
- Rice, C. (2015) – "Moving Beyond Causes: Optimality Models and Scientific Explanation", *Noûs,* 49: 589-615.
- Rice, C. (2019) – "Models Don't Decompose That Way: A Holistic View of Idealized Models", *The British Journal for the Philosophy of Science,* Volume 70, Issue 1, March 2019: 179–208.
- Rosen, R. (1991) – *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*, Columbia University Press: New York.
- Rheinberger, H. (1997) – "Experimental Complexity in Biology: Some Epistemological and Historical Remarks", *Philosophy of Science* 64 (Supplement): S245–S254.
- Rosslenbroich, B. (2016) – "Alvaro Moreno and Matteo Mossio: Biological autonomy: a philosophical and theoretical enquiry", *Biology & Philosophy.* 31: 591–601.
- Ruiz-Mirazo, K., and Moreno, A. (2004) – "Basic autonomy as a fundamental step in the synthesis of life", *Artificial Life,* 10: 235-259.
- Salazar-Ciudad, I. and Jernvall, J (2002) – "A gene network model accounting for development and evolution of mammalian teeth", *Proceedings of the National Academy of Sciences* Jun 2002, 99 (12): 8116-8120.
- Salmon, W.C. (1984) – *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.
- Salmon, W. (1997) – "Causality and Explanation: A Reply to Two Critiques", *Philosophy of Science,* 64: 461–477.
- Schoenwolf, G.C.; Bleyl, S.B.; Brauer, P.R. and Francis-West, P.H. (2009) – *Larsen's Human Embryology (5th Edition),* Elsevier: Philadelphia
- Schrödinger, E. (1944) – *What is Life? The Physical Aspect of the Living Cell*, Cambridge University Press.
- Seo, H.; Kim, W.; Lee, J.; and Youn, B. (2013) – "Network-based approaches for anticancer therapy (Review)". *International Journal of Oncology*, 43: 1737-1744.
- Shen-Orr, S.S.; Milo, R.; Mangan, S.; Alon, U. (2002) – "Network motifs in the transcriptional regulation network of *Escherichia coli*" *Nat Genet.* 2002;31(1): 64-68.

- Sheredos, B. (2016) – "Re-Reconciling the Epistemic and Ontic Views of Explanation (Or, Why the Ontic View Cannot Support Norms of Generality)," *Erkenntnis*, 81: 919–949.
- Simon, H.A (1962) – "The Architecture of Complexity", *Proceedings of the American Philosophical Society* 106 (6): 467–482.
- Singer, S.J. and Nicolson, G.L. (1972) – "The fluid mosaic model of the structure of cell membranes", *Science,* 175 (4023)*:* 720–31.
- Sober, E. (1983) – "Equilibrium explanation", *Philos Stud* 43: 201–210.
- Solé, R.V. and Valverde, S. (2006) – "Are network motifs the spandrels of cellular complexity?", *Trends in ecology & evolution* 21(8): 419–422.
- Stelling, J.; Sauer U.; Szallasi Z.; Doyle F.J. and Doyle, J. (2004) – 'Robustness of Cellular Functions', *Cell,* Volume 118, Issue 6, September 17, 2004, 675-685.
- Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.
- Suppes, P. (1960) – "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences," *Synthese*, 12 (2-3): 287–301.
- Tabery, J. (2004) – "Synthesizing Activities and Interactions in the Concept of a Mechanism", *Philosophy of Science*, 71: 1–15.
- Théry, F. (2015) – "Explaining in Contemporary Molecular Biology: Beyond Mechanisms", In *P.A. Braillard and C. Malaterre (eds.), Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Springer: 113-143.
- Tinbergen, N. (1963) - "On Aims and Methods in Ethology," *Zeitschrift für Tierpsychologie*, 20: 410–433.
- Tindall, M.J.; Porter, S.L.; Maini, P.K.; Gaglia, G. and Armitage, J.P. (2008) - 'Overview of Mathematical Approaches Used to Model Bacterial Chemotaxis I: The Single Cell', *Bulletin of Mathematical Biology,* 70: 1525–69.
- Trout, J.D. (2002) – "Scientific explanation and the sense of understanding", Science, 69: 212-33.
- Tyson, J.J. and Novák, B. (2010) – "Functional motifs in biochemical reaction networks", *Annu Rev Phys Chem.* 2010;61: 219-240.
- van Eck, D. (2015) – "Reconciling Ontic and Epistemic Constraints on Mechanistic Explanation, Epistemically", *Axiomathes* 25: 5–22.
- van Fraassen, B. (1970) – "On the Extension of Beth's Semantics of Physical Theories", *Philosophy of Science* 37: 325-39.
- van Fraassen, B. (1980) – *The Scientific Image*, Oxford University Press: Oxford, UK.
- von Bertalanffy, L. (1968) – *General System Theory: Foundations, Development, Applications*. New York: George Braziller.

- Waskan, J. (2011) – "Mechanistic explanation at the limit", *Synthese*, December 2011, Volume 183, Issue 3: 389–408.
- Waters, C.K. (2014) – "Shifting Attention from Theory to Practice in Philosophy of Biology", In: *Galavotti M., Dieks D., Gonzalez W., Hartmann S., Uebel T., Weber M. (eds) New Directions in the Philosophy of Science. The Philosophy of Science in a European Perspective, vol 5*, Berlin: Springer International Publishing: 121-39.
- Watts, D. and Strogatz, S. (1998) – "Collective dynamics of 'small-world' networks". *Nature* 393(1998): 440–442.
- Whitacre, J. M. (2012) – 'Biological robustness: paradigms, mechanisms, and systems principles', *Frontiers in genetics*, 3(2012), 67.
- Weisberg, M. (2007a) – "Three kinds of idealization", *Journal of Philosophy*, 104(12), 639–659.
- Weisberg, M. (2007b) – "Who Is a Modeler?", *The British Journal for the Philosophy of Science*, 58(2): 207-233.
- Weisberg, M. (2013) – *Simulation and Similarity: Using Models to Understand the World*, New York: Oxford University Press.
- Westerhoff, H.V. and Palsson, B.O (2004) – "The evolution of molecular biology into systems biology", *Nat Biotechnol.* 2004; 22(10): 1249-1252.
- Wilson, J. (2010) – "Non-reductive Physicalism and Degrees of Freedom", *The British Journal for the Philosophy of Science*, Volume 61, Issue 2, June 2010: 279–311.
- Wilson, J. (2011) – "Non-reductive Realization and the Powers-based Subset Strategy", *The Monist*, 94(1): 121-154.
- Wimsatt, W.C. (1972) – "Complexity and Organization", in *PSA: Proceedings of the BiennialMeeting of the Philosophy of Science Association*, Volume 1972, 67–86.
- Wimsatt, W.C. (2007) - *Re-engineering philosophy for limited beings: piecewise approximations to reality*. Cambridge: Harvard University Press.
- Winning, J. and Bechtel, W. (2018) – "Rethinking Causality in Biological and Neural Mechanisms: Constraints and Control", *Minds & Machines* 28: 287–310.
- Winther, R.G. (2003) – "Formal biology and compositional biology as two kinds of biological theorizing" PhD thesis submitted to Department of History and Philosophy of Science, Indiana University. Accessed from https://www.researchgate.net/publication/228537881
- Winther, R.G. (2006a) – "Parts and Theories in Compositional Biology", *Biol Philos* 21: 471–499.
- Winther, R.G. (2006b) – "On the dangers of making scientific models ontologically independent: taking Richard Levins' warnings seriously". *Biol Philos* 21(2006): 703–724.

- Woodward, J. (2000) – "Explanation and invariance in the special sciences*", The British Journal for the Philosophy of Science,* Volume 51, Issue 2, June 2000: 197–254.
- Woodward, J. (2002) – "What Is a Mechanism?: A Counterfactual Account", *Philosophy of Science,* 69: S366–S377.
- Woodward, J. (2003) – *Making Things Happen: A Theory of Causal Explanation,* New York: Oxford University Press.
- Woodward, J. (2013) – "Mechanistic Explanation: Its Scope and Limits", *Aristotelian Society Supplementary Volume,* Volume 87, Issue 1, 1 June 2013: 39–65.
- Woodward, J. (2018) – "Some Varieties of Non-Causal Explanation", in *Reutlinger, A. and Saatsi. J. (eds.) (2018) - Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations,* Oxford: Oxford University Press: 114-137.
- Wouters, A.G. (2007) – 'Design explanation: determining the constraints on what can be alive', *Erkenntnis* (2007) 67: 65–80.
- Wouters, A.G. (2013). Biology's functional perspective: roles, advantages, and organization. In *K. Kampourakis* (Ed.), *Philosophy of biology: A companion for educators* Berlin: Springer: 455–486
- Wright, C.D. (2012) - "Mechanistic explanation without the ontic conception", *Euro Jnl Phil Sci* 2, 2012: 375–394.
- Wright, C.D. (2015) – "The ontic conception of scientific explanation", *Studies In History and Philosophy of Science* Part A. 54: 20-30.
- Yi, T.; Huang, Y.; Simon, M. and Doyle, J. (2000) - 'Robust Perfect Adaptation in Bacterial Chemotaxis through Integral Feedback Control', *Proceedings of the National Academy of Sciences,* 97: 4649–53.
- Zarai, Y. and Tuller, T. (2019) Correction: "Computational analysis of the oscillatory behavior at the translation level induced by mRNA levels oscillations due to finite intracellular resources", *PLoS Comput Biol* 15(7): e1007192