Researchers' assumptions and mathematical models: A philosophical study of metabolic systems biology

Submitted by Josephine Donaghy, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Philosophy, August 2014.

Abstract

This thesis examines the philosophical implications of the assumptions made by researchers involved in the development of mathematical models of metabolism. It does this through an analysis of several detailed historical case studies of models between the 1960's and the present day, thus also contributing to the growing literature on the historiography of biochemical systems biology. The chapters focus on four main topics: the relationship between models and theory, temporal decomposition as a simplifying strategy for building models of complex metabolic systems, interactions between modellers and experimental biochemists, and the role of biochemical data. Four categories of assumptions are shown to play a significant role in these different aspects of model development; ontological assumptions, idealising assumptions, assumptions about data, and researchers' commitments. Building on this analysis, the thesis brings to light the importance of researcher's ontological and idealising assumptions about the temporal organisation, alongside the spatial organisation, of metabolic systems. It also offers an account of different forms of interactions between research groups – hostile interactions, closed collaboration, and open collaboration – on the basis of differences in the characteristics of researcher's commitments. Throughout the case studies, biological data play a powerful role in model development by virtue of the contents of available data sets, as well as researchers' perceptions of those data, which are in turn influenced by their ontological assumptions. The historical trajectories explored illustrate how the relationships between different facets of model building, and their associated philosophical abstractions, are often best understood as transient features within a highly dynamic research process, whose role depends on the specific stage of modelling in which they are enacted. This thesis provides an expanded perspective on the different types and roles of assumptions in the development of mathematical models of metabolism, which is firmly grounded in a historical analysis of scientific practice.

Table of Contents

List of Figures

Acknowledgements

My first supervisor Sabina Leonelli played several pivotal roles over my time as a Master's and later PhD student. She patiently read many very rough drafts and carefully guided me in developing the necessary research skills to think and communicate clearly. I always left our supervision meetings with far more work than I had anticipated, but also the energy and enthusiasm with which to tackle it. My interactions with her nourished a steady growth in my personal resilience and professional confidence. Equally as importantly, she accepted, even encouraged, the fact that my PhD research was not my main raison d'être allowing me the space to peruse other avenues of interest and enjoyment. I will make use of all of these wonderful experiences for many years to come.

John Dupré, my second supervisor, contributed a much appreciated additional pair of expert eyes. In the final stages he provided the reassurance that things were going well and that the end really was in sight. A special thanks to my mentor, Dario Castiglione, – initially my second supervisor – who tolerated a lot of early enthusiastic but confused ideas before I settled down to a more grounded research topic.

My apprehension in exposing my work to the wider academic community was often unfounded. I want to thank the anonymous reviewers of two published articles based on two of the chapters in this thesis for engaging with my work and helping it fulfil its potential. Being drawn towards performance arts, giving conference presentations was one the most enjoyable aspects of the PhD. Thank you to all the known and unknown faces who engaged with my work in this setting. In any future research I intend to make more use of social interactions and collaboration than I did in this PhD. My viva provided a forum for putting this sentiment into practice, Thank you to my examiners Ulrich Krohs and Staffan Müller-Wille for this productive experience.

The group of staff and PhD students at Egenis provided a very supportive atmosphere for this project. A special thank you to David Wyatt for calmly guiding me through the technical frustrations of word and endnote formatting and last minute copy-editing. Sara Green, a visiting PhD student also working on mathematical models and systems biology, provided much needed peer support and

encouragement. Other Egenis researchers past and present deserving special mention for their positive influence on this project include: Ann-Sophie Barwich, Louise Bezhuidenhout, Nicholas Binney, Tarquin Holmes, Nadine Levine, James Lowe, Maureen O'Malley, Pierre Olivier Méthot, Selina Nath, Mila Petrova, and Adam Toon.

One of the highlights of the past four years was organising the Systems Biology Skills X-change workshop in spring 2013 alongside PhD students from the biosciences and engineering departments (See Appendix A). This workshop was funded by the systems biology research cluster at the University of Exeter and aimed to provide early career researchers with a space to discuss the challenges and opportunities posed by interdisciplinary research. It was particularly rewarding to work alongside Varun Kothamachu whose open and enquiring mind helped create a space for dialogue.

At the beginning of my second year I embarked on a creative diversion by entering the annual Dance your PhD competition whilst I tried to figure out what a PhD research project actually involved (See Appendix B). It ended up being a delightfully ramshackle first foray into the arts, and a steep learning curve that unfortunately I have had no time to pursue since. A huge thank you to the large number of participants from Devon and Exeter's local dance and film making scene who volunteered a lot of time on the project. Particular thanks to Benjamin Borley for coordinating the film making, and Noel Perkins for guidance on choreography, I wish I had listened to you both more.

Moving even further astray from the confines of academia, my Buddhist sangha in Exeter provided a much needed weekly space for mindfulness and reflection keeping me centred throughout the up's and down's of life as a PhD student. As a counterpoint I want to thank the members of my family, past and present, who tend to live life at the extremes encouraging the exploration of the limits of possibility and impossibility.

Many friends have made my PhD experience not just bearable but enjoyable. Special thanks to: James King for dancing, Hannah Standen and Al Inglis for fires, and Josh Dart and Flo Crawley for looking on the bright side. And, of course, Jani, who held my hand on the wilds of Dartmoor.

1 Introduction

This thesis explores the entangled relationship between what we know and how we know. Specifically, it addressed the relationship between researchers' assumptions and the development of mathematical models of metabolism. Assumptions include what is being taken for granted rather than what is being questioned and investigated. They are often a low profile facet of research; by definition they will never have the status of technological or theoretical achievements. However, they play a significant role in the development of research tools used for the asking and answering of questions. This research was carried out in relation to an open ended question about what assumptions researchers make and how they play a role in the development of models of metabolism. Its analysis is informed by looking at the scientific practices of mathematical model development through historical case studies.

There are a wide variety of ways in which researchers assumptions may intersect with the development of mathematical models. Given that researchers anticipate that mathematical models will make a contribution to knowledge and understanding the epistemic evaluation of models is likely to play a significant role in their development. This process will involve researchers making assumptions about things like the relationship between completed versions of models and real world systems, and the relationship between results derived from models and epistemic claims about those systems. However, before this can take place models need to be constructed. Model building often involves the integration of diverse ingredients including; theory, data, equations, research questions, and researchers assumptions. The quality of these ingredients and the quality of the processes of model assembly are likely to impact on the epistemic contributions of the completed model. Therefore assumptions which play a role in the evaluation of ingredients and construction processes will also have an impact on model development. The fact that models are built from a diversity of things, and that those things need to be bought together into a coherent model, means that model development is also shaped by practicalities alongside epistemic concerns. Researchers may encounter constraints on model building stemming from the availability and quality of ingredients, or technical limitations encountered during their integration. Assumptions may also play

a significant role in how researchers address these constraints, again influencing the trajectory of model development. Shifting attention from how available ingredients are put together, to where those ingredients and the expertise required to put them together comes from, highlights another way in which assumptions can affect model development. Mathematical models of metabolism are often sites of interactions between different communities of researchers as diverse ingredients from mathematical formulations, to experimental results, and specific biological knowledge, come into contact with each other. Assumptions which have an impact on social interactions may also have important effects on the form and dynamics of model development.

The philosophical contributions of this thesis are based on a methodology involving the analysis of historical cases of science in practice which is introduced in the second chapter of this this thesis. Chapter's three to six are derived from paying attention to what appear to be the most significant factors affecting model development in each of the cases examined. Each of these chapters involves the philosophical analysis of between one to three historical case studies of specific approaches to building mathematical models of metabolism. They bring out a variety of assumptions which researchers make and the impact they have on mathematical modelling. Chapter two contextualises these chapters in relation to existing philosophical research on assumptions in mathematical modelling and biological research. The different assumptions highlighted throughout this thesis include: ontological, or ontic assumptions about the spatial, functional and temporal organisation of systems; idealising assumptions which facilitate the construction of working models within the limits of particular aims and constraints; assumptions about the contribution that particular inputs, most significantly biological data, make to the mathematical representation of the metabolic system; and researchers commitments, their assumptions about best practice in research. The specific impacts on model development examined in each chapter are: chapter three - the relationship between models and theory; chapter four – the role of assumptions about temporal organisation; chapter five – the relationship between assumptions and different types of social interactions between research communities; and chapter six – the role of assumptions in relation to changes in available data resources. Chapter seven brings together threads from all the chapters discussing the

philosophical implications of the various assumptions and their role in the development of mathematical models.

What is also characteristic of this thesis is the fact that I look at the interplay between different types of assumptions and the research elements mentioned above through the lens of history, and specifically by analysing in detail the way in which specific key findings and techniques in metabolic systems biology were developed and used between 1960 and today. Examining how these relationships unfold over time allows this thesis to draw out a variety of dynamics, from continuity and stasis to transience and innovation, that arise out of the intersection of assumptions and the various facets of research surrounding the construction of mathematical models of metabolism. This historical period provides a fertile ground for interrogating the role of assumptions in research as it is an era in which models are under continual development. As such, mathematical, experimental, and social aspects of research are being exposed to each other and coming together in different configurations, bringing assumptions to the forefront as they are exposed, challenged, held on to, or revised. Using history strongly informs my philosophical views, and it also means that the historical analysis reported in this thesis contributes to the growing historiography on biochemistry. These historical aspects of the thesis are contextualised in chapter two.

The historical cases span the period between the small scale interest in using mathematical models to investigate biochemical systems in the 1960's through to the current period of extensive growth in data intensive biochemical systems biology. The shift towards data intensive systems biology has led to qualitative and quantitative changes in many aspects of research practices: the production, management, and use of biological data (Leonelli, 2012), social and political organisation of research, including changes in international and interdisciplinary collaboration (Calvert and Fujimura, 2011), the form of research questions and their relationship to data production (Krohs, 2011), shifts in the significance and meaning of biological systems (Nicholson, 2014; O'Malley and Dupré, 2005), and the relationship between mathematical models and biological explanations (Brigandt, 2013b). Yet the historical development of the mathematical models used in this area of research has received little detailed attention. Mathematical modelling, and the

assumptions upon which model development is based, play a crucial role in bringing together these different dimensions of research; data, social organisation, ontology, mathematics, research questions, and explanations. This thesis also contributes to the historical contextualisation of philosophical perspectives on current research practices involving mathematical models of biochemical systems.

1.1 Thesis Overview

The second chapter provides the methodological, philosophical and historical contextualisation of work carried out in the thesis. It highlights what work has already been carried out and where this thesis will contribute to and extend current perspectives. I examine the methodological issues surrounding research in the history and philosophy of science. Next I provide an overview of current philosophical perspectives on the role of assumptions in mathematical modelling and biological research. Finally, I introduce historical literature on the history of biochemistry and biochemical systems biology.

The third chapter in this thesis looks at the history of the model metabolic control analysis and examines the models relationship to developments in theories of metabolic control. The pre-cursor to metabolic control analysis was initially developed in the early 1960's and it is an early attempt to interest experimental biochemists in the using mathematical models to investigate the control of metabolism. The chapter contributes to the models as mediator's framework by proposing a clarification of the potentially paradoxical claim that mathematical models have a partially independent and an autonomous relationship with theory (Morrison and Morgan, 1999b). It examines the role that discrepancies in the ontological and idealising assumptions involved in model building can play in facilitating theoretical change. I argue that a coherent account of partial independence and autonomy requires a historical perspective bringing to light how the relationship between model and theory changes over time. A version of this chapter has been published in The History and Philosophy of the Life Sciences (Donaghy, 2013).

The fourth chapter of this thesis involves an analysis of three approaches to building mathematical models: relaxation times, biochemical systems theory, and

constraint based optimisation. It examines the important role that ontological assumptions about the temporal organisation of metabolic reactions play in facilitating idealising assumptions used during mathematical model development. In relation to William Bechtel and colleagues' framework of structural and functional decomposition, I argue that temporal decomposition should be recognised as another important strategy for investigating complex biological systems (Bechtel and Richardson, 1993; Bechtel and Abrahamsen, 2005). A version of this chapter has been accepted for publication in Studies C: The History and Philosophy of the Biological and Biomedical Sciences (Donaghy, Forthcoming).

Chapter five looks at the relationship between different communities of researchers in early biochemical systems biology – 1) metabolic control analysts and experimental biochemists, and 2) metabolic control analysts and biochemical systems theorists. In this period interactions were frequently hostile. Different groups occupied attacking and defensive positions leading to stagnation in research. I provide a comparative analysis of these cases studies with work on collaboration by Peter Galison and Nancy Nersessian, highlighting three different types of interaction: hostile interactions, closed collaboration, and open collaboration. I argue that these are related to different characteristics of researchers commitments; their assumptions about best practice in research. The form of interactions depends on whether researcher's commitments are rigid or flexible, and whether they view differences with other cultures in a negative, neutral, or positive light.

Chapter six, examines the development of a mathematical model, constraint based optimisation, from its initial inception in the 1980's through its transition into current data intensive biochemical systems biology. I argue that the impact of data on research is affected in two ways: Firstly, by the contents of available data, secondly, researchers' perception of that data. The second dimension is affected by researcher's ontological and idealising assumptions as well as their assumptions about the contribution which data makes to the mathematical representation of the system. I highlight the important role that open collaborative interactions play in mediating between these two aspects of research. Open collaborative interactions help researchers address constraints at the same time as imposing new standards for them to operate within. I explore how researchers assumptions and perception of data change over time and highlight the multiple and transient relationships which

can occur between mathematical models, data, research communities, and researchers assumptions.

Chapter seven draws material from the different chapters of the thesis together and discusses its philosophical and historical contributions to understanding how researchers' assumptions affect the development of mathematical models of metabolism. Firstly, I review four categories of assumptions which have played significant roles; ontological assumptions, idealising assumptions, assumptions about data, and researchers' commitments, and illustrate them with examples from the chapters. Secondly, I highlight the main philosophical contributions of the thesis, reassessing the role of reductionism in modelling in light of assumptions about data, examining the status, idealising or ontological, of assumptions about the temporal organisation of metabolic systems, and looking at the role of discrepancies in assumptions in relation to research dynamics ranging from stagnation to innovation. Finally, I look at the historical implications of this thesis including its perspective on the history of contemporary systems biology, and the value of historical case studies for philosophy of science.

The final concluding chapter provides a summary of the main philosophical and historical contributions of the thesis. It also highlights limitations with the work carried out and potential avenues for further research.

2 Methodological, philosophical, and historical context

This chapter lays out the territory of this thesis, providing the methodological, philosophical, and historical contextualisation required to make sense of the work carried out in the subsequent chapters. In section 2.1.1 I discuss methodological issues relevant for research in the history and philosophy of science and highlight current concerns about discerning the particular contribution that historical case studies can offer within a philosophy of science informed by multiple disciplinary approaches to examining science in practice. Section 2.1.2 discusses the particular methodology used to generate the material in this thesis and highlights the value that a historical approach adds to the philosophical perspectives generated in this thesis. Section 2.2 reviews the bodies of philosophical literature to which this thesis contributes and expands, including work on; mathematical modelling and assumptions in biological research, with a particular focus on work concerning assumptions about the temporal organisation of biological systems. The final section 2.3 presents historically informed philosophical work on biochemistry and biochemical systems biology, and illustrates a gap is this literature to which this thesis makes a contribution.

2.1 Methodology

2.1.1 Why historical case studies?

This thesis fits under the umbrella of integrated history and philosophy of science. It involves the integration of historical case studies and philosophical analysis. This formulation highlights two sets of interrelated methodological questions. Firstly, what are historical case studies, how are they constructed, and what is their relationship to history of science as a whole? Secondly, what is the relationship between historical case studies and philosophical issues? The first question focuses on the process of building a historical case study, and the second on the philosophical contribution that historical case studies can make. These two aspects of research in integrated history and philosophy of science may be more or less intertwined. The 1960's and 1970's saw an increase in interest in the relationship between historical and philosophical perspectives on science. Ideas about what this relationship is and the appropriate research methodology have

continued to shift over time. In this section I review these changes, closing with an identification of the current pertinent questions for integrated history and philosophy of science which this thesis contributes to. My presentation of this history is influenced by the recent work of Jutta Schickore (2011). Following Schickore, I use this thesis to explore the particular role that a historicist approach, that which explores the temporal dimensions of how things 'came into being' (Arabatzis and Schickore, 2012 p399), has in the current context where multiple disciplinary approaches are used for case study construction in conjunction with philosophical analysis of science in practice. In the following section, 2.1.2 My Methodology, I reflect on the particular role that historical case studies play in relation to the philosophical analysis in this thesis.

During the 1970's this debate was framed in terms of Ronald Giere's formulation: Do the philosophy and history of science have an intimate relationship or was it merely a marriage of convenience? Giere (1973) argued that it was the latter, and that claims for a relationship between the two lacked any strong conceptual rationale. Analysing the work of Thomas Kuhn, Giere claimed that elements of Kuhn's thesis were substantiated not by the historical character of the case studies, but because they were case studies of actual science. As such, philosophers should preferably invoke case studies of current scientific research in their analysis, which will provide the most accurate information about science, and need not look to the history of science. This claim indicates that Giere thought of historical case studies as those relating to past events, rather than those which examined the temporal unfolding of science regardless of their proximity to the present day. Giere's defensive attitude related to his belief at the time that philosophy of science aimed at a normative account which he saw as being threatened by the historicist approach.

Several philosophers presented broadly similar counter arguments presenting history and philosophy of science as being 'inextricably linked' (McMullin, 1976; Burian, 1977; Shapere, 1984). These contributions all focused on the dynamics of theory change, in particular theory evaluation, to illustrate the case that in order to adequately understand these processes philosophers could not but look to history. They presented material illustrating serious difficulties with giving logical and universal accounts of theory evaluation. Instead they highlighted, how the

historical context, for instance 'the background knowledge against which these developments took place' (Burian, 1977 p2), needed to be taken into account in order to understand these decision making processes. This cluster of responses clearly proposes that history can make a contribution because it examines the temporal unfolding of events, of the 'before and after' (McMullin, 1976 p593). Schikore highlights that these responses embody an approach which sees philosophy as an attempt to understand science as it is, rather than aim at building a normative account, and what's more that 'understanding something involves understanding how it came about.' (Schickore, 2011 p462). She proposes that they represent a Anglo-American parallel to the rich continental approach of historical epistemology (Rheinberger, 1997b)[1].

In the late 1970's and 1980's there were more widespread changes in the focus of philosophy of science which led to further repositioning of the role of history with respect to philosophical analysis. As I highlighted above, the resistance to the use of historical cases in philosophy of science related to normative ambitions. Logical empiricists sought to uncover a general logic and rationale characterising the scientific enquiry process, establishing a clear account of what constituted science. Part of this agenda involved a demarcation between the context of discovery and the context of justification. The context of discovery was regarded as a precursor to justification and something which involved subjective and contingent decisions and actions. The process of justification, of constructing theories and evaluating theory and evidence, was regarded as the location of the rational scientific process and as such was the focus of philosophy of science. As we saw above, this emphasis on theory evaluation was initially also the primary target of philosophers advocating an inextricable relationship between the history and philosophy of science. Philosophers began to challenge the distinction between the modes of working and thinking involved in discovery and justification, and their clear cut temporal ordering, previously assumed to characterise scientific research (Nickles, 1990). The attention of philosophers expanded from theory to the whole range of practices scientists engaged in and they began to embrace the disunity of scientific theory and practice

[1] Historical epistemology perhaps has a more extended history of paying attention to the non-theoretical aspects of scientific research (Méthot, 2013).

(Darden and Maull, 1977; Dupré, 1995). Prominent work was carried out on the cognitive aspects of scientific research (Simon, 1977), the role of experiments (Hacking, 1983), the importance of scientific models (Cartwright, 1983), scientific instruments (Galison, 1997), and the wider social and cultural context's in which scientists operated (Longino, 1990; Knorr Cetina, 1999). Philosophy of science began to be influenced by studies of science originating from a wider range of disciplines including; anthropology, sociology, and cognitive science (Simon, 1962; Foucault, 1973; Latour and Woolgar, 1986). Therefore, the prominent role for historical case studies began to be displaced as a range of disciplinary approaches were embraced by philosophers for integrating accounts of actual scientific practice into their work (Wylie, 1994).

This widening of the diversity of processes of constructing case studies and philosophical contributions led to a shift in methodological concerns. A wide community now accepted that case studies of actual scientific practice had an integral role in philosophical analysis. The major methodological question that arose was how? What was the relationship between a case study and the philosophical claims it was associated with? Schikore identifies a publication by Larry Laudan in 1989 as heralding a shift from what she calls a historicist to a confrontationist approach to integrated history and philosophy of science. Laudan argued for the 'historical evaluation of philosophical claims' (Laudan, 1989 p11). The approach mirrored a scientific model in which philosophers construct theories about science and then use historical cases as evidence for assessing their accuracy. He implied a clear division of labour between the tasks of philosophical analysis and the production of historical accounts of science.

In 2001 Joseph Pitt wrote a critique of this approach highlighting 'The Dilemma of Case studies' (Pitt, 2001). He argued that if you start with a philosophical point and then use a case study to evaluate it, the philosophical point will inevitably effect the selection and interpretation of the case study, meaning that any conclusions drawn from this comparison are invalid. However, if you start from one, or any number of historical case studies, which involve attention to the contextual particularity of each case, then these cannot be used to justify making general philosophical claims (See also Galison, 2008 p119-122). This is because science is such a rich and diverse enterprise that it is very difficult to justify making

generalisations from a limited number of case studies to all scientific research. Friedrich Steinle and Richard Burian (2002) suggested that this tension between the particular and general has led an overall divergence, rather than integration, between historical and philosophical accounts of science since initial interest in their relationship in the 1970's and 1980's. In an earlier article Burian (2001) proposed a solution to Pitt's dilemma. He suggested that the dilemma only arises if you assume that there is a general universal scientific method. He see what he calls the 'top down approach', starting with general philosophical theories and using historical case studies to evaluate them, as always being based on this assumption. However, the 'bottom up approach', working from the particularities of case studies, does not encounter the dilemma Pitt proposes if you appreciate that there is diversity in scientific practices. Burian uses the case of exploratory experimentation to make the point that case studies can make a useful contribution to philosophy by highlighting aspects of science not yet captured in philosophical analysis. If philosophers are not aiming to make general assertions about science then particular case studies play a role in that they 'yield local, or better *regional* standards.' (Burian, 2001 p400).

Hasok Chang's (2011) account of 'epistemic iterativity', provides another perspective on the relationship between history and philosophy which moves beyond the linear relationship's at stake in top down and bottom up approaches. My account of the methodology used in this thesis (See section 2.1.2) illustrates a productive application of Chang's position. Chang makes a very useful distinction – instead of thinking about the relationship between history and philosophy as that between the particular and the general, we should think of it as the relationship between the concrete and the abstract. He is indicating that even the description of a single concrete historical episode of science requires the use of abstract philosophical concepts, such as measurement and explanation, even if these concepts are only abstracting what these terms mean in relation to the specific concrete instance. Like Burian, he points to the creative potential of case studies. They can bring to light things which philosophers of science have yet to appreciate, and help philosophers form abstract accounts of them. At the same time, Chang emphasises the importance of moving both ways between accounts of the concrete and the abstract. An iterative relationship between the two is required to refine the account of the concrete instance and its abstraction into philosophical concepts. Additionally,

Chang gives a role to the application of these developing abstractions to novel case studies in order to facilitate further iterative development. This approach, which Chang terms epistemic iteration, requires that the historical and philosophical dimensions of research are truly integrated, they cannot happen separately from each other, and must be carried out by the same person, or through close collaboration between a group of people.

The debates about the epistemic role of case studies in integrative history and philosophy of science have produced valuable insights. However, these insights pertain to the use of case studies in philosophy of science in general, and do not address the epistemic contribution that historical case studies in particular can make to philosophy of science. The debate about the role of case studies has attracted contributions from researchers operating across the field of interdisciplinary science studies (E.g. Morgan, 2012). Given this, Schickore and Arabatzis recently argued that researchers practicing integrated history and philosophy of science need to address a new set of methodological questions:

> "Historical study becomes just one option among many other empirical approaches to science, and given the pressing problems current science is facing, one might think not a particularly relevant one. Given this situation, the main challenge for scholars of HPS is not how to combine "general" philosophical theses and "particular" historical cases. Scholars of HPS are under increased pressure to legitimize the historical perspective. They need to show what, exactly, historical study can contribute to the understanding of current science, and how the historical perspective may aid and augment philosophical reflection." (Arabatzis and Schickore, 2012 p399)

It is this new variety of questions which this thesis makes a contribution to. Schickore proposes her own answer – history and philosophy of science needs to move away from the confrontationist approach and reinvigorate a historicist perspective which emphasises that historical case studies involve an appreciation of the temporal dimensions through which things come into being (Schickore, 2011).

Hans-Jörg Rheinberger's work is exemplary of the historicist approach to philosophy of science. His historical epistemology of the development of an experimental system for synthesising proteins during the middle of the twentieth century is a rich analysis of the complex temporal dynamics of scientific research (Rheinberger, 1997b; Rheinberger, 1997a). Rheinberger's epistemology involves

following the details of scientific practices. He identifies experimental systems, rather than theories, as being the attractor around which research coalesces. Productive experimental systems consist of technical objects which provide a coherent structure within which research can take place. Importantly they provide a clear background around which the elusive epistemic objects of research can start to be discerned and eventually come to constitute the technical facets of research programs. Rheinberger notes that in order to provide a means of investigating complex biological systems experimental systems must involve a simplification of the system. However, this means that the internal constitution of the experimental system is not enough to understand the historical dynamics of scientific practices and the emergence of epistemic things. The history of experimental systems needs to be understood in relation to the wider epistemic landscape in which they are located:

> 'It is the network of surrounding experimental systems that makes each of its elements take on its epistemic value. If ontic complexity has to be reduced in order to make experimental research possible, this very complexity is epistemically retained in the rich context of an experimental landscape.' (Rheinberger, 1997a p274)

This point will become particularly relevant in my discussion of the role of assumptions about data in chapter seven. Rheinberger (1997b) highlights three processes resulting in significant epistemic transformations. The first, conjunctures, focuses more on the unfolding of the internal dynamics of the experimental system in giving rise to the unexpected and allowing research objects to be recognised and formalised. The second, hybridisation, refers to the formation of relationships between the experimental system and aspects of the landscape in which it is operating to produce new research arrangements. Finally bifurcations, consist of experimental systems splitting to produce different research contexts. Rheinberger highlights that these often occur when the degree of internal complexity of the experimental system requires specialisation. In the next section I will introduce the particular temporal dynamics that this thesis examines. I also continue to expand on this topic in section 2.3 where I examine existing philosophical analysis based on the history of biochemistry.

Recognising the plurality of case study approaches leads to other interesting methodological questions. For instance, can historical perspectives intersect with other case study approaches, and how (Nersessian, 2005; Leonelli, 2010a)?

Integrated history and philosophy of science has embraced a pluralistic stance on questions about the methods of historical case study construction and the ways that these case studies can contribute to philosophy (Arabatzis and Schickore, 2012). Research has not led to the development of heavily prescriptive methodologies defining, for instance, a strict relationship between the parameters of the historical case study and the philosophical questions which it is best suited to contribute to. However, pluralism does not negate the importance of methodological reflection. It is important to ask questions about how the process of historical case study construction is related to the philosophical outcome of analysis. Frederic Holmes book 'Investigative Pathways' provides a reflection on how the temporal dimensions, both length and resolution, of historical case studies of biochemical research affects the analysis they are situated in (Holmes, 2004). Given the importance of methodological choices for contextualising the outcomes of research the next section involves a somewhat more personal account of the methodology used to generate the analysis in this thesis.

2.1.2 My Method

The broad area of research for this PhD was established as an historical and philosophical analysis of the role of mathematical models in biochemical systems biology, specifically models of metabolism. It was motivated by the current climate of philosophical interest in mathematical modelling in the context of data intensive biology. To begin with, choices needed to be made about the specific case studies which came to constitute the historical backbone of the thesis. Initially I started out by examining constraint based optimisation, the focus of the sixth chapter of the thesis. This choice came about through an earlier interest in evolutionary systems biology, in particular the work of Andreas Wagner (2005b; 2011) who's theory of robustness and evolvability relies heavily on analysis carried out using the constraint based optimisation technique. Constraint based optimisation was a widely used modelling approach in data intensive systems biology and this contributed to my choosing it as a starting point to trace the history of mathematical modelling of metabolism. Looking into the history of constraint based optimisation led me back to metabolic control analysis, the focus of the third chapter. This decision was based on the prominent role that Bernhard Palsson, the major proponent of constraint based

optimisation, gave to metabolic control analysis in his own history of systems biology (Westerhoff and Palsson, 2004). Working on this history drew my attention to the heated debates between researchers involved in metabolic control analysis, and biochemical systems theory – the subject of the fifth chapter of this thesis. My growing familiarity with the case studies led me to be interested in examining the significant constraints which the availability of different types of data seemed to impose on mathematical modelling and the role of assumptions about temporal organisation in addressing them. This gave rise to the fourth chapter on temporal decomposition.

Once a case study, or case studies, had been identified for a chapter the next set of decisions to be made were those about the parameters of the case study. My thesis was to be based on research articles, reviews, and commentaries in academic publications. This decision was not made on the basis of historical or philosophical considerations, but was largely driven by practical constraints like time scale and a lack of existing expertise in archival or field research. It should be recognised that certain important limitations arose out of these methodological choices. As the chapters largely analyse historical material which has not yet been the subject of historical and philosophical investigation their contribution cannot be directly compared to other interpretations. Additionally, within the scientific publication record there exists minimal commentary on the historical events analysed in this thesis by researchers from other areas. This is partly because many of the cases are from small scale research projects which did not receive the attention of the wider community beyond researchers who were directly involved. Subsequently the analysis is perhaps overly reliant on the auto-historiography of the scientists directly involved. This limitation could be addressed by including the analysis of sources beyond the academic publication record, for instance oral history.

Within the remit of working with publications, I had to determine the parameters, including time scale, breadth of publications examined, depth of attention paid to different aspects of research. The material itself was often the main influence suggesting fairly obvious demarcations. The parameters of the case studies could often follow contours derived from the publication record. For instance, for chapter five, the publications marking the beginning and end of the debates between metabolic control analysts and biochemical systems theorists could easily

be identified, and it was possible to exhaustively read, even if I didn't comment on, all the directly relevant material (Kacser and Porteous, 1987b; Savageau, 1992). As most of the case studies examine the early stages of mathematical model development, involving very small groups of researchers, this made it easy to be fairly comprehensive in my engagement with directly relevant primary literature. When large research communities coalesced around mathematical models, for instance around the mathematical reconstructions of genome scale metabolic networks examined in chapter six, I became more dependent on scientific review articles to give me an overview of the field. As this thesis is based around a series of interrelated case studies, even though the individual case studies could be considered rather narrow, the chapters provide a broader contextualisation for each other. For instance, chapter fives examination of early problems in the relationships between experimenters and modellers contextualises chapter six which examines the changes in these relationships in the context of data intensive biology.

My choice of working at this relatively fine grained level with only a handful of specific case studies reflected my comfort zone. At this level I felt I had the depth awareness, and feeling for the material, from which I could feel confident in the historical and philosophical assertions I wished to make. I was aware that much literature in the area of philosophy of systems biology appeared to be covering more ground than my work (Boogerd et al., 2007). Initially I was concerned by this, was it a reflection of my lack of knowledge about recent scientific developments? Eventually this tendency to stick very closely to the details of the historical movements became one of the strong points of the thesis. A comparison with work by James Griesemer will help me explain why.

Griesemer (2007) develops a processual perspective on the historical relationship between embryology and classical genetics to make a methodological point about historical and philosophical narratives of scientific episodes. Griesemer argues that narratives which follow the intricacies of the development of scientific practices, rather than a series of theoretical developments, lead researchers to pay attention not just to research configurations occupying the 'limelight' but also those which exemplify periods of stasis or decline. The pattern of the historical relationship he examines between embryology and classical genetics overlaps somewhat with

the relationship this thesis examines between biochemistry, or more specifically biochemical mathematical models, and the wider context of molecular biology:

> 'Whilst it may be true that the fortunes of classical genetics rose, while those of descriptive embryology fell, the continued practice of scientific styles out of fashion requires historical investigation if we are to understand the emergence, problems, and prospects, and historical continuity of hybrid intersectional fields such as evo-devo at the end of the twentieth century.' (Griesemer, 2007 p418)

Griesemer is examining the processes, involving both internal and external relations, which kept a minimal level of embryological work ongoing between 1940 and 1980 and provided the conditions enabling its reintegration with developmental biology to take off as a major research program in the subsequent period.

This thesis examines the practice of building mathematical models of metabolism which went on in the background to more high profile work in experimental biochemistry and molecular genetics. It looks at the factors which motivated this research program as well as those that maintained its small scale. Chapter five looks at some of the factors which inhibited growth in research in early biochemical systems biology. Following this, chapter six examines how the use of one of these models exploded in the context of changes in the wider landscape of biological data in which it was situated. Picking up on Schikore's challenge regarding the value of a historicist approach, chapters three and six in particular highlight how important a historical perspective is for perceiving the transience in relationships between different aspects of research and the reflection of this in our philosophical abstractions regarding scientific research. The point is that a methodological focus on the fine grained details of the case studies allowed me to investigate these aspects of research dynamics which remain hidden to coarser perspectives which are more likely to only pick up on high profile events. I reflect on this point in section 2.3 of this chapter where I situate my research amidst existing philosophical and historical work on biochemistry and contemporary systems biology.

After an initial draft of the historical case study was constructed, the next step was to think about the philosophical framework in which it would be situated. My ability in working with scientific and philosophical material affected this ordering of attention to history and philosophy. My undergraduate degree in 'Human sciences' was based in the biosciences department at Sussex university, and my

Masters in the ‘History and Philosophy of Biology’ at the University of Exeter in the EGENIS research centre. When I started my PhD I still felt more at home in scientific journals and less sure on my feet in literature in philosophy of science. It was the historical case studies which initially provided the threads connecting the different chapters together rather than particular philosophical themes.

Despite starting out with such a heavy emphasis on the historical narrative, as soon as I started to develop the philosophical analysis an iterative relationship was established between this and the historical case study. This led to a process along the lines of Chang’s (2011) description of epistemic iterativity involving an ongoing movement back and forth between the philosophical and historical aspects. Chang’s distinction between the concrete and the abstract is particular useful for thinking about how the historical cases in this thesis were constructed and presented. The case studies are highly conceptualised; the concrete historical narratives are explicitly developed in conjunction with philosophical abstractions. They should not be interpreted and evaluated as aiming to provide comprehensive historical accounts, from one perspective or another, of the mathematical models these philosophical contributions are developed around.

The choice of a specific case study was largely motivated by the fact that my attention had been drawn to something potentially interesting. Usually interesting meant that the ‘something’ seemed to be playing an important role in the development of the mathematical model. Not that the ‘something’ was necessarily noteworthy philosophically. There was always an element of risk to these choices. Whether that something would transpire into a clearer thing from which I would be able to extract philosophical relevance only became apparent in the later stages of drafting each chapter. Rarely was it immediately obvious what was philosophically interesting about a particular historical case. I experienced my lack of sure footing in philosophy of science as contributing to my difficulty in immediately making these connections, which with hindsight always somehow seemed obvious. I often experimented with several different philosophical framings before making a final choice. For instance, chapter five was initially framed in relation to the debate about the epistemic status of the results of experiments carried out in laboratories and experiments carried out with mathematical models (Morgan, 2003), and ended up as

an examination of social interactions between mathematical modellers and experimental biochemists.

During this phase of research the chapters were always re-drafted multiple times. The philosophical framing of the case study in the introduction and analysis was frequently reworked. My decisions about philosophical analysis would feedback into my account of the historical case. As the pertinent philosophical issue at stake became more apparent I would often go back and not only re-write the historical case study to reflect this focus, but also return to the primary historical literature in order to assess it through this new lens. Often this would involve keeping the timescale and breadth of the case study fairly constant, but altering my focus on the relevant features within the same demarcation of historical material, for example shifting my attention from the results section to the methods section of primary research articles. In turn, re-working the historical material would often feedback and alter the focus and outcome of its philosophical analysis. Gradually through these iterations an integrated historical and philosophical perspective on a case study would emerge which was stable enough to put to one side and move on to the next chapter. I did not explicitly think about a broader philosophical theme which would run through the thesis whilst I was writing each chapter. Asides from connections between the historical material used, initially I treated them as distinct pieces of philosophical work. Drawing out the dominant philosophical themes that ran throughout the thesis, outlined in the following section, only occurred as a result of active reflection once a stable draft of each chapter had been established.

2.2 Philosophical issues: Researchers assumptions, mathematical modelling and molecular biology

2.2.1 Assumptions and mathematical modelling

This section reviews philosophical work on assumptions in mathematical modelling across the sciences situating the contributions of this thesis in relation to this area of enquiry. The Models as Mediators account raised the profile of the role of assumptions in model building in philosophy of science. Margaret Morrison and Mary Morgan (1999b) argued that models were not just versions of theory or versions of

data, instead they should be regarded as an autonomous instrument which could be used in various ways to mediate between the theoretical and empirical dimensions of research. Part of their argument rested on an analysis of model construction. They suggested that the independent instrumental capacity of models arose because multiple diverse components, alongside theory and data, were used during model building. These diverse components could include assumptions. The role of idealising assumptions in model building in general has generated extensive work in philosophy of science (Suárez, 2008). Morgan and Knuttila, in an overview of mathematical modelling in economics, point out that many different meanings of idealising assumptions have been employed and provide a useful general definition of idealisation 'as involving processes of generalising, simplifying, abstracting, and isolating, following technical, substantive or conceptual aims or requirements.' (Morgan and Knuuttila, 2012 p51). Much of the discussion in philosophy of science has focused on the epistemic status of versions of models given these idealisations, their relationship to the systems they represent, and their role in knowledge production (Morgan, 1988; Boumans, 1999; Morgan, 2005; Winsberg, 2006; Krohs, 2008). In the case studies examined in this thesis researchers assumptions primarily play an important role in model development by influencing the construction of models, rather than the epistemic evaluation of completed versions of models. In the rest of this section I focus on what philosophers have said in relation to assumptions and model construction, although the processes of construction and evaluation are often intertwined (Boumans, 1999).

Michael Weisberg's analysis of the philosophical literature classifies three different kinds of idealisation, Galilean idealisation, minimalist idealisation, and multiple-model idealisation. I will use the first two of these to frame a distinction between two different types of assumption, ontological and idealising, used in the construction of mathematical models (Weisberg, 2007). Firstly Galilean idealisation, the practice of reducing the complexity of mathematical models in order to make them mathematically tractable; Secondly minimalist idealisation, the practice of only including factors thought to be 'causally relevant' in the model[2]. Throughout the thesis I will refer to both of these as idealising assumptions because they are used

[2] The third, multiple-model idealisation, is not a strategy which is directly relevant for the philosophical analysis in this thesis. It involves the construction of multiple models using different assumptions and method's in order to mitigate the impact of idealising assumptions through triangulation.

strategically to facilitate the construction of a working version of a model. However, what Weisberg describes as minimalist idealisations are based on researchers ontological assumptions, often established knowledge claims, about what is causally relevant. By ontological I am referring to assumptions about actual things and actual processes such as metabolites and metabolic reactions rather than assumptions about the fundamental constitution of reality[3]. I will use ontological assumptions as a distinct category as the research in this thesis highlights a variety of ways in which ontological assumptions can impact on the development of mathematical models alongside facilitating minimalist idealisations. Several philosophers have commented on the role of ontological, or what are otherwise referred to as theoretical assumptions, and idealising assumption in the construction of mathematical models. I class theoretical assumptions as ontological assumptions because they relate to assumptions about the specific objects, relations, and processes constituting metabolic systems which are not subject to confirmation.

Rasmus Winther's account of mathematical model building in evolutionary genetics provides an analysis of the importance that ontological assumptions play in mathematical modelling in biology. Winther provides the following definition:

> "Theoretical assumptions of various sorts pick out what are interpreted as the important material structures and processes of the system under study. These assumptions concern views both about basic ontology, including basic structures and basic causes, and about legitimate ways of abstracting; they are themselves the product of previous theoretical and empirical activity." (Winther, 2006a p221)

He argues that the differences between two models of parasite-mediated cytoplasmic incompatibility based on the perspectives of Fisher and Wright stem primarily from differences in the theoretical assumptions employed in model building rather than differences in the mathematical methods or data used. Part of his argument involves pointing out that theoretical assumptions play a crucial role in determining the 'content and form' of the model, which components and variables should be included and how they should be described.

[3] I recognise that some philosophers refer to such assumptions as ontic. Thank you to my examiners for pointing this out.

Marcel Boumans (1999; 2005) uses the analogy of baking a cake without a recipe to describe the process of building mathematical model's using examples of business cycle models in economics. He suggests that a modeller starts out with a range of diverse ingredients including things such as, empirical data, theoretical notions, stylised facts, mathematical techniques, and a notion of what they want to achieve. They must then find a way of putting these ingredients together, or integrating them, so as to construct an appropriate model. Boumans uses the analogy of a cake to convey the fact that it is impossible to distinguish the original ingredients once the model is constructed. He proposes that integration of the ingredients takes place through a process of mathematical moulding. The mathematical form of the model functions as an 'impartial mediator' (Boumans, 2005 p4) around which 'the disparate ingredients can be harmonised and homogenised into one effective model.' (Boumans, 2005 p13). Moulding the diverse ingredients together around the chosen mathematical formulation involves the use of idealising assumptions, 'elements of convenience or fiction' (Boumans, 2005 p14), in order to produce a seamless working model. In Boumans account it is the selected mathematical form, rather than ontological or idealising assumptions, which plays the most central role in shaping model building.

Others have emphasised the role of ontological and idealising assumptions in mathematical model building. Richard Levins (1966) paper, 'The strategy of model building in population biology' is foundational for philosophical discussion of idealisation of mathematical models of complex biological systems (Weisberg, 2006; Wimsatt, 2007b). Levin's argues that models based on the brute force approach which tries to establish 'a faithful, one-to-one reflection of this complexity' (Levins, 1966 p421) often fail for three reasons. Firstly, limitations in the availability and collection of data for all the different parameters, secondly, intractability of the mathematical equations, thirdly, difficulty in meaningfully interpreting the results. (The first two of these reasons are recurrent themes in the case studies examined in this thesis). Levins then describes how simplified models are built with researchers selecting the most relevant factors, a process partly based on researchers ontological assumptions, according to the particular phenomenon they are interested in. For Levins, during this decision making process researchers must make a

compromise between achieving the desirable qualities of generality, realism, and precision.

Eric Winsberg (2001) makes an argument along similar lines in his analysis of the role that theoretical and idealising assumptions play in the construction of simulations of physical systems using an example from astronomy of a mathematical simulation of the properties of red giant stars. He describes the process of constructing a simulation as starting from ontological, or what he calls theoretical, assumptions about the components, their properties, and interactions. These assumptions are used to build an initial mathematical model of the system using partial differential equations to describe the complex non-linear dynamics that the system exhibits. However, these initial models are often too complex. The system of coupled differential equations is intractable, there is insufficient computational power to find an analytic solution within a reasonable timeframe. Winsberg describes how in order to produce a tractable mathematical model 'tricks of the trade' must be employed to convert infinite differential equations into finite difference equations. These involve making 'simplifying assumptions, removal of degrees of freedom, and even substitution of simpler empirical relationship for more complex but also more theoretically founded laws.' (Winsberg, 2001 p445; See also Gramelsberger, 2011). Whilst making these idealising assumptions modellers are striving to retain the pertinent features of the system for simulating the phenomenon they are interested in. An interesting feature of Winsberg's account is that he regards simulations of complex physical systems as something which is carried out when there is a lack of empirical data, he regards model construction as a process working downwards from theory. He emphasises that empirical data plays no significant role in the construction or evaluation of mathematical simulations. In comparison to Levin's account, assumptions employed during model building do not intersect in any significant way with data resources.

Existing philosophical accounts have emphasised the role of ontological assumptions in facilitating researchers choices about what to include in their mathematical representation of the system. Idealising assumptions have primarily been examined in their role of addressing constraints encountered during model building. Minimal attention has been paid to the role of idealising assumptions in addressing constraints encountered due to data availability. The emphasis has

largely been on using idealising assumptions to address technical constraints imposed by the mathematical intractability of models involving multiple differential equations. Philosophers have observed that idealising assumptions and can be applied according to 'tricks of the trade' (Winsberg, 2001), or made on the basis of ontological assumptions which guide 'legitimate ways of abstracting' (Winther, 2006a). Building on this work, this thesis expands the role that ontological assumptions play in the development of mathematical models. For instance, chapter six looks at the role of ontological assumptions in the evaluation of data resources used during model building. Work in chapters four and six pays attention to how idealising strategies are used to address constraints stemming from the availability of data. Chapter four also extends the analysis of the use of idealising assumptions to address problems stemming from mathematical intractability by focusing on the role of ontological assumptions about the temporal organisation of metabolic systems in facilitating these idealisations.

2.2.2 Assumptions and research strategies in biology

This thesis is about the role of assumptions in biological mathematical models, those of metabolic systems. I will now turn to a particular group of assumptions that are prevalent in philosophical discussions of biological research in general, those which relate to the relationship between parts and wholes, and point out how this thesis contributes to their discussion. These types of assumptions play a significant role in different forms of biological reductionism; ontological, methodological, and epistemic (Sarkar, 1992; Brigandt and Love, 2012). They also underlie several major areas of research in philosophy of biology including; emergence, robustness, and mechanistic explanation. This section focuses on assumptions about the organisation of parts and wholes because they have been particularly significant in discussions of cellular and biochemical complexity. Significant philosophical examinations have been carried out of other common ontological and idealising biological assumptions, for instance that natural selection is an optimising process (e.g. Resnik, 1996), which I will not examine here. After briefly discussing epistemic reductionism, I will focus on a specific body of philosophical literature which illustrates the close relationship between reductionist ontological assumptions about the relationships between parts and wholes, and

idealising assumptions, employed during the construction of representations of biological systems. Following from this, I discuss work on the issues that context-dependency raises for biological reductionism. I then go on to examine the recent debate about the relationship between research strategies which involve part-whole assumptions and mathematical models of biological systems, and the role of reductionism in contemporary biochemical systems biology.

Initially the philosophical debate about biological reductionism focused on theory reduction, in particular the debate stemming from Kenneth Schaffner's (1969) work about whether theory in classical genetics could be reduced to the theory of molecular genetics. The debate later moved towards explanatory reduction focusing on the reductionist reasoning biologists appealed to when providing explanations based on the assumption of a part-whole organisation (Brigandt and Love, 2012; See Kauffman, 1970 for an earlier articulation of this position). An extreme form of epistemic reductionism is the claim that explanations in molecular biology can be reduced to explanations in physics and chemistry (Keller, 2010). Marcel Weber recently argued for a strong form of physical reduction in molecular biology using the illustration of explanations of the transmission of nerve signals. He argues that whilst it is impossible to reduce higher level theory from neuroscience to theory from physics and chemistry, physical and chemical explanations play a major role in many explanations in molecular biology that needs to be accounted for. He gives the example of the use of chemical concentration gradients to explain the transmission of action potentials through neurons. Weber goes so far as to claim that:

> 'Experimental biologists must apply theories from physics and chemistry in order to provide explanations of biological phenomenon. The explanatory force is provided solely by the physicochemical theories – theories that describe how molecules interact and how they behave in bulk. There is no specifically *biological* explanatory import; all the explanatory concepts are physicochemical ones.' (Weber, 2004 p28)

Weber argues that biological explanations involve the application of physical and chemical theories in a specific biological context. Biological concepts only serve to describe the context in which those theories are being applied; they do none of the explanatory work. For Weber, biological context is descriptively relevant but irrelevant for explaining the properties of molecular interactions. All the case studies

in this thesis, and in particular chapters four and six, involve researchers applying assumptions of physico-chemical determination to some properties of metabolic systems but not others. This thesis does not evaluate these kinds of assumptions, but highlights when researchers make them, how they change over time, and their role in the development of mathematical models.

I will now examine philosophical perspectives on research in molecular biology and biochemistry which primarily focus on assumptions about the spatial hierarchical organisation of parts and wholes. I will discuss the aspects of this work which look at the production of representations of biological systems, rather than how these representations are connected to explanations to contextualise this focus of this thesis on the role of assumptions in the development of mathematical models. As I discussed in section 2.2.1, building these representations often involves making ontological assumptions about the system which facilitate choices about which aspects of the system need to be included in the representation. I will begin by reviewing Sahotra Sarkar's general account of the substantive assumptions underlying different types of reductionist explanations, and then present a trajectory of work, running through, Herbert Simon, William Wimsatt, and William Bechtel and Robert Richardson, which illustrates a very close relationship between ontological and idealising assumptions about the organisation of parts and wholes. Following from this I will discuss the current philosophical discussions related to this work which this thesis makes a contribution to.

Sarkar (1998) describes three substantive criteria, or assumptions, which form the basis of representations around which reductionist explanations are developed. Firstly, fundamentalism, the assumptions that a different realm is going to play a more fundamental role in understanding than the actual realm that is the focus of enquiry. Secondly, abstract hierarchy, the system is represented as a hierarchy arranged according to a criterion which is independent from the explanation of the system. A lower level of the hierarchy will play a fundamental role in explaining higher levels. Thirdly, spatial hierarchy, this results when the 'independent criteria' for determining the levels of the hierarchy is 'spatial containment'. The hierarchy consists of spatial parts and the wholes which they constitute. Sarkar argues that strong reductionism, which fulfils all of these criteria, is found 'in many explanations in molecular biology.' (Sarkar, 1998 p45). The generality of Sarkar's three criteria

leaves open the possibility that biological reductionism need not involve explanations in terms of physics and chemistry. In molecular biology, systems level behaviour is often explained at the fundamental level of its macromolecular constituents (Sarkar, 2003).

Herbert Simon (1962) argued that a broad class of complex systems, including social systems, biological systems, and physical systems, exhibit a hierarchical organisation. He defined a particular type of hierarchy, a nearly decomposable hierarchy, which could guide decisions about what needed to be included in the investigation of complex systems by facilitating simplifications of the system. Nearly decomposable hierarchies are those where intra-level interactions were stronger and more frequent than inter-level interactions, and interactions within components in the same level were stronger and more frequent than interactions between components in the same level. As the internal interactions of levels and components were more significant than their connections to other parts of the system Simon proposed that systems could be decomposed into distinct levels and components which could be studied in isolation from the rest of the system. Simons approach is often thought of as a reductionist research strategy as it promotes the investigation of systems in terms of their constituent parts, and the investigation of these components in isolation from the wider systems context. As the literature summarised below illustrates, Simons account of the organisation of complexity has largely been interpreted as a presentation of complex biological systems as a hierarchy of spatial parts, 'we find cells organised into tissues, tissues into organs, organs into systems.' (Simon, 1962 p469). In chapter four I offer a different perspective on Simons account of the nearly decomposable hierarchies observed in complex systems, highlighting the significant role also played by assumptions about temporal organisation (Simon, 1977).

William Wimsatt critically extended Simon's ideas. He emphasised that viewing complex biological systems as a decomposable hierarchy did not entail theoretical or ontological simplicity by specifying a single correct hierarchical decomposition of levels and components (Wimsatt, 1976; Wimsatt, 1972). Instead, Wimsatt emphasised that there will be numerous appropriate ways to decompose a system depending on the interests of the researcher and the state the system is in at the point of investigation. Near decomposability is compatible with theoretical and

ontological pluralism. In a later paper “Forms of Aggregativity” Wimsatt (1986) reinforces an interpretation of a nearly decomposable hierarchy which focuses on the organisation of parts and their spatial relationships. Wimsatt’s aggregative system is one where, 1) the behaviour of the system is an aggregate of the behaviour of its parts, 2) the behaviour of the system is robust to changes in its parts, and 3) there are minimal interactions between component parts.

Simon and Wimsatt’s work was a significant influence on William Bechtel and Robert Richardson’s (1993) work ‘Discovering Complexity’. In this work Bechtel and Richardson were interested in how scientists investigated complex biochemical and cellular systems and how these approaches to investigation affected the form of the explanations they gave. They proposed that near decomposability and aggregativity enabled systems to be decomposed into particular functions localised in component parts which could be studied in isolation from the rest of the system. They argued that this approach gave rise to mechanistic explanations commonly found in biology which ‘account for the behaviour of a system in terms of the functions performed by its parts and the interactions between these parts’ (Bechtel and Richardson, 1993 p17). In Bechtel and Richardson’s work there is a close connection between their ontological assumptions about the spatial and functional organisation of the parts of biochemical systems and idealising assumptions which facilitate the simplification and investigation of complex biochemical systems. Chapter four offers an extension of the decompositional strategy used to investigate biological complexity by examining assumptions about the temporal organisation of metabolic systems which researchers make during the construction of mathematical models.

The research in this thesis contributes to current philosophical discussion around assumptions about part-whole organisation in three ways which I now go on to discuss. Firstly, it feeds into analysis of the impact that assumptions about the context dependency of properties of component parts have on model development. Secondly, it is relevant for current debates about the relationship between part/whole assumptions and mathematical models. This debate has particular significance in light of the current growth in biochemical systems biology. Thirdly, in the next section (2.2.3), I contextualise the contribution this thesis makes to discussions on the

relationship between assumptions about the spatial organisation and the temporal organisation of biological systems.

A common feature of reductionist perspectives based on the analysis of a hierarchy of spatial parts is that the context of the part is deemed insignificant for understanding the behaviour of components and the contribution their role in biological systems. Several philosophers have raised significant problems with the assumption that the biological context of component parts is irrelevant for understanding their role in the particular biological wholes in which they are located (Gilbert and Sarkar, 2000; Laubichler and Wagner, 2001; Burian, 2004). John Dupré makes the case that the properties of components of biological systems are relational to the wider biological context in which they are situated (Dupré, 2009; See also Powell and Dupré, 2009). He points out that whilst upward causation from parts to wholes is widely accepted, the occurrence of downward causation from wholes to parts is treated with widespread scepticism. Using the example of moonlighting proteins - proteins whose functional capacities depend on features of its wider environment - Dupré argues that appeals to downward causation are becoming common in molecular biology. He claims that the properties of biological components derive from features of those components and how those features interact with the biological context in which they are located (See also Barnes and Dupré, 2008). If biological context is an important part of understanding the properties of component parts then this has serious implications for investigative and explanatory practices based on Simon's assumption that biological systems constitute nearly-decomposable hierarchies.

Assumptions about the relevance of biological context have significant implications for the methods used to investigate biological systems. In experimental biology, the problematic relationship between *in vivo* and *in vitro* experimental interventions and knowledge claims is a pertinent example (Rheinberger, 1997b; Strand, 1999). The rise of reductive explanations in molecular biology has been associated with the development of new experimental methods for isolating component parts and studying them in isolation from biochemical systems (Morange, 1998; Powell and Dupré, 2009). Chapters four and six highlight the particular significance that assumptions about the context dependency of component

properties have for the evaluation of data resources used to build mathematical models.

Recently there has been significant philosophical debate about the assumption of part-whole biological organisation, the forms of explanation which it elicits, and its relationship to mathematical models of biological systems. Rasmus Winther (2006b; 2011) has made a distinction between compositional biology and formal biology as two distinct styles of research involving different modes of explanation, modelling, and partitioning biological systems. According to Winther, compositional biology is based on the understanding and investigation of concrete parts and wholes their organisation and interactions using a variety of 'material, diagrammatic, and narrative models.' (Winther, 2006b p474). Winther includes molecular biology in this compositional category. For Winther formal biology involves the use of mathematical modelling and mathematical laws and centres around the understanding and investigation of phenomena in terms of the values and relationships between parameters and variables. Here, a clear distinction is made between methodological approaches to biological research which are based around assumptions of part-whole organisation and those involving mathematical modelling. Carl Craver has also argued for a distinction between mechanistic explanation in terms of the entities, activities, and their organisation and the perspective on biological systems provided by dynamic mathematical models. (Craver, 2006; Kaplan and Craver, 2011). Using examples from neuroscience, he claims that dynamical mathematical models usually only provide phenomenological descriptions of biological systems rather than explanations because they 'characterise the behaviour of systems, not in terms of their component parts but in terms of emergent or high level variables describing global states of the system.' (Kaplan and Craver, 2011 p604). Craver (2006) claims that whilst dynamical mathematical models may be able to reproduce systems level inputs and outputs he does not class them as explanatory because they do not provide a causal mechanical account of 'how actually' that systems level behaviour was generated through detailed descriptions of parts and their interactions. Again, there is a distinction been made between approaches based around assumptions of parts and wholes and those involving mathematical modelling.

Others have argued that mechanistic explanations based around the assumption that biological systems can be decomposed into parts and wholes are compatible with mathematical modelling approaches. William Bechtel and Adele Abrahamsen (2010) have offered an account of dynamic mechanistic explanation. They propose that mathematical modelling is used during the re-composition of the system in order to provide an explanation of the temporal orchestration of the component parts and operations comprising a biological mechanism. Ingo Brigandt (2013b) has recently made a similar argument using models from contemporary systems biology. He proposes that many modelling approaches in data intensive biochemical systems biology involve a very fine grained description of the component parts and operations included in the mathematical model. As such, they involve a combination of mechanistic explanation and mathematical explanation which goes beyond the description of a static mechanism. They provide an understanding of dynamic aspects such as how the levels of component parts change, and how parts and operations appear and disappear during the execution of a mechanism. The emphasis in this thesis is not on the type of explanations of metabolic systems which mathematical models do and don't provide. However, it does focus on the assumptions about biological systems which are important for a wide range of factors involved in the development of mathematical models. I have highlighted this current debate about the relationship between mechanistic explanations based on assumptions about the organisation of parts and wholes and mathematical models given the observation that methodological decisions about how to investigate complex and the form of subsequent explanations are heavily intertwined. Chapter four focuses on the role of assumptions about temporal organisation alongside those about spatial organisation is particularly relevant for the debate concerning the relationship between research involving assumptions about the organisation of parts and wholes and research involving dynamic mathematical models.

More recently, data intensive systems biology and the increased use of mathematical modellinghas raised fresh philosophical questions about the role of reductionism in biology. Systems biology is often associated with a reversal in the trend of reductionism in molecular biology and increase in the significance of the organismal context for biological research (Cornish-Bowden and Cárdenas, 2005;

Nicholson, 2014). Mathematical models in systems biology are frequently divided into two broad categories. 1) Bottom up models, those of relatively small biochemical pathways based on detailed data about the compositional and dynamics properties of individual components (Krohs, 2010; O'Malley and Dupré, 2005). These models are often presented as a continuation of reductionism in molecular biology. Whilst they play a role in understanding how the properties of biochemical systems emerge from interactions between their component parts, they are often seen as still being based upon the assumption that the properties of biochemical components can explain higher levels of biological organisation. 2) Top down models, those of large scale biochemical networks frequently based on large omics data sets about components and data about systems level dynamics (Krohs, 2010; O'Malley and Dupré, 2005)**.** These models are often used to pick out features of systems level organisation, for instance network topologies, which can account for the properties of the network. They are associated with a 'systems-theoretic' perspective which attempts to provide explanations at the level of properties of the biochemical system, not properties of its component parts. The historical case studies in this thesis, particular that in chapter six, contribute a more nuanced perspective on the relationship between particular approaches to mathematical modelling and different ontological assumptions about biological systems.

2.2.3 Temporal organisation

A common feature of both reductionist and anti-reductionist positions is that they are based on ontological and idealising assumptions of a spatial hierarchy between wholes and parts, regardless of their differences in opinion about the nature of the relationship between these different levels and the importance of biological context. The models of metabolism examined in this thesis involve assumptions about a spatial hierarchy. However, in many instances assumptions about the temporal organisation of metabolic systems play an equally significant role in model development. This issue is addressed directly in chapter four. The role of temporality in biological science has recently only attracted the attention of a handful of philosophers (Mitchell, 2003). I have decided to pay particularly close attention to this topic in this contextualising chapter in order to ensure familiarity of the reader with this work. Below I firstly review work focusing on the ontological dimensions and

highlight two different conceptions of temporal organisation which emerge in this literature. I refer to these as, 1. a causal temporal organisation, and 2. a processual temporal organisation. This is a thesis about the relationship between ontological assumptions and mathematical modelling as a particular methodological approach for analysing metabolic systems. At the end of this section I also review integrated historical and philosophical work which examines the relationship between perspectives on time and research methodologies in biochemistry and cell biology.

Andreas Hüttemann and Alan Love's recent collaborative work picks out a causal temporal organisation which they argue is distinct to the biological sciences (Hüttemann and Love, 2011; Love and Hüttemann, 2011). Hüttemann and Love are interested in further analysing the reductive reasoning practices used in biology. They suggest that notions of temporality in biology have been side-lined by the emphasis on the spatial relationships between parts and wholes which provide an a-temporal time slice and seem to make 'temporal relations appear beside the point' (Hüttemann and Love, 2011 p521). Additionally, they claim that explanations in physics and biology are characterised by different notions of temporality and that the legacy of physics in philosophy of science has led philosophers to be blinkered to the particularity of the role of time in biological explanations. In order to appreciate reasoning about time in reductive explanations they argue that we need to make a clear distinction between composition and causality in biological systems: '*composition*: Higher level entities are composed of, realised by, or nothing but lower level entities. … *causation*: some higher level entities are caused, bought about, or determined by lower level entities.' (Hüttemann and Love, 2011 p522). They regard part-whole explanations as often conflating composition and causation and neglecting the fact that causation is a process which unfolds in time.

Using the example of protein folding they exemplify how reductionist reasoning in biology involves an assumption of a determinate causal temporal relationships . For Hüttemann and Love:

> 'Temporality refers to the aspect of reductive explanation whereby a property of a whole at time *t** is explained in terms of properties of its parts at an earlier time *t*. If a *temporal* relation is one in which a property or a state at *t is related to another property at* t*, then a *causal* relation is one in which a property or a state at *t* determines or

> influences another property or state at *t**…' (Hüttemann and Love, 2011 p531)

In reductionist reasoning about spatial organisation, the constituents of lower levels of a spatial organisation determine the properties and composition of higher levels. Hüttermann and Love argue that in reductionist reasoning about temporality, the properties of the parts of a system at an earlier time are all that is needed to explain the properties of the parts of the system at a later time They argue that whilst reductionist reasoning in biology invokes a determinate causal temporal organisation reductionist reasoning in physics is a-temporal, even when physics presents dynamic models of systems. This is because, whilst these models involve the description of the temporal dynamics of their parts, they are still *a*-temporal because they involve an account of the higher level in terms of the properties of the parts at the same point in time, rather than invoking a temporal causal relationship to explain the properties of the whole in terms of the properties of the parts at an earlier time (Hüttemann, 2005; Love and Hüttemann, 2011).

William Bechtel and Adele Abrahamsen, in their recent work on dynamic mechanistic explanation, have also given an analysis of the importance of temporal causal order for biological explanations (Bechtel and Abrahamsen, 2010). Bechtel argues that constructing a mechanistic explanation first involves the reductive process of decomposing a system into its relevant constituent parts and operations. Once characterised and investigated in isolation researchers must then recompose these parts and operations in order to produce the mechanistic behaviour of the system. Bechtel emphasises that this process does not just involve the spatial organisation of the parts with respect to the whole. Researchers must investigate how the constituent operations are orchestrated in time and space so as to produce the mechanism:

> "A basic mechanistic explanation combines these parts and operations, qualitatively specifying the spatial organization of the parts and the temporal sequence of operations that are performed in succession until the termination conditions are satisfied, in this way producing the phenomenon of interest." (Bechtel, 2011 p537)

Recomposing a mechanism involves figuring out a causal temporal sequence for the constituent operations so as they produce the phenomena to be explained.

There are several differences between Bechtel's, and Hüttemann and Love's work. Bechtel is discussing the reasoning processes of researchers after they have carried out reductionist analysis, whereas Hüttemann and Love are analysing reductive reasoning processes. Bechtel is interested in a causal temporal sequence at one level of organisation, the constituent operations. Whereas Hüttemann and Love, are interested in a causal temporal sequence between two levels, the level of earlier and later time points, and the levels of the parts and the level of the whole. Hüttemann and Love are explicitly appealing to a causal temporal organisation in which properties at earlier times explain properties at later times, it is not so obvious whether Bechtel's work, although looking at causal temporal sequences of operations, can be construed as appealing to these kinds of determinate temporal relationships. Whilst Hüttemann and Love dismiss mathematical models in physics using differential equations to describe the dynamics of individual parts as *a*-temporal, Bechtel's work involves the analysis of mathematical models in circadian rhythm research which are based on coupled differential equations (See also Brigandt, 2013b). As I said at the beginning of this analysis, philosophers of biology have only recently started to pay attention to temporal dynamics. I am not going to argue whether the differences I have pointed out between these two accounts are significant. My point here is to group Bechtel, and Love and Hüttemann's, work together because they are both interested in the role of temporality in biology in relation to causal biological explanation involving the temporal order of interactions between parts, and between parts and wholes.

James Griesemer's (2001) and John Dupré's (2012) independent work on temporal organisation in relation to biology constitute more radical reworking's of philosophical assumptions about biological systems. Love and Hüttemann's, and Bechtel's, accounts assume a biological ontology involving biological objects, parts and wholes, which are distinct from the operations, or biological processes, which mediate interactions between these objects. As I outlined above they are interested in the causal temporal organisation of interactions between these parts and wholes and the relationship of this to biological explanations. Griesemer and Dupré's work involves developing a notion of a processual temporal organisation in which biological processes are presented as more ontologically fundamental than biological

wholes and their constituent parts. Griesemer and Dupré's work involves rethinking the ontological assumptions made in biology.

Griesemer (2001) develops an in-depth processual perspective on genetics and its relationship to development and reproduction. He argues that:

> 'The failure to analyse process is widespread in the units of selection literature and reflects the fact that analysts of the structuralist perspective have often been more interested in units that map neatly onto hierarchies of structural organisation than in tracing the ontological implications for evolution as a process.' (Griesemer, 2001 p243)

His central claim is that analysing genetics from the perspective that 'The fundamental entities of biology are processes rather than structure or functions.' (Griesemer, 2001 p240) allows for a significant explanatory shift. The structural perspective starts with a hierarchy of parts and wholes where cells and organisms are constituted by biochemical molecules including genetic material. This provides a rigid framework around which explanations of developmental and evolutionary processes must be built and encourages a tendency towards reductionist accounts. Griesemer claims that the alternative process orientated perspective he provides frees up the space of possible explanations. He argues for a position from which developmental and reproductive processes are more fundamental than the genetic processes to which they are related. As such, there is no straightforward overlap in the hierarchical relationships between biological objects and the hierarchical relationships between biological processes. Following from this, the perspective provides a different space for biological and philosophical reasoning about biological evolution and development.

Dupré's work involves a revision of the clear distinction between biological objects and biological processes through developing a notion of the temporal organisation of biological processes. He is interested in developing a processual ontology for biology examining 'the nature of living systems and the causality that they exhibit' (Dupré, 2013 p19). Dupré's interest in the temporal organisation of biological phenomena is related to an account of causation, but he presents some ideas about the processual nature of living systems as a basis for his account of causality. It is this first part of the account which I am interested in here. His analysis

of biological temporal organisation re-examines the fundamental nature of the constituents of biological systems.

Dupré argues that whether something is considered a thing or process depends on the timescale it is being examined from:

> '… the designation of an entity as a thing or a process is often best seen as relative to timescale. … An entity stabilized over a particular timescale may, relative to that timescale, be considered as thing.' (Dupré, 2013 p31)

He uses the example of a mountain as an illustration of something which when examined over a relatively short time scale may appear to be a very stable thing, but over a longer timescale appears to be undergoing significant changes. Similarly, over a relatively short timescale a cell might appear to be a stable object, but it changes significantly over the life span of an organism. Dupré argues that processes are more fundamental to living systems. It is possible to consider a biological phenomenon, such as a biochemical pathway, through an *a*temporal time slice where it appears to be composed fundamentally of stable things. However, a system which is frozen at a moment in time is not a living system. Fundamental to understanding biological systems is appreciating the multiplicity of different processes, operating at different timescales, which are involved in maintaining things which appear stable over particular timescales, and facilitating changes in those systems which are apparent at certain timescale resolutions. This processual organisation is related to a methodological hierarchy relating to the timescale the phenomenon is being examined at, with longer and possibly coarser grained timescale at the top, and shorter and possibly finer grained timescale at the bottom. My use of the term hierarchy, in relation to Dupré's work, does not refer to anything about the significance of processes operating at different timescales, or interactions between different levels of the hierarchy. The idea that different biological processes occur at different levels of a time hierarchy can be found in the work of several biologists and philosophers (Waddington and Kacser, 1957; Rheinberger, 2002; Burian, 2013). In chapter seven, where I discuss the philosophical implications of this thesis I will discuss the relationship between my analysis of researchers assumptions about temporal organisation carried out in chapter four and a focus on causal temporal organisation versus processual temporal organisation.

Hannah Landecker, Angela Creager, and James Griesemer's, separate bodies of work examine the impact of the development of new research tools and techniques on conceptions of temporality in genetics, biochemistry, and cellular biology. They clearly illustrate that different modes of investigating biological systems allow researchers to intervene and conceptualise biological temporal organisation in different ways. Landecker examines the development of experimental techniques in cellular biology, looking at the transition from histology to cell culturing and cell cloning (Landecker, 2002; Landecker, 2009; Landecker, 2013). She argues that in the early part of the twentieth century the move from *in vivo* histological samples, to *in vitro* cell culturing shifted the spatial and temporal dimensions of cellular biology. Histological samples presented static *a*temporal snapshots of cells situated within an organism, comparing samples from different stages of an organisms development enabled biologists to build a picture of a series of developmental stages. The ability to sustain cell cultures in the laboratory provided biologists with a means of watching development unfold in continuous time; however this was development of cells spatially relocated outside of the organism (Landecker, 2002). Creager addresses the impact of the use of radioactive isotopes in biochemistry after World War Two (Creager, 2013a; Creager, 2013b). Prior to this, metabolic pathways had been represented as static networks of metabolites and interactions. Radioisotopes enabled biochemists to study the temporality of the movement of molecules through metabolic networks. The nature of radioactive decay meant that this involved researchers 'tracking the appearance and disappearance of the spots over time.' (Creager, 2013b p84). Griesemer looks at the historical relationship of genetics and embryology from the perspective that following processes, as opposed to the structural relationship between components and operations, is a 'characteristic activity of science.' (Griesemer, 2007 p375; Griesemer, 2002). He looks at the 'marks' which geneticists and embryologists attach to hereditary and developmental process, how they use these marks to track these processes in real time and use them to construct representations of processes. He emphasises how changes in the methodologies for marking, tracking, and representing processes affect how these processes are understood and are association with splitting and diversification in the fields of early twentieth century genetics and embryology. In a similar vein to these three accounts this thesis examines how the use of mathematical tools in modelling metabolism is related to

researcher's assumptions about the temporal organisation of metabolic systems. In particular chapter four highlights how different sorts of mathematical equations are associated with the representation of features of the systems which are assumed to exhibit different properties as they change over time: e.g. the use of algebraic equations to represent discontinuous changes and the use of differential equations to represent continuous changes.

2.2.4 An expanded perspective on the role of assumptions in biological research

Philosophical discussion of the assumptions guiding research in molecular biology has focused on the relationship between parts and wholes in biological systems. These assumption have an ontological dimension, they are things which researchers take for granted. They also have an idealising dimension as they guide the strategies researchers use to investigate biological complexity. The philosophical discussion of assumptions in biological research is considerably narrower than the philosophical discussion of the role of assumptions in models, and mathematical models, across the sciences. In this broader context, philosophers have highlighted the importance idealising assumptions facilitating the mathematical representation of dynamic behaviour. Additionally, the models as mediators account argues for the important transformative role that diverse modelling ingredients, including assumptions can play in scientific research.

As I have already mentioned one of the significant ways that this thesis expands the philosophical discussion of the role of assumptions in biology is by paying attention to assumptions about the temporal organisation of biological systems. Alongside ontological and idealising assumptions, the case studies in this thesis also show the important role which assumptions about biological data resources, and modellers' research commitments, their assumptions about how research should be carried out, play in the development of mathematical models of metabolism. I keep on referring to researchers assumptions as playing a role, what I mean by this is that researchers' assumptions, despite often being in the background, have an impact on how research is carried out. The use of historical case studies examining the research practices of mathematical modelling communities allows me to draw out the ways in which assumptions affect research dynamics. As I have previously mentioned, one of the interesting things about

examining the history of these mathematical models is that it allows the observation of mathematical modelling, biochemical data, biochemical theory, and biochemical experiments coming into a relationship with each other. These relationships frequently expose assumptions which researchers may not have previously been aware of or questioned. Another significant contribution of this thesis is the analysis of the role that assumptions play in facilitating different dynamics during the period before biochemical systems biology became main stream, and how they contribute to its expansion in the context of data intensive systems biology. The historical perspective on the temporal dynamics of research means that this thesis also makes a significant contribution to a connected area of research, philosophical perspectives on the history of biochemical systems biology, and I go on to review relevant literature in this area in the following section.

2.3 Philosophical interactions with the history of biochemistry

The previous sections introducing the main philosophical themes from this thesis shows that historical cases from cell biology, genetics, and biochemistry have been used in the development and substantiation of philosophical perspectives on a variety of topics related to explanation, reductionism, ontology, and methodology. Philosophy of biology has historically focused on work in evolution and genetics. Much work in the history and philosophy of molecular biology has predominantly focused on molecular genetics (Morange, 1998; Sarkar, 2001). Although the title molecular biology is also frequently used synonymously with molecular genetics in scientific contexts, as far as it is the study of biological molecules it can be seen as encompassing a much wider range of disciplines including biochemistry. In light of data intensive systems biology, more and more philosophers are examining cases from what they term molecular or biochemical systems biology which clearly lie beyond the narrow definition of molecular biology as molecular genetics (O'Malley and Soyer, 2012; MacLeod and Nersessian, 2013c). The history of molecular biology in the wider sense has been noted to receive less philosophical attention than genetics, a topic on which numerous monographs have been published (Sarkar, 1998; Kay, 2000; Moss, 2004; Barnes and Dupré, 2008; Griffiths and Stotz, 2013). The recent flurry of philosophical work on molecular systems biology has prompted interest in the history of the multiple disciplines which now contribute to this highly

integrative line of work. Following Griesemer's methodological approach of paying attention to work going on in the background (Griesemer, 2007 see section 2.1.2), this thesis contributes a philosophical analysis of historical cases of mathematical models of metabolism which are now a widely used research tool in systems biology. In this section I firstly give an overview of existing work on the history and philosophy of biochemistry which contextualises some of the themes examined in this thesis. Secondly, I outline recent work on the history and philosophy of contemporary systems biology.

There are several historiographies of biochemistry by biochemists and historians written from a variety of different perspectives - biographic, institutional, instrumental, conceptual (For review's see Kohler, 1975; Morgan, 1990). Robert Kohler's (1973) own perspective gives an account of the emergence of biochemistry as a discipline at the beginning of the 19th century. Kohler emphasises the development of enzyme theory as playing a significant role of demarcating biochemistry as a discipline. The proposal of specific relationships between metabolites and enzymes mediating their transformation separated biochemistry from the earlier thesis of a uniform protoplasm which mediated biochemical changes. From the outset, a significant theme singling out biochemistry from other related disciplines was the fact that it examined '"dynamic biochemistry"... Physiological chemistry was the chemical statics, biochemistry the chemical dynamics of living systems" (Kohler, 1973 p183). As far as biochemistry was heavily equated with the study of metabolism, it examined changes in molecules mediated by enzymes. The emphasis on enzyme substrate specificity, rather than a homogenous protoplasm, intersected with important developments in experimental work involving the ability to isolate and study metabolic interactions *in vitro*. Frederic Holme's (1986; 1992) historical work focuses on the study of intermediary metabolism as the defining feature of biochemistry. Studying intermediary metabolism poses many difficulties; metabolic intermediaries are transient and rapidly metabolised through another step in a metabolic pathway. Research focusing on discerning these transient stages marked biochemical research out from previous work based on the assumption of uniform protoplasm. During the first half of the 20th century biochemists were primarily concerned with identifying the stages in now iconic biochemical pathways such as glycolysis. Kärin Nickelsen's historical and philosophical analysis of the

development of Otto Warburgs model of the mechanism of photosynthesis shows how these models were constructed using a 'building block' strategy combining aspects of existing knowledge and techniques (Nickelsen, 2007).

From the 1950's, after the ground breaking work of Watson and Crick (1953), molecular genetics took centre stage in scientific research and in the later retrospective philosophical gaze. Several philosophers and historians have looked at the turbulent relationship between biochemists and molecular geneticists in the 1950's and 1960's (De Chadarevian, 1996; Abir-Am, 1992; De Chadarevian, 2002). The sequencing of proteins by biochemists made a significant contribution to the work of unravelling the relationship between the genetic code and biological molecules. However, biochemists felt that their contributions to genetics were being down played and their disciplinary authority being displaced by the up and coming discipline of molecular genetics. Hans-Jörg Rheinberger's (1997b) work 'Towards a History of Epistemic Things' develops his concept of an experimental system through examining the trajectory from *in vitro* protein synthesis to the interpretation of the genetic code. Angela Creager and Jean-Paul Gaudillière (1996) have written a detailed historical account of the development of allosteric regulation in two laboratories in Paris and Berkley from 1959-1968. They pay particular attention to the relationship of allosteric regulation to biochemistry and molecular genetics. They conclude that molecular genetics reduced the 'jurisdiction of biochemists' (Creager and Gaudillière, 1996 p87) because it assumed that gene expression was prior to metabolic interactions. However, work on allosteric regulation 'helped to rescue a few biochemists from the shadow of molecular genetics' (Creager and Gaudillière, 1996 p89) as it readdressed the balance through giving allosteric proteins a significant role in regulating protein synthesis [4].

Philosophical work using cases from the history of biochemistry after the 1950's, which is not examining the relationship of biochemistry to molecular genetics, is relatively sparse. An exception is William Bechtel's analysis of historical cases from biochemistry in the 1950's and 1960's which plays an important role in his philosophical work on biological mechanisms. In 'Discovering cell mechanisms' Bechtel (2006) focuses on the relationship between cell biology and biochemistry in

[4] Michel Morange has recently written an account of the trajectory of allostery from its origin to the present day (Morange, 2012)

this period. He examines how understanding oxidative phosphorylation required integrating work in biochemistry and cellular biology in order to locate biochemical processes in particular organelles, the structure of which played a significant role in facilitating certain biochemical transformations. Roger Strand, (1999) in his article "Towards a useful philosophy of Biochemistry: Sketches and Examples", attempts to highlight the problems which biochemists regularly encounter in their work which are of philosophical interest. He focuses on the *in vivo / in vitro* problem, 'the problem of assessing the biological relevance of biochemical data.' (Strand, 1999 p237). This theme is taken up throughout this thesis, particularly in chapters four and six. The thesis contributes towards historical work on biochemistry in the period between the 1960's and data intensive system biology which is not directly related to molecular genetics.

Philosophers and biologists have contributed several broad historical overviews of contemporary systems biology. The scientists Hans Westerhoff and Bernhard Palsson (2004) describe systems biology as having two separate roots. The first is the more well-known root of molecular genetics which led to the high-throughput sequencing of genomes and the subsequent production of multiple other types of large omics data sets about molecular components. They identify the second as developing from non-equilibrium thermodynamics in the 1930's and leading to the development of mathematical models from the 1970's onwards which attempt to simulate the dynamics of interactions between multiple biochemical components interacting as systems. Systems biology resulted from merger of large biochemical data sets with these mathematical modelling practices. O'Malley and Dupré (2005) similarly describe a distinct two root history of systems biology which is reflected in two divergent approaches to mathematical modelling found in contemporary practices. On the one hand, pragmatic systems biologists, whilst focusing on interactions between components in systems, use bottom up modelling techniques working from detailed descriptions of individual components up to analysis of systems level behaviour. This focus on explaining the whole in terms of properties of the parts leads O'Malley and Dupré to claim that this is often a reductionist strategy which has much in common which the 'reductionist aspirations of genomics' (O'Malley and Dupré, 2005 p1270). On the other hand, systems theoretic approaches, use top down modelling approaches, starting from a

description of the behaviour of the system as a whole and attempt to understand this through systems principles. This approach is historically associated with the work of earlier systems theorists and cybernetics. Ulrich Krohs and Werner Callebaut (2007) provide a historical account based around three roots, biochemical pathway modelling, molecular genetics, and biological cybernetics. They describe systems biology as emerging out of the merger of omics data stemming from molecular genetics with pathway modelling and cybernetics. Pathway models are a bottom up approach to model building involving detailed kinetic descriptions of individual components. In Krohs and Callebaut's account, this approach is not heavily associated with the use of omics data, as omics provides relatively poor structural data not the rich kinetic detail these models require. Cybernetic modelling has led to top down modelling approaches which are based on minimal data about systems level dynamics, but after the availability of omics data also rely heavily on large structural data sets to build large scale reconstructions of metabolic networks. The fine grained case studies analysed in this thesis contribute a perspective on the history of biochemical mathematical models which is less clear cut and picks up on the blurred boundaries and transient relationships which exist between different modelling approaches, biochemical data sets, and assumptions about biological systems.

The use of mathematical modelling approaches to analyse large biological data sets is often seen as one of the distinguishing features of contemporary systems biology (Kitano, 2001). Philosophers have analysed the practices and epistemic implications of this style of research which integrates mathematical modelling and biochemical data from a number of different perspectives (O'Malley and Soyer, 2012; Brigandt, 2013b; Green, 2013; Leonelli, 2013; MacLeod and Nersessian, 2013c). There has been significant detailed philosophical and historical attention on the production, management, and use of large scale data sets including work on the history of whole genome sequence production and open-access data sharing infrastructures (Hilgartner, 1995; Bostanci, 2004; Strasser, 2008; Leonelli and Ankeny, 2011; Leonelli, 2012; Leonelli, Forthcoming). The histories of the mathematical techniques which are used to analyse and model this data have received less thorough philosophical consideration. Biochemistry has a lengthy history of using mathematical models in research, at least from Michaelis-Menten's

(1913) introduction of a mathematical models of reaction kinetics. I have not identified any historical or philosophical work on mathematical models of the kinetics of individual reactions, and historical and philosophical accounts of biochemistry, both before and after molecular genetics, have tended to focus on it as an experimental laboratory based science (Holmes, 1986; Strand, 1999; Bechtel, 2006). This lack of attention to the history of mathematical models of individual biochemical transformations is problematic for developing accounts of mathematical models of systems of interacting molecules which characterise contemporary systems biology. Sara Green and Olaf Wolkenhauer (2013) have recently examined the history of organising principles associated with the more systems theoretic approach to mathematical modelling in contemporary systems biology. The scientists involved in developing the mathematical models have contributed reflections on the histories of some of the models examined in this thesis (Westerhoff and Palsson, 2004; Fell, 2007). The historical case studies in this thesis make a contribution to this gap in the literature concerning the history of mathematical modelling of metabolic systems from the 1960's to the present day.

3 Autonomous mathematical models: Constructing theories of metabolic control

Abstract

This chapter considers how the relationship between mathematical models and theories in biology may change over time, on the basis of a historical analysis of the development of a mathematical model of metabolism, metabolic control analysis, and its relationship to theories of metabolic control. I argue that one can distinguish two ways of characterising the relationship between models and theories, depending on the stage of model and/or theory development that one is considering: partial independence and autonomy. Partial independence describes a model's relationship with existing theory, thus referring to relationships that have already been established between model and theory during model construction. By contrast, autonomy is a feature of relationships which may become established between model and theory in the future, and is expressed by a model's open ended role in constructing emerging theory. Idealising assumptions used during model construction play a crucial role in transforming the models relationship to ontological assumptions about metabolic control. These characteristics of partial independence and autonomy have often been conflated by existing philosophical accounts, partly because they can only be identified and analysed when adopting a historical perspective on scientific research. Adopting a clear distinction between partial independence and autonomy improves philosophical insight into the changing relationship between models and theories.

3.1 Introduction

During the 1970's an important aspect of biochemists' understanding of metabolic control was the concept of a rate limiting step – a single reaction in a metabolic pathway determining the pathways overall metabolic flux. In the same period an alternative theory began to emerge, one which suggested that metabolic flux was controlled through the interactions of multiple reactions in a metabolic system. The construction of this new theory was facilitated by the development of a mathematical model, metabolic control analysis. However, the initial development of this model was motivated by a need to clarify the concept of a rate limiting step. The

mathematical model was constructed in relation to one theory yet later led to the construction of a different theory. In the following, I analyse this historical case in order to elucidate the difference between two characterisations of the relationship between mathematical models and theories – partial independence and autonomy.

Margret Morrison and Mary Morgan (1999b) argue that models have an independent epistemic role in scientific research, and are instruments which mediate between its theoretical and empirical aspects. One key part of their argument is that mathematical models are not just versions of theories, instead they are partially independent from theories. They do not attempt to give a hard and fast distinction between models and theories, but point out that whereas "in some cases … theories consist of general principles which govern the behaviour of large groups of phenomena; models are usually more circumscribed…" (Morrison and Morgan, 1999b p12). Their argument for partial independence centres on looking at how these more circumscribed models are constructed. They situate their argument in relation to previous philosophical accounts which regarded models to be singularly influenced by theory, and saw models as constituting often simplified versions of theory in a variety of formats (Morrison and Morgan, 1999a). Morrison and Morgan regard the processes of simplification and approximation to be important aspects of model building. However, they point out that these processes are influenced by a large number of factors alongside theory, for instance mathematical tools, and data availability (see also Boumans, 1999). The importance and inclusion of these diverse factors during model building means that models are more than versions of theory. Instead they are partially independent from the theory which influenced their construction.

Nancy Cartwright's (1999b) contribution to the models as mediator's framework provides a useful basis for understanding partial independence in the specific case examined in the next section of this chapter. In *The Dappled World* (1999a) Cartwright argues that theories are comprised of abstract concepts and abstract relations between those concepts; she focuses her analysis on the concept of force in physics. Abstract concepts do not exist separately from their concrete, or particular, applications. In her account, the process of model building involves the use of bridge principles to concretise these abstract theoretical concepts. These provide the grounds for prediction and mediation between theory and the world by

connecting the theory to empirical content. Bridge principles contain more concrete concepts which often provide the basis for measurement. In her example, she shows how modelling force involves using more concrete concepts such as acceleration and mass. In line with Morrison and Morgan, Cartwright points out that, as multiple factors alongside theory influence this process of concretisation, models bear a partially independent relationship to the theory which influenced their construction. Cartwright's position provides a suitable framework for analysing the material in the next section for two reasons; firstly, it addresses instances when models are intended to strengthen theoretical inferences from experimental results; secondly, as mathematics is frequently an important aspect of concretisation it provides an account of why we should refer to the relevant equations as comprising mathematical models rather than mathematized theories.

A second key part of Morrison and Morgan's argument is that models play an autonomous role in research, and this is something which is facilitated by their partial independence. In fact,

> …if models are to play an autonomous role allowing them to mediate between our theories and the world, and allowing us to learn about one or the other, they *require* such partial independence. (Morrison and Morgan, 1999b p17)

In their account the two characterisations of the model theory relationship are playing quite distinct roles– partial independence is acquired during the process of model construction and this is what enables models to function as autonomous agents in research. Morgan and Morrison argue that the autonomous capacities of models are revealed in their use as instruments performing a variety of different functions, including theory construction and exploration, measurement, and design and intervention. Through these uses a model can become "an autonomous source of knowledge" (Morrison, 1999 p47) potentially facilitating learning about its associated theory. This characteristic is central to the claim that models should be recognised as playing an independent epistemic role in scientific research.

Morgan and Morrison's assertion that models are partially independent from theory has been largely accepted by philosophers adopting the models as mediator's framework. However, there has been some scepticism over whether it makes sense

to claim that models also have an autonomous relationship with theory; as Daniela Bailer-Jones puts it:

> I find talk of autonomy of models misleading. There is no denying that there always exists some relationship between a model and some theory from which the model draws, and between a model and the phenomenon of which it is a model. In short, there always exist constraints for the relationship between model and theory and model and phenomenon. (Bailer-Jones, 2009 p35)

Eric Winsberg has expressed a similar concern:

> For our purposes, the term "autonomous models" is somewhat misleading. A better term would be "semiautonomous." The claim frequently made by Morrison and Morgan that models are autonomous or independent of theory is meant to emphasize the fact that there is no algorithm for reading models off from theory […]. But to call these models completely "autonomous," at least in this context, is to deny the obvious and strong connections these models have to theory. (Winsberg, 2003 p105)

Bailer Jones and Winsberg appear to be suggesting that it doesn't make sense for models to have both a partially independent and an autonomous relationship with theory. They imply that because models will always have a relationship to the theory which influenced their construction they can never be regarded as fully autonomous and we should just refer to the model theory relationship as partially independent or semiautonomous. Given that partial independence and autonomy play such distinct roles in Morrison and Morgan's account conflating these two properties appears to be a problematic response to a lack of clarity in the distinction between them.

In the section 3.2 I present my historical material outlining how the theory of the rate limiting step influenced the construction of a mathematical model, metabolic control analysis, which then gave rise to a new systemic theory of metabolic control. In the section 3.3 I use this material to elucidate a distinction between the partially independent and the autonomous relationship between models and theories. The analysis of the historical trajectory of a model is crucial to my argument as I focus on identifying different relationships which models can have with theory over time.

3.2 Understanding metabolic control

3.2.1 From metabolic structure to metabolic regulation

Prior to the 1960's biochemists were primarily occupied with investigating the structure of major metabolic pathways and constructing metabolic maps showing the relationships between constituent reactions and metabolites (Newsholme and Start, 1973; Holmes, 1992). Philosophical and historical accounts have focused on the trajectory leading from multiple strands of interdisciplinary research to the emergence of biochemistry as a coherent discipline from the 1930's with this focus on discerning the structure of intermediary metabolism (Kohler, 1975; Bechtel, 1986; Holmes, 1986; Morgan, 1990). By the late 50's this project was widely regarded as being complete. Historians have then examined biochemistries struggle to maintain and establish an identity in relation to the newly emerging molecular biology (Abir-Am, 1992; De Chadarevian, 1996). Laboratory experiments were integral to this area of research, and several philosophers have examined their role in discerning between competing theories of oxidative phosphorylation – the conversion of energy released during respiration into biologically useful ATP (Allchin, 1996; Weber, 2002).

After the 1960's biochemists increasingly turned attention to the regulation of metabolism – how functional metabolism was achieved in the face of on-going perturbations, including changes in available inputs and changes in the demands for end products. As I document in more detail in section 3.2.2, prior to the 1960's biochemists understanding of metabolic regulation was dominated by the idea of a rate limiting step – the slowest step in a metabolic pathway which would set the pace for the activity of the other steps in the pathway. The most well-known development in understanding of metabolic regulation in the 1960's is the Monod, Changeaux, Jacob theory of allosteric regulation and end product inhibition (Monod et al., 1963). This increased attention to the role of systems level interactions rather than the properties of individual components for achieving metabolic control. Existing historical and philosophical accounts of allosteric regulation indicate a smooth transition between research around the rate limiting step and that of end product inhibition (Creager and Gaudillière, 1996; Morange, 2012). David Fell even goes so far as to claim that "the rate limiting step concept was strengthened by the discovery of feedback inhibition in metabolic pathways." (Fell, 2007 p88). However, there is

scope for historical and philosophical work directly addressing the transition between these two perspectives.

The development of metabolic control analysis in the early 1970's is connected to the development of a systemic theory of metabolic control, which is presented by some researchers as being a distinct alternative to the theory of a rate limiting step (Kacser and Burns, 1973). As I detail in section 2.2.3, these researchers explicitly acknowledge that their theory does not overlap smoothly with the theory of end product inhibition. Despite this, metabolic control analysis led to the development of an extensive area of research and is still used as an approach in contemporary systems biology (Fell, 1992; Cascante et al., 2002; Moreno-Sánchez et al., 2008). So far philosophers have paid attention to the impact of metabolic control analysis on theories of genetic dominance (Falk, 2001; Plutynski, 2008). This chapter examines the history and impact of metabolic control analysis in relation to theories of metabolic control.

3.2.2 Joseph Higgins and the rate limiting step

The theory that control over the flux through metabolic pathways was exerted by a single rate limiting step had been around since 1905 when F.F. Blackman introduced it in relation to photosynthesis, "When a process is conditioned as to its rapidity by a number of separate factors, the rate of the process is limited by the pace of the slowest factor" (Blackman, 1905 p289). The idea that one step in a metabolic pathway would be the locus of metabolic control lasted well into the latter half of the twentieth century. However, its theoretical characterisation underwent several qualitative amendments including those by leading figures in biochemistry. For example, Sir Hans Krebs (1957) suggested that it would be the first step in a metabolic pathway. The diversity of different characteristics, alongside a wide range of experimental techniques for detecting rate limiting steps led to confusion amongst researchers who struggled to agree upon the identification of rate limiting steps in major metabolic pathways (Fell, 1997).

Joseph Higgins (1990) first became involved in problems with research on the rate limiting step whilst working with Britton Chance at the Johnson Foundation, part of the University of Pennsylvania. He later recalled a rare atmosphere at the Foundation where Chance, then the director, encouraged interactions between

experimentalists, physicists, and mathematicians. Higgins was involved in the development of analog and digital computers at the Johnson Foundation for investigating biochemical problems from 1949 when he was initially employed as an electronics technician during his undergraduate studies (Higgins, 1961; Higgins, 1990). Through working on a project about control in glycolysis Higgins became aware first hand of the assumptions experimentalists made about the rate limiting step and the range of different characteristics attributed to it (Higgins, 1964):

> Perhaps I overstate, but the concept was that the first reaction of any sequence or at the beginning of a branch point was an irreversible reaction (unaffected by its products). This was the so-called "committed" step and it was also the "rate-limiting" step. It was also the slowest reaction (Whatever that meant). (Higgins, 1990 p46).

In the same reflections on his experience, Higgins describes how the problems with research in this area were brought to light by experimental work indicating that some of what were considered to be the first steps or branch points in this pathway were either reversible reactions or allosterically regulated by their product. In addition he was confused by what researchers meant by the slowest reaction, given that reaction rates are balanced when a metabolic pathway is in a steady state.

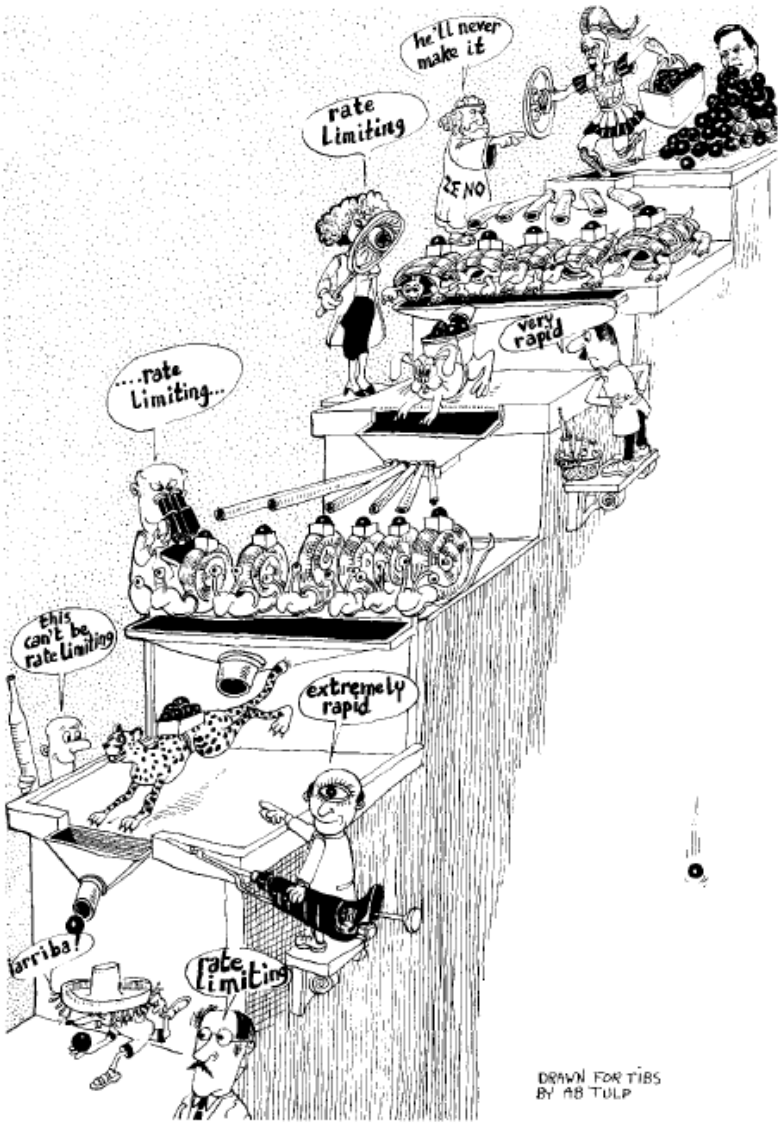


Figure 3-1: Cartoon illustrating the confusion amongst experimental biochemists surrounding attempts to identify the rate limiting step in metabolic pathways (Van Dam, 1986).

At a colloquium on metabolic control at the Johnson Foundation in 1965 Higgins presented an article in which he tried to address these issues. One of his major aims was to use mathematics in order to quantify the concept of control and provide a basis for making clearer inferences from experimental to theoretical work:

> Over the last fifteen years the application of mechanical and electronic analogies has introduced many new concepts such as energy transfer, control and feedback. Such concepts, though fruitful, have generally been introduced and utilized in a qualitative and vague manner. At this stage, it seems necessary to make these concepts more precise in order to attack the general problem of cellular dynamics. (Higgins, 1965 p13)

I think Higgins work can usefully be interpreted within Cartwright's philosophical framework as a process of developing bridging principles in order to concretise the abstract concept of metabolic control. Existing work giving qualitative characterisations of the rate limiting step lacked such bridging principles resulting in a lack of clarity in research. As Cartwright observed when abstract concepts are not concretised "then their introduction is *ad hoc* and the power of the derived predication to confirm theory is much reduced" (Cartwright, 1999a p195).

> Higgins concretised the notion of control in the following manner:
>
> Since "control" is normally used in the sense of "to effect," it is useful to take some function of the change in flux for a given change in one of the SCV as a measure of the control strength. ….

Then, Control Strength $$C = \begin{cases} \dfrac{\partial v}{\partial \alpha_i} \\ \text{or} \\ \dfrac{\partial \ln_v}{\partial \ln \alpha_i} = \underset{R}{\underset{V}{\alpha_i}} \end{cases}$$ (Higgins, 1965 p34)

Higgins bridge principle concretised metabolic control as the control strength of each reaction – the impact of changing the concentration of a particular reaction on flux through the entire metabolic system. In addition he provided equations for calculating the control strength from the relevant experimental measurements. He hoped this would provide a way of clearly interpreting experimental results to determine which reaction had the largest control strength and could be considered the rate limiting step.

With regard to the rate limiting step, Higgins "did not set out to destroy that view" (1990 p47) and did not anticipate any significant repercussions for the theory of metabolic control from his work. He saw himself as working entirely within this framework and attempting to provide experimentalists with a tool which would help

them clarify which steps were rate controlling. He even states that in his analysis “the relationship of the rate controlling step to the fast and slow steps (defined through the values of characteristic times) is easily demonstrated.” (Higgins, 1965 p39). Whilst Higgins saw his analysis as confirming that slow steps would be rate controlling steps he used a particular definition of slow, a slow characteristic time. This referred to the time it took for a reaction to return to steady state after perturbation, rather than the reaction rate. In chapter four I focus on analysing the assumptions that researchers make about the temporal organisation of metabolic systems. Higgins concept of a characteristic time is the equivalent of the concept of relaxations times around which one of the modelling approaches examined is based.

The concept of the rate limiting step was significantly modified during its concretisation. Higgins formulation of the control strength did not rigidly maintain the experimentalists’ assumption that only one reaction would be rate controlling. It left open the possibility that multiple reactions would have large control strengths, and that the control strengths of reactions could vary across different conditions. Higgins himself observes that “there need not be any great difference in the control strengths implying that no one step has dominant rate control” (Higgins, 1965 p39). Such observations indicate the potential of his work to lead to a reassessment of the experimentalists’ assumptions about metabolic control. However, at the time these observations could be considered boundary phenomena as whilst they were noted their potential impact on the theory of the rate limiting step was not recognised (Rheinberger, 1997b p21).

Higgins work was not only influenced by trying to clarify theory. The control strengths were a development from the more general notion of reflection coefficients he had worked on during his PhD – “…when one of the fundamental variables is changed, it *causes* changes in all the other variables of the system. Or conversely, all the variables in the system *reflect* changes in the primary variable.” (Higgins, 1961 p330; see also Higgins, 1963) In developing reflection coefficients Higgins aimed to provide a way of mathematically analysing the behaviour of biochemical systems without having to have the rate constants of all the reactions involved. This allowed the mathematics and experimental data required to be greatly simplified – “The convenience of the reflection coefficient lies in the simplicity of these equations

and the fact that they can be directly tested experimentally." (Higgins, 1961 p335). Higgins was incredibly keen to develop a quantitative analysis which would be useful to experimentalists, but later recalls how experimentalists chanted "chain Higgins to the (lab) bench" (Higgins, 1990 p48) in response to his perceived lack of awareness of what was experimentally feasible (Interactions between modellers developing metabolic control analysis and experimental biochemists are further examined in chapter five of this thesis). As these other factors affected the model building process the equations can be considered to constitute a model which is partially independent from theory rather than a mathematized theory.

3.2.3 Metabolic control analysis

Higgins work was developed into metabolic control analysis simultaneously by two independent teams of researchers; Henrik Kacser and Jim Burns at the University of Edinburgh, and Tom Rapoport and Reinhart Heinrich at the Humboldt University. The two teams acknowledged joint responsibility for developing the model and published a standardisation of the associated terminology in 1985 (Burns et al., 1985). The teams also shared similar motivations whilst developing this mathematical model. However, the two trajectories of model development illustrate different relationships with theories of metabolic control.

In my account I will use the 1985 standardised terminology rather than the terminology used by the two teams in their initial publications (Kacser and Burns, 1973; Heinrich and Rapoport, 1974b). Both teams retain Higgins control strength, which became known as the control coefficient – the influence of a change in enzyme concentration on metabolic flux through an intact metabolic system. Another coefficient was added, the elasticity coefficient, which referred to the influence of an effector on the reaction rate of an enzyme in isolation from the system. This coefficient was included in order to give an idea of how easily the system could manipulate the activity of particular reactions. Two of what were referred to as mathematical theorems were also added. Firstly, the summation theorem, in which the sum of all the sensitivity coefficients in a system add up to one. This provided an important means of assessing the distribution of control amongst reactions and affirming claims about whether only one or multiple reactions had a significant role in metabolic control. Secondly, the connectivity theorem, which provided a means of

relating the elasticity coefficients to the sensitivity coefficients. Even though these parts of the model are referred to as theorems, due to the diverse influences on model building I think it is legitimate to call it a model rather than a mathematized theory. Even the researchers involved engaged in a long running debate about whether to refer to the model as metabolic control analysis or metabolic control theory (Burns et al., 1995).

The two teams had the same three major motivations as Higgins whilst constructing the model. Firstly, they wanted to provide a quantitative tool in order to deal with confusion amongst researchers and clarify theoretical inferences from experimental work; they "attempted to set the problems of biochemical control on a conceptually sound basis" (Kacser and Burns, 1973 p91). Secondly, they wanted to simplify the mathematics required to build models for investigating metabolism at the systems level partly because "such a computation is often impracticable for experimentalists" (Heinrich and Rapoport, 1974b p89). In the initial papers of these two teams it becomes clear that part of this simplification is due to the fact that the coefficients are based on linear approximations of nonlinear relationships. Thirdly, they wanted to build a model for which it is possible to obtain the experimental inputs and provide a useful tool for laboratory researchers and "marry theory to experiment and observation" (Kacser and Burns, 1973 p91). In Kacser and Burns' (1973) paper the importance of the quality of experimentally obtained input measurements for the model was emphasised, in particular, the need for the *in vitro* conditions to mimic the *in vivo* state.

Heinrich and Rapoport were based in the Institute for Physiology and Biological Chemistry at the Humboldt University in Berlin, and their work on metabolic control analysis was submitted as a joint habilitation thesis in 1977. Like Higgins, they promoted it as a tool for discerning the rate limiting step – "C_i equals 1 if the enzyme E_i fully control the flux through the chain, while Ci = 0 indicates the complete unimportance of that enzyme for the flux." (Heinrich and Rapoport, 1974b). They recognised that their approach to discerning the rate limiting step differed significantly from previous work because it required "that an analysis of the features of the regulation of a metabolic pathway has to start from a consideration of the whole chain rather than from the detailed description of single enzymes." (Heinrich and Rapoport, 1974b p95). Despite the recognition that it was a systemic approach,

they initially used their work to discredit particular characterisations of the rate limiting step associated with investigating the properties of individual reactions, not the theory of a rate limiting step in general (Heinrich and Rapoport, 1974a)

Henrik Kacser was based in the genetics department of Edinburgh University. He had a longstanding interest in questions about the specificity of genes for determining the phenotype of the organism and the related phenomena of genetic dominance – the exhibition of the wild type phenotype by heterozygous mutants (Kacser, 1957; Kacser, 1960). He favoured an account which focused on the non-specificity of genotype-phenotype relations and provided an explanation of genetic dominance in terms of the biochemical and kinetic properties of metabolic organisation. Prior to the late 1960's Kacser had not developed a clear or substantiated account of this kind. In 1967 Kacser, along with his PhD student Jim Burns, published an article utilising the equivalent of Higgins' control strengths which began to build such an account (Kacser and Burns, 1967). They suggested that the observation that changes in enzyme concentration can have minimal impact on phenotype is a reflection of the fact that no one enzyme in a system has a dominant control coefficient. In other words, changing the concentration of any individual enzyme in a system does not necessarily have an impact on the systems metabolic flux.

A few years later Kacser and Burns (1973) published their extended version of the mathematical model, containing both coefficients and theorems. This paper focused on presenting the mathematical model in relation to metabolic control, questions about genetic dominance are side lined until the publication of their highly cited paper, "The molecular basis of dominance", in 1981 (Kacser and Burns, 1981). In this paper they use the model to explicitly reject the theory of a rate limiting step and suggest an alternative systemic theory of metabolic control.

> The sum of all the sensitivity coefficients is equal to unity. … Equally possible is that none of the enzymes is of major importance. 'Pacemakers' or 'bottlenecks' do not therefore necessarily exist in a system. It is interesting to note that this conclusion is consistent with the general evidence from inborn errors of metabolism. (Kacser and Burns, 1973 p77)

They proposed that metabolic control is distributed amongst reactions in a pathway and results from interactions between those reactions. The summation theorem

helps support this claim, and also means that the control coefficients are context dependent as their values are determined in relation to the control coefficients of other reactions in the system. It is noteworthy that at this point their assertions about the distribution of control amongst reactions are mainly derived from the form of the mathematical model and only backed up by circumstantial experimental evidence. Even though metabolic control analysis was built as a tool for experimentalists, due to the novelty of many of the experiments required to obtain, the measurements needed to calculate the two coefficients substantive experimental data did not begin to be produced in association with metabolic control analysis until the early 1980's (Groen et al., 1982).

In addition Kacser and Burns provide a hypothetical analysis of what the distribution of sensitivity and elasticity coefficients would be in a pathway involving end product inhibition. The theory of end product inhibition was introduced into work on metabolism in the 1950s (Umbarger, 1956; Yates and Pardee, 1956). It can be regarded as another move towards a systemic theory of metabolic control as it involves explaining the control of metabolic flux through the interaction of reactions, i.e. end product inhibiting the first reaction in the sequence thus regulating its production. Kacser and Burns point out that their analysis does not fit smoothly with this perspective:

> This means, of course that the last enzyme has a Sensitivity Coefficient of almost unity, i.e. it 'controls' flux. Since the sum of the Sensitivity Coefficients in a chain has been shown to be equal to unity, it immediately follows that the rest of the enzymes, *including the 'controlled' one*, have very low Sensitivities and therefore no 'control'… The enzyme has a high Elasticity Coefficient with respect to the controlling pool, but its Sensitivity Coefficient with respect to the system flux is low. (Kacser and Burns, 1973 p86-87 the term sensitivity coefficient is earlier terminology for the control coefficient)

Kacser and Burns are pointing out here that under the framework they set out, the last reaction would have a high control coefficient, i.e. appear to be controlling, as changing the concentration of the reaction would have a large effect on the amount of the end product. However, the first reaction would have a low control coefficient, as changing its concentration would not have such an immediate impact on the end product, yet a high elasticity coefficient as its rate would be incredibly sensitive to the concentration of end product.

3.3 Autonomy of models and theories

3.3.1 Models with lives of their own

I am going to argue that the characteristic of partial independence and autonomy of models and theories can be distinguished as they describe different qualities of the model theory relationship as it changes over time. I will highlight the role that idealising assumptions used during model building play in changing the models relationship to ontological assumptions about metabolic control. In the introduction to this chapter, I examined how partial independence refers to the relationship between the model and existing theory which is established during model construction. In this section I am going to suggest that the autonomy of models from theories refers to the relationship of the model to emerging and future theory.

Morrison and Morgan often extend their description of models functioning as autonomous agents with the claim that a model can have a life of its own. They are of course borrowing this phrase from Ian Hacking who used it relation to the role of experiments in research, and Morrison claims she wants "to argue for a similar characterisation of models as independent entities." (1999 p46). Hacking introduced the phrase "Experimentation has a life of its own" (1983 p150) in order to assert that experiments play more diverse roles in research than theory testing. He is drawing attention to the significant role of experiments in knowledge production, just as Morgan and Morrison are highlighting this role for models. He asserts that laboratory experiments have lives of their own because of their capacity to elicit meaningless phenomena - those which are not anticipated and cannot be made sense of within the existing research framework (Hacking, 1983 p158). In drawing attention to such phenomena, experiments play the role of introducing unanticipated novelties which impact upon the epistemic trajectory of research. There is an on-going debate between philosophers of science about whether experiments, or simulations, carried out with mathematical models can also play this role (Morgan, 2005; Parker, 2009). In the following analysis of the development of mathematical models of metabolic control I want to support my argument that autonomy refers to the relationship of models to future theory by pointing out that the process of constructing models, in

particular the use of idealising assumptions, can sometimes facilitate the introduction of meaningless phenomena into research.

Another strand to my position on the distinction between the partial independence and autonomy of models builds upon the work of Peter Galison. In *Image and Logic*, Galison (1997) examines the development of micro-physics from the perspective of the material instruments, the detectors used in this area of research. Like Hacking with respect to experiments and Morgan and Morrison with respect to models, he wants to emphasise the significance of aspects of scientific practice, other than theory, in shaping the dynamics of research and knowledge production. Galison also adopts Hacking's notion, claiming that material instruments can have lives of their own, and suggests that a signature of this is the distinct dynamics of change exhibited by different aspects of research:

> Instead of depicting the practices of instrumentation, experimentation, and theory as changing synchronously, I want to leave open the possibility that each has its own tempo and dynamics of change. Put in short form: *The periodizing breaks of the various subcultures of physics are intercalated, not necessarily coincident*. (Galison, 1997 p14)

These distinct dynamics indicate that theory, instruments, and experiments are not tied to each other but can exhibit their own independent historical trajectories. Instruments can go on to be used in a different theoretical and experimental context than that with which they were initially associated. Whilst Galison's work relates to material instruments, Morrison and Morgan "claim that what it means for a model to function autonomously is to function like a tool or an instrument." (Morrison and Morgan, 1999b p11) – models can exhibit a variety of instrumental uses in relation to theoretical and empirical aspects of research. The material status of the model is not so important for the question of whether it can be regarded as an instrument. In the next section analysing the case of metabolic control, I will illustrate how models and theories can also sometimes exhibit non-synchronous dynamics of change.

3.3.2 Autonomy of models and theories of metabolic control

In the case of metabolic control presented in section 3.2, the trajectory of the relationship between model and theory clearly illustrates a pattern of non-synchronous changes. There are two stages of model development and two different theories of metabolic control associated with it. However the change in the model

and the change in the theory do not overlap. Higgins developed a model consisting of one coefficient, his control strength, which was associated with the rate limiting step. Heinrich and Rapoport expanded this to include another coefficient and the two mathematical theorems, and initially associated this with the rate limiting step. Kacser and Burns developed a different systemic theory of metabolic control in association with the first stage of model development, the single coefficient, and the second stage of model development, the two coefficients and two theorems.

As I have already suggested, I think that Higgins work can be regarded as giving rise to boundary phenomena. I take these to be similar to Hackings meaningless phenomena. As Rheinberger describes it,

> […] there is again a continuous generation of new phenomena, which need not have anything to do either with the preceding assumptions or with the proposed goals of the experimenter. They usually begin their lives as recalcitrant "noise" as boundary phenomena, before they move on stage as "significant units".(Rheinberger, 1997b p21)

In the case examined they arise from a mathematical model rather than an experiment. Higgins model opened up the possibility that multiple reactions in a system could be playing an important role in metabolic control. Whilst this phenomenon was acknowledged by Higgins it was side-lined and the primary role of the model to help to identify the rate limiting step was maintained. This phenomenon was later given meaning by Kacser and Burns in their development of a systemic theory of metabolic control based upon Higgins work. This potential of the model to support the rejection of the rate limiting step and lead to the development of an alternative theory was not anticipated by Higgins who was intending to clarify the existing research framework.

As Morrison and Morgan emphasise the partial independence of the model from existing theory was crucial for allowing the model to function in this manner. The idealising assumptions introduced during the process of model building which establish this partially independent relationship played a crucial role in transforming the relationship of the model to a particular theoretical perspective on metabolic control and the ontological assumptions about metabolic control associated with these. Higgins initially developed the model to facilitate identification of the rate limiting step, rather than test the theory of the rate limiting step, as such it was based

on the ontological assumption that a single step in a metabolic pathway was responsible for metabolic control. Higgins used control coefficients to simplify the mathematical equations and to connect the model to hypothetically obtainable experimental data sets. Control coefficients involved the idealising assumption that the impact of changing the concentration of a single reaction on the behaviour of the entire system was a reflection of that reactions role in controlling metabolic flux. Crucially this idealising assumption did not overlap exactly with the ontological assumptions of a single rate limiting step. This meant that the definition of Higgins control strength was loosened from a one to one connection with the definition of a rate limiting step, and this partial independence opened up the space in which the model could function autonomously as an instrument for theory construction. It enabled Kacser and Burns to change the ontological assumptions associated with the model. The model was no longer attached to the ontological assumption that only a single reaction was involved in metabolic control, instead the distribution of control amongst metabolic reactions became what was being tested by the model. New ontological assumptions were attached to the model during its development. For instance, the summation theorem, which stated that all the control coefficients in a system must add up to one, introduced the assumption that the control coefficients of individual reactions were dependent on their context in a particular system rather than being a context independent property.

Importantly Kacser and Burns' grounds for rejecting the rate limiting step were primarily based on thinking about and working with the model. They rejected the notion that mathematical and computational modelling only involved working out consequences of a "system constructed in a known manner" (Kacser and Burns, 1967 p23), and thought of it as a creative process which could stimulate the development of novel theoretical perspectives. Their extension of the model and change in theory of metabolic control was not significantly influenced by an alternative, already existing systemic perspective on the control of metabolism. As I illustrated they explicitly pointed out how their position did not sit comfortably with the theory of end product inhibition. Additionally, the new theory was constructed prior to the use of the model to analyse experimental data sets. Remember, that the novelty

of the experimental techniques required to obtain the relevant input measurements meant that this data was not available until the early 1980's.[5]

3.3.3 Autonomy and partial independence

Clear non-synchronous dynamics of change in theory and model help to illustrate that a model can have two very different relationships with theory. The theory which pre-exists the model and was an important factor influencing model construction – in the case developed previously, the theory of the rate limiting step – has a relationship of partial independence with the model. The fact that this relationship is partially independent enables the same model to establish relationships with other theories. Partial independence means that models do not have a one-to-one relationship with their associated theory. As theories are underdetermined by models, one model can be associated with multiple different theories. The capacity for models to be associated with multiple domains of enquiry has also been analysed by Sergio Sismondo in relation to island biogeography. Sismondo suggests that "we should expect solid models to have multiple interpretations." (Sismondo, 2000 p253). On the other hand, when the model serves as an instrument facilitating the construction of a new theory, this is an example of an autonomous relationship between model and theory. This characteristic of autonomy is particular to the process of constructing an emerging theory. Once a theory is established, if that theory influences changes in the model, then a partially independent relationship will become established between the model and theory.

The sense in which the relationship between a model and an emerging theory can be considered autonomous needs to be examined further. For similar reasons as Bailer-Jones and Winsberg in relation to models, Galison also chooses to emphasise that the lives of instrumentation, experimentation and theorisation are not completely independent, but semi-autonomous, because they are constrained and connected to each other somehow. He is not arguing that the different dynamics of changes in these areas of research are a mark of their full autonomy. However, Galison recognises features of this semi-autonomy that I will argue enable instruments to be considered as fully autonomous in a different sense. He pays

[5] Nancy Nersessian (1999) provides another philosophical perspective on the role of model construction as a tool for scientific problem solving, and the connection of this particular form of 'model based reasoning' to innovative conceptual change.

attention to the fact that constraints should not be regarded as negative because they play an important role in creative aspects of research: “Because constraints restrict moves, they shape the theorist’s positive research program – giving a problem-domain form, structure, and direction.”(Galison, 1997 p16).

The importance of constraints in scientific research for knowledge production and generating epistemic novelty has been explored by philosophers of science (Rheinberger, 1997b; Galison, 1997; Rheinberger, 2010). Constraints provide specific boundaries for creative work and facilitate the formation of novel ideas and practices. In the case examined the constraints provided by Higgins mathematical model allowed Kacser to start developing more coherent ideas about the relationship between metabolic organisation and genetic dominance. Tarja Knuuttila has recently argued that it is important to pay attention to how the particular form of the model, which is established during the model building process, constrains and facilitates the future use and development of the model (Knuuttila and Voutilainen, 2003; Knuuttila, 2011). However, along with Atro Voutilainen, she also points out that:

> …although models bear traces of their intended use in their construction, they can also be used in many other ways. As epistemic artefacts, models are open-ended things that have their own history and dwell in our research practices in manifold ways as both tools and objects of enquiry. (Knuuttila and Voutilainen, 2003 p1494)

It is the open-ended character as to what might occur whilst working within these constraints that is so important for regarding the relationship between models and emerging theory as autonomous. The process of constructing models can elicit meaningless phenomena, in a comparable way to experiments. As I discussed above, the introduction of idealising assumptions which do not exactly overlap with the theoretical ontological assumptions upon which the model is built play a crucial role in loosening the relationship between a model and particular theory. These are phenomena which are not anticipated and may initially be downplayed by researchers working within the established framework. However their recognition as important can initiate changes in understanding and is an example of the independent epistemic role that models can play in research. Importantly as this process is open rather than closed there is an element of surprise and unpredictability which I think warrants recognising an autonomous relationship between model and theory.

3.4 Summary

In this chapter I have provided a clarification of Morrison and Morgan's distinction between models as both partially independent and autonomous from theory in scientific research. By examining the case of the development of a mathematical model, metabolic control analysis, and its relationship to theories of metabolic control, I have made a distinction in terms of how the model theory relationship changes over time. A model has a relationship of partial independence with theory which pre-exists the model and influences its construction. This relationship is partial as theory, and its associated ontological assumptions, are only one of many factors influencing model development. A model has a relationship of autonomy with emerging and future theory which the model may elicit the construction of. As a model is partially independent from pre-existing theory the process of construction involving the introduction of idealising assumptions can result in the introduction of meaningless phenomena into research. These phenomena may be picked up on and given meaning in the context of a new theory. As this process is an unanticipated and open-ended outcome of the initial model development I consider it to illustrate an autonomous relationship between model and theory. It is important to distinguish this from the partially independent relationship which models have with pre-existing theory in order to understand how models can play an independent role in knowledge production.

4 Temporal decomposition: A strategy for building mathematical models of complex metabolic systems

> "While the requirement of wholeness urges us to consider more and more interactions in describing a given system, time, insight, and precision of instruments place limits on the number of interactions that can be considered. … Fortunately, there are at least three different types of simplification - spatial, temporal, and functional - that arise naturally to limit the complexity of systems and make their analysis feasible.' (Savageau, 1976 p81)

Abstract

In 'Discovering complexity' Bechtel and Richardson (1993) highlighted the connection between how biologists investigate the world and the type of explanations they give. This chapter extends their account of how we investigate the world by examining the strategies used by researchers to build mathematical models of complex metabolic systems between the 1970's and 1990's. Bechtel and Richardson analysed how researchers decompose complex systems by reducing the number of variables included in the model, thus simplifying them and making them suitable objects for research and understanding. Bechtel and Abrahamsen (2005) later distinguished two types of decomposition: 1) Structural decomposition, starting with the identification of the relevant component parts and 2) functional decomposition, starting with the identification of the relevant component operations. I use my case studies to argue that temporal decomposition should be recognised as an additional strategy for investigating complex metabolic systems. Temporal decomposition involves the identification of the relevant dynamic variables. Existing accounts of decomposition are based on the assumption of a spatial hierarchy which classifies modules according to the frequency of interactions between components. Temporal decomposition is based on the assumption of a time hierarchy which classifies variables as dynamic or constant according to the relative speed with which properties of the system change.

4.1 Introduction

In 'Discovering Complexity' (1993) William Bechtel and Robert Richardson focus on exploring the strategies used by scientists to research complex biological systems. Their interest in this is motivated by an awareness that the strategies used by scientists also shape the form of the explanations they give:

> 'These strategies are, from one perspective, the procedures that define how humans approach the problem of understanding the world. They define *how* we think about the world. From another perspective, the procedures humans embody constitute assumptions about the structure of the world, or of part of it to be explained. They define *what* we think about the world.' (Bechtel and Richardson, 1993 p11)

Bechtel and Richardson's work is foundational for the philosophy of mechanism, which has become an expansive area of research in philosophy of science over the past 20 years (Machamer et al., 2000; Glennan, 2002; Bechtel and Abrahamsen, 2005). This research has focused on developing an account of the mechanistic explanations frequently found in biology to replace a law-based perspective on biological explanations. Whilst much critical philosophical enquiry has paid attention to evaluating *'what* we think about the world', Bechtel and Richardson's account of '*how* we think about the world' has been left largely undeveloped by philosophers working in this area and received little direct critical attention. In this chapter I evaluate and extend Bechtel and Richardson's account of decomposition as a strategy for investigating complex systems. I argue for a specific form of decomposition, temporal decomposition, using three case studies of mathematical models of metabolism from between the 1970's to the early 1990's.

Researching complex systems presents the issue of how to delimit the appropriate features and make them manageable foci for investigation. Bechtel and Richardson provide an insightful account of how scientists approach this problem of simplification, by 'finding laws and variables sufficient to explain what does and does not happen' (Bechtel and Richardson, 1993 p12). They describe the reduction of variables included in these models as crucial to this process as it limits the space of possible explanations. The reduction of variables usually involves the imposition of assumptions about the systems organisation. These assumptions both reflect researcher's ideas about how the system may work and the variety of constraints,

‘such as limitations on available mathematical models, or simple technological limitations on what data can be gathered.’ (Bechtel and Richardson, 1993 p12), that they have to work within. The assumptions frequently have related ontological and idealising dimensions. This process of simplification aims to retain the pertinent features of the systems for the particular property being investigated. These features can then become the focus of researcher’s investigations which they build their understanding of the system and its properties around. Bechtel and Richardson introduce these strategies in the context of problem solving in general indicating they are relevant for understanding scientific investigations involving either experiments and / or mathematical models.

Bechtel and Richardson go on to develop an account of a particular strategy, decomposition:

> ‘*Decomposition* allows the subdivision of the explanatory task so that the task becomes manageable and the system intelligible. Decomposition assumes that one activity of the whole system is the product of a set of subordinate functions performed in the system.’ (Bechtel and Richardson, 1993 p23)

They focus on a particular type of decomposition, which in later work is referred to as functional decomposition. This involves first decomposing the complex behaviour into a ‘set of subordinate functions’ and then locating those functions in particular parts of the system which carry them out. The major assumption involved is that the system is decomposable into discrete functional behaviours and discrete component parts. This also implies that the system is in a sense hierarchical.

Bechtel and Richardson’s account of decomposability is developed from work by Herbert Simon (1962) and William Wimsatt (1972). Simon gives a description of the hierarchical organisational structure which allows systems to be decomposed. He argues that a decomposable system will be one where interactions within sub-components of the system will be more frequent than interactions between subcomponents. That fact that some interactions do occur between subcomponents makes the system nearly decomposable rather than fully decomposable. He exemplifies what it means for biological systems to be decomposable by listing levels of spatial organisation ‘Taking the cell as a building block, we find cells organised into tissues, tissues into organs, and organs into systems.’ (Simon, 1962 p469).

Bechtel and Richardson utilise this criterion of minimal interaction between sub-components as one of the primary reasons for the assumption that biological systems are decomposable. For the purposes of my argument it is important to note that they define a decomposable system as one which is 'modular in character, with each component operating primarily according to its own intrinsically determined principles.' (Bechtel and Richardson, 1993 p25). They discuss the hierarchical structure which enables biological systems to be decomposed in terms of the spatial organisation of systems. The parts of the system can be decomposed and studied independently because their behaviour is assumed to be largely determined internally, not through interactions with other components of the system. This assumption facilitates the reduction of , firstly 1) the number of components and 2) the number of interactions between components.

In his early work Bechtel (1984; 1986) analysed biochemistry in the first half of the twentieth century. During this period biochemists were largely occupied with discerning constituent reactions and metabolites and the functional relationships between them in order to construct metabolic maps (see also Kohler, 1975; Holmes, 1986). This work forms an important basis for Bechtel and Adele Abrahamsen's later distinction between two different strategies for decomposing complex biological systems, functional decomposition and structural decomposition (Bechtel and Abrahamsen, 2005). To perform functional decomposition researchers initially focus on component operations. They start by decomposing the main operation of the system into sub-operations, following this with an identification of the active component parts performing those operations. Performing a structural decomposition involves researchers initially focussing on the component parts. Researchers begin by identifying parts which may be involved in performing operations and then go on to identify parts and their organisation with particular operations. Bechtel and Abrahamsen illustrate these two modes of decomposition with a metabolic example:

> 'The component operations are characterised differently in different domains, but often involve transformations to some substrate. The biochemical system that performs metabolism in cells, for example, catabolises glucose into carbon dioxide and water. The component operations are then characterised in terms of individual chemical reactions on a series of substrates (e.g. oxidising or reducing them, adding or removing H2O, etc.). A successful functional decomposition of the system will identify each operation and its passive parts (the

> substrate and the resulting product). What it lacks is a specification of the active parts - that is, the enzyme that initiates and guides each reaction. Once all parts are identified a structural decomposition accompanies the functional decomposition.' (Bechtel and Abrahamsen, 2005 p433)

For them, the component operations are the general types of metabolic reactions, and the parts are the passive metabolites and active enzymes which are involved in these reactions. Investigating and understanding complex metabolic systems using the strategy of functional decomposition involves beginning by identifying the relevant types of reaction. The strategy of structural decomposition involves beginning by identifying the relevant passive and active parts.

I am going to argue for an additional type of idealising decompositional strategy, one which is based on an ontological assumption of a time hierarchy of biological processes, in addition to structural and functional decomposition based around an ontological assumption of a spatial hierarchy of biological parts. As I discussed in the thesis introduction, section 1.2.3, comparatively little philosophical attention has been paid to the temporal organisation compared to the spatial organisation of biological systems, this chapter contributes to this under examined area of philosophy of biology.

Bechtel and Abrahamsen's account of structural and functional decomposition has continued to be used to frame analysis of developments in biochemical systems biology. They have been used in the debate about whether mathematical models of the dynamics of metabolism provide mechanistic explanations (Bechtel and Abrahamsen, 2010). These two categories of decomposition have also been used to structure philosophical histories of the different types of approaches to modelling used in contemporary systems biology (Krohs, 2010; Krohs and Callebaut, 2007). Both of these topics involve paying attention to mathematical models which involve representations of how biochemical processes unfold over time. Through paying attention to the decompositional strategies used during the production of dynamic mathematical models this chapter is contributing a fresh perspective from which these issues could be analysed.

In this chapter I explore historical material which starts where Bechtel's early work left off. After work on the structure of metabolic pathways was considered

mostly complete, biochemists began to focus on metabolic regulation and the temporal dynamics of metabolism. This was coupled with an increase in the use and development of mathematical models. I firstly explore the two major constraints upon building mathematical models of metabolism: 1) availability and reliability of kinetic data resources and 2) the complexity of mathematical techniques for representing reaction kinetics. Secondly, I analyse three approaches to building metabolic models given these constraints: 1) relaxation times 2) descriptive rate laws, and 3) constraint based optimisation. I then use this material to develop my philosophical argument that these three approaches to building simplified mathematical models involve the strategy of temporal decomposition which is based upon the assumption of a time hierarchy of biological processes[6]. In the final section of this chapter I address the relationship between the spatial and temporal decomposition of the system.

4.2 Historical development of mathematical models of metabolism

From the 1960's onwards biochemists became increasingly interested in moving beyond research which focused on the structural relationships between reactions and metabolites and the kinetic behaviour of individual systems components and towards research which investigates how more complex behaviours, such as metabolic regulation, arose out of interactions between components in systems (Newsholme and Start, 1973; Creager and Gaudillière, 1996). This focus on metabolic regulation increased the need for researchers to pay attention to the temporal dynamics of metabolic behaviour. Rather than looking at metabolic systems as static pathway structures, they wanted to understand how the time sensitive behaviour of multiple reactions was co-ordinated to achieve and maintain metabolic functionality.

This increase in attention to the temporal dynamics of metabolic behaviour involved the development of mathematical models as a research tool. Laboratory biochemistry is confronted with the methodological issue of constructing appropriately simplified *in* vitro experimental systems which could be used to make inferences about the more complex *in vivo* state (Strand, 1999). Hans Jörg

[6] My distinction between strategies and approaches is based on the observation that one approach to mathematical modelling can involve several different strategies and the same strategy can be used in multiple different approaches.

Rheinberger (1997b) gives an insightful description of the process of 'Establishing the *in vitro* system of protein synthesis 1949-52' where 'The solution had to be found somewhere between the bog and the simple system'(Rheinberger, 1997b p55). Now appropriate mathematical models needed to be constructed which would aid understanding of *in vivo* metabolism. In a comparable fashion to *in vitro* experimental systems, mathematical models needed to be established which involved a functional degree of simplification yet retained enough detail for them to inform understanding of *in vivo* metabolic behaviour.

4.2.1 Constraints on model building: Metabolic data and its mathematical representation

The strategies employed by researchers during their development of mathematical models for investigating the temporal dynamics of metabolic behaviour were shaped by two major constraints. Firstly, the availability of different types of empirical data resources and assumptions about the objects and process to which different types of data pertained. Secondly, the complexity of the mathematical tools required to represent these different data types.

I have identified in the literature three major types of empirical data which mathematical modellers were working with (See Kacser and Burns, 1967 p11-12). The first, compositional, refers to the components which are found in a metabolic system. It provided information about the identity of enzymes and metabolites. The second, stoichiometric, refers to the structural relations between systems components. It gives information about which metabolites identity is transformed by which reactions and was used to construct maps of metabolic pathways. The third, kinetic, refers to the dynamic behaviour of individual reactions. It provided information about how the rate of a particular reaction in isolation changes in relation to changes in substrate or effector concentration (Kacser and Burns, 1973 p65). These three data types were taken to exhibit a hierarchy, where knowledge of kinetic parameters presupposes knowledge of stoichiometry, and knowledge of stoichiometry in turn implies knowledge of systems composition (Kacser and Burns, 1967 p13-14; Heinrich et al., 1977 p4). Researchers developing mathematical models initially assumed that all three types of data were required to build useful

mathematical models of metabolic systems. Perceptions of the availability of data and assumptions about the compositional and stoichiometric properties of metabolic systems differed significantly from the perceived availability and assumptions about the kinetic properties of metabolic reactions. These differences had significant repercussions for mathematical model building.

By the 1970's those developing mathematical models perceived the compositional data set and knowledge of reaction stoichiometry to be almost complete (Heinrich et al., 1977 p2). Acquiring this information was considered to be a once fruitful area of research which was drawing to a close – '…activity in this area of exploration may have reached its zenith. Thus completion of the molecular inventory of the cell is in view...' (Savageau, 1972 p122). Now that the metabolic map of well-studied organisms such as *E.coli* was considered known it was time to move research towards investigating more complex behaviours arising from interactions between systems components. Roger Strand states that 'The *in vivo - in vitro* problem is defined as the problem of justifying claims about the *in vivo* system on the basis of evidence obtained with the *in vitro* system.' (Strand, 1999 p274). He argues that the identity of biomolecules and biochemical processes, including reaction stoichiometry, are data types which are potentially subject to this issue. However, initially researchers developing mathematical models of metabolism do not refer to the *in vivo in vitro* problem in relation to compositional and stoichiometric data. These data types were often evaluated in terms of an ontological assumption that the structural relationship between reactions and metabolites were invariant properties of metabolic systems which would remain constant in the *in vitro* and *in vivo* context. I refer to this assumption as ontological because it is part of researchers assumed knowledge about metabolic systems. (I will examine the historical development of ontological assumptions about compositional and stoichiometric data types and their connection to idealising assumptions which facilitate model building further in chapter five.) This meant that once data had been collected about stoichiometric relations in one context those relations were assumed to be correct in all contexts. This ontological assumption contributed to the perceived completeness of stoichiometric data. A second ontological assumption which affected the procedures used to build mathematical models was that stoichiometric data referred to discrete changes in metabolite identity. Reactions were regarded as

converting metabolites from one identifiable state directly into another, as opposed to a gradual change in metabolite identity. This assumption is based upon the understanding of metabolites as thermodynamically relatively stable structures which require an increase in energy facilitated by an enzyme mediated reaction in order to transform them into different thermodynamically stable metabolites (Savageau, 1976 p23). The speed of the transformation was so rapid that any intermediate states could be ignored and the process modelled as a direct transformation from one metabolite to another. This meant that reaction stoichiometry could be represented using simple algebraic equations giving the quantities and identities of reaction inputs and outputs.

The availability of biochemical data about individual reaction kinetics was perceived to be reasonable, but not as comprehensive as compositional and stoichiometric data types. Kinetic data refers to the change in reaction rates as a function of changes in concentrations of substrates and effectors. Awareness of the *in vivo in vitro* problem featured heavily in discussion about kinetic data amongst those developing mathematical models:

> 'elegant and definitive investigations can be carried out with isolated, well characterised, enzyme systems; however the experimental conditions must differ from those existing *in vivo* and the relevance of such work is therefore uncertain.' (Wright and Kelly, 1981 p105)

Researchers regarded kinetic properties as context variable in comparison to the context invariant stoichiometric relations. In particular, contemporary experimental studies challenging assumptions about the relative concentrations of enzymes and metabolites usually used in experimental investigations of enzyme kinetics were frequently mentioned as cause for concern (Srere, 1967; Sols and Marco, 1970). Researchers were worried about the validity of using *in vitro* kinetic data in order to build mathematical models of metabolic systems which could be used to enhance knowledge of *in vivo* metabolism. Reaction rates were assumed to vary in a continuous manner compared to the discrete changes of reaction stoichiometry. The relationship between reaction rates and concentrations of substrates and effectors was often complex and non-linear. These ontological assumptions meant that more complicated mathematics involving ordinary differential equations was required to represent the kinetic properties of individual reactions.

Given these constraints of data availability and the complexity of its mathematical representation researchers took a variety of approaches towards developing mathematical models. The different approaches contained various levels of detail trying to find the appropriate space between the incredibly simple and the very complex:

> ‘On the one hand, the model may be intended to give a minutely detailed representation of the biological system which includes the properties of every component regardless of its role … The other extreme constitutes idealized skeleton models which are intended to represent the basic features of a biological system without direct confirmation with experimental data.’ (Heinrich et al., 1977 p5)

The first models of metabolic systems veered towards the first extreme of ‘minutely detailed representation’. These modelling approaches aimed to build dynamic models involving full kinetic parameters for all component reactions. The basis of the models parameter values on high quality detailed experimental kinetic data was considered important for model quality (Garfinkel, 1969). These models involved the use of coupled ordinary differential equations to reconstruct the dynamic behaviour of metabolic systems (Garfinkel et al., 1970). However, the vast majority of these models ended up with what were referred to as intractable or stiff equations, ‘Differential equations whose computer solution is very slow when numerical procedures depending on difference methods are used.’ (Garfinkel et al., 1970 p494). An intractable mathematical problem is one where the solution is in principle obtainable but would take an impractical amount of time to reach (Hopcroft and Ullman, 1979). Numerical procedures are used to find approximate rather than exact solutions to differential equations. In stiff equations the solution is unstable; it varies significantly depending on the step size used during the calculation. The way to get around this is to reduce the step size used during the calculation. However this vastly increases the number of calculations which need to be carried out extending the time it takes to solve the equation. A single stiff equation will have a significant impact on the solution of a system of coupled differential equations. This level of detail in mathematical models quickly came to be regarded as problematic as the solutions to the models could rarely be achieved, and the kinetic data resources were not sufficiently available.

Each of the three modelling approaches outlined below exhibits a different response to problems with these initial highly detailed models[7]. They all attempt to achieve an appropriate simplification for exploring metabolic regulation involving making idealising assumptions whilst working within these two constraints. At this stage building mathematical models of metabolism was predominantly a very abstract area of research and this is reflected in my accounts of the three approaches. Most of the work in this area did not involve using actual biochemical data sets to build mathematical reconstructions of specific metabolic systems. As I've pointed out, modellers, or theoreticians as they were commonly referred to, were aware of the constraints arising from differences in data availability. However, these constraints provided a general rather than a specific motivation for their work. Modeller's methods for tackling them were largely based on abstract assumptions rather than actual biochemical data sets. At this time mathematical modelling was a fringe area of biochemistry which had an often fraught relationship with laboratory based researchers (Cornish-Bowden and Cárdenas, 1990; Cornish-Bowden, 1989). I will examine interactions between modellers and experimental biochemists in detail in chapter five of this thesis. In this chapter I focus particularly on approaches which attempt to model complex systems involving relatively large numbers of components, rather than those which simplified by focusing on two or three component systems. I also focus on highlighting how the models were appropriate simplifications in the sense of being mathematically soluble and capable of being built using experimental data, rather than assessing their epistemic value.

4.2.1.1 Relaxation times

The approach of analysing relaxation times aimed to construct models which could be used to simulate the dynamic behaviour of particular metabolic systems, and to eliminate the problems caused by stiff equations. It stems from the observation that a time hierarchy exists amongst the relaxation times of metabolic reactions. A metabolic system is assumed to be in a steady state when there is no accumulation of intermediary metabolites, i.e. no change in concentration of intermediary reactions substrates and products. If, for instance, the availability of

[7] This process of simplifying the complexity of highly detailed models is comparable to Eric Winsberg's work on simplifying intractable models of physical systems discussed in the section 1.2.1 of the thesis introduction (Winsberg, 2001).

external systems inputs increases then this may perturb the reactions from their steady state. The relaxation time is the amount of time it takes for a reaction to regain a reach a steady state or equilibrium after perturbation. The relaxation time of reactions in a system can differ significantly, some reactions take a very long time to return to a steady state, others do so almost immediately (Higgins, 1965; Park, 1974; Heinrich et al., 1977). The existence of multiple different time scales of reaction rates in models of metabolic systems was identified as a major cause of stiff equations which prevented the solution of models containing large numbers of coupled differential equations. Researchers developed the analysis of relaxation times as a means of working with this time hierarchy in order to simplify the mathematical simulation of the systems behaviour and avoid stiff equations.

This approach to model building exploited the potential to simplify mathematical representation of the behaviour of reactions existing at either extreme of the time scale hierarchy:

> ‘Some variables are so slow that they remain approximately constant during a specified time-period, and others are so fast that they are in a steady state. Thus only a few essential variables need to be considered.’ (Rapoport et al., 1976 p450)

Reactions which are slow i.e. had a long relaxation time were assumed to maintain a constant reaction rate. In this case the concentrations of substrates or products associated with reactions accumulate or decline but the reaction rate appears to remain constant. The concentrations of substrates or products will take a long time to return to the pre-perturbation steady state levels as the reaction rate is not responding to the perturbation. Reactions which were fast i.e. had a short relaxation time could be assumed to maintain a steady state, i.e. constant metabolite concentrations. In this case the concentrations of substrates or products associated with reactions appear to remain constant but the rate of the reaction varies. The reaction rate has responded quickly to the perturbation and almost immediately returned substrate and product concentrations to the pre-perturbation steady state. These assumptions meant that the behaviour of these reactions could be represented by differential equations with analytic solutions (Park, 1974). Only the remaining kinetic parameters required representation using differential equations which had no straight forward analytic form. This massively simplified the mathematical representation of the system resulting in more efficient simulations of

dynamic systems behaviour. Variations of this approach to simplified metabolic model building continued to be developed through the 1980's (Liao and Lightfoot, 1987; Joshi and Palsson, 1989). In the later models researchers explicitly refer to what they are doing as carrying out a 'temporal decomposition' of the different time scales of metabolic reactions (Joshi and Palsson, 1989 p516).

It provided an approach for building simplified mathematical models by making idealising assumptions about the kinetic behaviour of reactions based on ontological assumptions about the temporal organisation of metabolic reactions. However, it did not directly address the constraints that modeller's perceptions of biochemical data placed upon model building. Full experimental characterisation of the kinetics of all the reactions in a system was still required to identify pools of reactions exhibiting the two extremes of relaxation times. In order to address the limited amount of kinetic data available, in particular high quality kinetic data obtained in *in vitro* conditions which mimicked the *in vivo* state, researchers focused on building models of a very simple biochemical system, the red blood cell (Rapoport et al., 1976 p499). Instead of reducing the amount of kinetic data required to construct the model, model builders chose simple biological systems to model. The simplicity of the actual biological system being modelled would have had the additional benefit of further simplifying the mathematical representation of the system, as well as ensuring that minimal biochemical data would be required for model building (Liao and Lightfoot, 1987 p254).

4.2.1.2 Descriptive rate laws

Descriptive rate laws provide a means of building an approximate mathematical model of the kinetic behaviour of an individual reaction from a minimal amount of experimentally determined input parameters. The most famous of these is Michaelis-Menten kinetics (Michaelis and Menten, 1913). This is frequently used to estimate the behaviour of enzymes involving just a single substrate. It requires two experimentally determined parameters: the maximum reaction rate achieved by the system at any substrate concentration and the Michaelis constant, the substrate concentration at which the reaction reaches half its maximum rate (Cornish-Bowden, 1995). From these parameters the Michaelis-Menten equation can be used to

provide an analytic solution for the differential equations describing the nonlinear rate of the reaction over a wide range of possible substrate concentrations. The Hill equation (Brown and Hill, 1923) is another widely used descriptive rate law developed to provide an estimation of the behaviour of allosteric reactions, reactions where the reaction rate is also influenced by the concentration of an effector.

Michaelis-Menten kinetics and the Hill equation are examples of steady state rate laws. They are based upon the quasi steady state assumption, which says that the concentration of the intermediary enzyme substrate complex remains constant, but other variables i.e. concentrations of each reactions substrate and product can vary. This assumption reduces the number of dynamic variables requiring description in the model. This assumption is justified by an additional assumption:

> "let *a* be the initial concentration of *A*, *e* the total concentration of enzymes, *x* the concentration of *B* produced after time *t*, and *p* the concentration of enzyme combined with substrate at time *t*. We suppose *e* and *p* to be negligibly small compared to *a* and *x* ....Now since *p* is always negligible compared with *x* and *a* - *x*, its rate of change must, except during the first instant of the reaction, be negligible compared with theirs.' (Briggs and Haldane, 1925 p338)

In other words, it is assumed that the concentration of substrates will greatly exceed the concentration of enzymes. In which case saturation of the enzyme will be so rapid that the concentration of enzyme-substrate complexes can assumed to be stable justifying the removal of this concentration as a dynamic variable from the model (Eigen, 1968). [8]

Descriptive rate laws for the behaviour of individual reactions were used as a method for building simplified models of metabolic systems containing multiple reactions. Biochemical Systems Theory, developed by Michael Savageau (1969a) from the University of Michigan, is an example of an approach which used descriptive rate laws to simultaneously try and address problems arising from stiff equations and the lack of available kinetic data. It is a method for building models which can be used to perform dynamic simulations of systems level behaviour.

[8] This assumption was challenged by contemporary experimental evidence suggesting that cells could often exhibit similar concentrations of substrates and associated enzymes. *In vitro* experiments pertaining to the kinetics of reactions in isolation were carried out under conditions relating to this assumption, i.e. low enzyme concentrations and high substrate concentrations. The fact that these conditions did not accurately reflect the *in vivo* state was one of the major reasons researchers were sceptical that kinetic data could be used to inform understanding of the behaviour of intact metabolic systems. (Eigen, 1968; Sols and Marco, 1970)

Savageau argued that whilst existing descriptive rate laws could be used to build simplified models of metabolic systems, numerous types of more complex reactions existed for which descriptive rate laws did not exist:

> 'Although these approximations reflect the basic nonlinearity of the rate law, they are not mathematically simple enough. Even at the level of the individual enzyme catalysed reactions these approximations are inadequate for dealing with multiple reactants and modifiers.... The complex relations that result from combining just a few such reactions quickly become mathematically intractable.' (Savageau, 1972 p72)

The lack of descriptive rate laws for complex reactions was not only regarded as problematic because it still resulted in insoluble systems of equations, but also because the more complex reactions also required the most experimental data to describe reaction kinetics (Savageau, 1972 p67).

In order to deal with these issues Savageau developed a new mode of approximating non-linear reaction kinetics which could be used to write a mathematical description of the behaviour of any reaction regardless of its complexity. The rate law was also based on the quasi steady state assumption (Savageau, 1969a p366), and for convenience additionally assumed a steady state for all except one of the concentration variables – substrates and effectors – associated with a reaction (Savageau, 1969b). These assumptions were applied indiscriminately to all reactions in a metabolic system. The method involved the use of bode analysis, splitting the kinetic behaviour of the reaction into four simplified stages, and then using a power law to obtain a non-linear description of the overall behaviour. This provided a simplified mathematical description for the kinetics of any reaction eliminating stiff equations from the model. Savageau's descriptive rate laws could be used to build mathematical models based on coupled differential equations for simulating either steady state or dynamic systems level behaviour.

Savageau was also keen to point out the extent to which his non-linear power law approximation reduced the amount of kinetic data required during model building. He compared it to the use of descriptive rate laws based on a linear approximation used in other mathematical models. These approaches used an approximation technique which assumed a linear relationship between reaction rate and substrate concentration. However, this meant the range of changes in

concentrations across which the approximation was valid was very low and the amount of experimental data required for model building relatively high. Savageau emphasised that his non-linear approximation technique allowed for approximation across a large range of changes in concentrations from relatively little experimental input data. He argued that his approximation technique was experimentally validated 'It is evident that these systems are accurately described by the approximation theory over at least a 100-fold change in their input concentration.' (Savageau, 1971 p617). The use of descriptive rate laws to build simple mathematical models made them both mathematically tractable, and massively reduced the experimental data required for model construction.

4.2.1.3 Constraint based stoichiometric models

The formation of constraint based stoichiometric models was a major new development in attempts to build models of metabolic systems. It simultaneously created a new solution to the two major problems detailed kinetic models faced, 1) the complex mathematical representation of systems and 2) the lack of available kinetic data. It did this by getting rid of the requirement for kinetic data and correspondingly the need to use differential equations in the mathematical representation of the system. As we shall see, this resulted in a completely different type of mathematical representation of the system using only the algebraic equations required to represent discrete stoichiometric interactions, and discontinuous changes between systems states.

> 'The present models of the discrete type representing metabolic change as a discontinuous process of transition between different metabolic states. Most existing metabolic models are of the continuous (kinetic) type, representing metabolic change as a smooth progression in reactions rates and metabolite concentrations.'(Watson, 1986 p26)

Constraint based stoichiometric models are based only on stoichiometric data and involve the use of a different mathematical tool, linear programming. These models provided a way of using stoichiometric data to give something more than just structural understanding of the metabolic map. They could be used to analyse the distribution of metabolic flux between intermediary reactions in different conditions[9].

[9] The term metabolic flux is sometimes used in to refer to the reaction rate when the reaction is located in a system rather than in isolation from the system (Fell, 1997 p4).

These models were initially simultaneously developed in several distinct contexts (Watson, 1984; Fell and Small, 1986; Majewski and Domach, 1990). These strands of research were picked up by Bernhard Palsson who was based at the University of Michigan. Palsson had a long standing interest in developing metabolic models of metabolism, and as we have seen was particularly involved in approaches based on the analysis of relaxation times. He became the main researcher developing and advocating the potential of stoichiometric constraint based models of metabolic systems (Varma and Palsson, 1994). In 1992 and 1993 along with two PhD students Palsson constructed two large scale models, one of a hybridoma cell line and the other of *E.coli* (Savinell and Palsson, 1992; Varma and Palsson, 1993).

The first stage in building these types of models was the reconstruction of a stoichiometric matrix for the entire system which provided a coded description of the metabolic map. If a metabolite was consumed by a reaction a -1 would be entered in the appropriate column, if it was produced then this would be recorded as a +1, and if a metabolite and reaction did not directly interact then a 0 would be entered. The ability to represent the network structure in this simple binary manner stems from the understanding that reactions facilitate discrete changes in metabolite identity. Secondly, constraints are applied to how metabolic flux can be distributed in the network. The main constraint is a steady state assumption which in this case is applied to the concentrations of all the intermediary metabolites which are assumed to remain constant. The values of the input and output metabolites for the system were the only metabolite concentrations allowed to vary. This steady state assumption is based upon the assumption that:

> 'Metabolic flux models are based on the assumption that metabolic transients are more rapid than both cellular growth rates and dynamic changes in the organism's environment. Metabolism typically has transients which are shorter than a few minutes and therefore metabolic fluxes are in a quasi steady state relative to growth and typical process transients.' (Varma and Palsson, 1994 p994)

In other words, it is assumed that over the timescale addressed in the model there is a slow growth rate for the system compared to the speed of the flux of intermediary metabolites and so there will not be any accumulation or growth which would deviate the system from the steady state (Savinell and Palsson, 1992 p424). It also assumed that no environmental perturbations will occur which may also disturb the system

from the steady state. These assumptions enabled the conversion of the differential equations used to represent dynamic mass balances in the algebraic equations giving steady state mass balances simplifying the mathematical representation of the system (Varma and Palsson, 1994). As the system is in a steady state it is also assumed that the rate of flux through each reaction remains constant.

The steady state assumption reduces the number of potential flux distributions through the network. However, network structure alone still underdetermines steady state flux distribution and multiple feasible possibilities remained. In the third stage, linear optimisation was used to identify a unique steady state flux distribution. Metabolic flux distribution is affected by the kinetic behaviour of the reactions in a system, but,

> 'in the absence of detailed knowledge of enzyme kinetics, we can estimate the metabolic distribution if we postulate the "objectives" that underlie the cell's behaviour. An underdetermined set of equations can be solved uniquely, given an objective function using linear optimisation techniques.' (Savinell and Palsson, 1992 p424)

An optimisation criterion would be set for the network, such as minimise input consumption and maximise output production. An algorithm was then used to quickly search through the feasible steady state flux distributions and identify the steady state flux distribution which would best achieve this criterion. The use of an optimisation criterion mitigated the requirement for kinetic data to be included in the model.

Constraint based stoichiometric models could be used to identify different steady state flux distributions that a metabolic system might be able to achieve. However, because no kinetic information was included in the model it didn't say anything about whether a system would actually be able to regulate itself so as to achieve this distribution. Additionally, the model couldn't be used to investigate how a metabolic system made a continuous transition from one steady state to another. Unlike kinetic models, these gave representations of discrete and discontinuous distributions of metabolic flux.

The elimination of kinetic information from the model and the use of linear optimisation greatly reduced the complexity of the mathematical representation of metabolic systems (Savinell and Palsson, 1992 p499). However, the development of this stoichiometric constraint based modelling approach by Palsson and colleagues was primarily motivated by the difference in availability and reliability of

stoichiometric and kinetic data, rather than the mathematical complexity associated with kinetic models based on differential equations. Whilst enough data existed to construct kinetic models of red blood cells by analysing relaxations times, with constraints based optimisation Palsson aimed to increase the diversity of types of systems for which large scale metabolic models could be built (Varma and Palsson, 1993 p478).

4.3 Temporal Decomposition

I argue that temporal decomposition is an important strategy used by researchers when developing mathematical models to explore the behaviour of metabolic systems. Temporal decomposition is distinct from structural decomposition, which takes its starting point as discerning the relevant parts, and functional decomposition, which takes its starting point as discerning the relevant functions. In comparison, researchers using temporal decomposition are trying to discern the relevant time scale. These three types of strategy were often recognised by researchers developing these models of metabolism (See epigraph to this chapter). I hope my analysis will also convince philosophers of science of their existence. I am proposing that temporal decomposition is a strategy for investigating complexity which neatly fits into Bechtel and Richardson's initial interest in understanding how researchers worked within constraints to construct simplified explanatory models by 'limiting the relevant variables and imposing assumptions about the form of relevant laws' (Bechtel and Richardson, 1993 p16). The major constraints explored above were differences in the availability of stoichiometric and kinetic metabolic data, and the mathematical tools available to represent these different sorts of metabolic data.

Researchers carry out a temporal decomposition of a metabolic system when they make assumptions which facilitate the reduction of the number of dynamic variables which need to be included in the model. Dynamic variables, in the case of metabolism, are those describing continuous changes, either in concentrations of components or the rates of reactions, over time in the model of the system. In a complex biological system dynamic variables will often exhibit non-linear behaviour. The major category of assumptions about systems organisation involved in reducing

the numbers of dynamic variables are various steady state assumptions, but other assumptions about metabolic regulation are also relevant.

Steady state assumptions are idealising assumptions based on the claim that the concentration of a part of the system can be assumed to remain constant. This claim is based on ontological assumptions about the organisation of metabolic processes into a time hierarchy. The case studies illustrated three different versions of this. In the relaxation time approach, for fast reactions the concentration of substrates was assumed to be in a steady state. Descriptive rate laws assume that the concentration of enzyme-substrate complex is in a steady state. Constraint based optimisation assumes that the concentration of substrates and products associated with all reactions are in a steady state. Other assumptions are involved in the removal of dynamic variables. Analysing relaxation times also involved the assumption that slow reaction rates are constant and constraint based optimisation assumes that all metabolic fluxes are constant. This assumption is similar to the steady state assumption as it involves the assumption that something which could be considered a dynamic variable is constant, just that in this case it is the rate of a reaction rather than the concentration of a part. I will leave the optimisation assumption used in the constraint based optimisation approach out of my analysis because it is not involved in the removal of dynamic variables from the model.

The decomposition of metabolic systems into variables which are to be considered constant and those which are to be considered dynamic involves the ontological assumption that organisation of metabolic processes constitutes a time hierarchy. Researchers explicitly use the notion of a time hierarchy to describe the organisation of dynamic variables and their approach to investigating complex metabolic systems.

> ‘Time hierarchies are a general feature in nature. Any theoretical or experimental approach requires confinement of the system to be investigated with respect to the time ranges. These restrictions determine the experimental methods to be applied as well as the structure of the models assumed.’ (Heinrich et al., 1977 p467)

Researchers invoke a metabolic time hierarchy in order to classify processes according to relative timescales. In the relaxation time approach three levels of a time hierarchy are identified: 1) fast reactions – constant metabolite concentrations and dynamic reaction rates, 2) average reactions – dynamic metabolite

concentrations and dynamic reaction rates and 3) slow reactions – dynamic metabolite concentrations and constant reaction rates.

> 'This temporal separation of phenomena can greatly simplify the analysis of complex systems. The variables in such a system that respond much faster than the phenomena of interest can be assumed to be at their steady state value, those variables responding more slowly than the phenomena of interest can be assumed to be constants or slowly varying parameters' (Savageau, 1976 p81-82)

Descriptive rate laws involve the identification of two levels of a metabolic time hierarchy: 1) fast processes – saturation of enzyme by substrate, constant concentration of enzyme substrate complex, and 2) average processes – dynamic metabolite concentrations and reaction rates. Constraint based optimisation methods are based on a comparison of two levels of a metabolic time hierarchy: 1) slower processes – growth rate and environmental perturbations, constant metabolite concentrations 2) faster processes – flux of metabolic intermediaries, constant metabolic fluxes. Identifying a time hierarchy of metabolic processes facilitates researcher's discrimination and justification of which variables need to be treated as dynamic, and which can be treated as constant in the model. Within a particular timescale the levels of the temporal hierarchy are assumed to be relatively independent from each other. Researchers assume that they can treat variables at one level as constant without significantly affecting the behaviour of variables occupying other levels of the time hierarchy.

My analysis of these three different approaches to temporal decomposition confirms that it can be approached in two different ways:

> 'First, the idealized model may be extracted from a detailed one which was able to describe actual experimental data. The simplification can either be arrived at by the use of the time hierarchy or topological contraction. A second approach which may be considered the reverse of the preceding one is to set up an idealized model as a preliminary step to a more detailed description. It is particular appropriate if a few details such as rate law and parameter values of a metabolic system are known.' (Heinrich et al., 1977 p11)

Analysing relaxation times involves detailed experimental kinetic data sets to inform the selection of specific dynamic variables which can be removed from the model of the system. Models based on descriptive rate laws or constraint based optimisation involve the application of assumptions to all reactions in a system reducing or

eliminating dynamic variables in the mathematical model. These assumptions are often only validated later against experimental data.

4.3.1 Three types of decomposition: Structural, Functional, and Temporal

Let us return to Bechtel and Richardson's definition of decomposition quoted in the introduction to this chapter:

> '*Decomposition* allows the subdivision of the explanatory task so that the task becomes manageable and the system intelligible. Decomposition assumes that one activity of the whole system is the product of a set of subordinate functions performed in the system.' (Bechtel and Richardson, 1993 p23)

In the introduction I outlined Bechtel and Abrahamsen's example of decomposition applied to metabolism. Structural decomposition discerns the relevant active and passive parts, enzymes and metabolites. Functional decomposition discerns the relevant types of reaction, the operations those component parts are involved in. These two forms of decomposition simplify the model of the system by reducing the number of components and the number of operations. I have argued that temporal decomposition, in relation to metabolism, discerns the relevant dynamic variables, concentrations of enzymes and metabolites, and rates of reactions. It simplifies the model of the system by reducing the number of dynamic variables which need to be included in the model. It is still based upon the same assumption that a set of subordinate functions produce the behaviour of the system as a whole. In my case studies, temporal decomposition involves decomposing the time dynamics of those subordinate functions, rather than the structural relations between types of operations and component parts. It should be made clear that the three decompositional strategies are not mutually exclusive. In Bechtel and Abrahamsen's account of decomposition in relation to metabolism one was followed by the other. For instance, a structural decomposition would be followed by a functional decomposition. Similarly, temporal decomposition is usually preceded by structural and / or functional decomposition. Researchers will have made decisions regarding which component parts and which functional operations to include in their model before they make decisions about which of the variables associated with those parts and operations they are going to treat as dynamic or constant.

In this chapter I have made the case that temporal decomposition should be recognised alongside structural and functional decomposition as a strategy used to model complex metabolic systems. However, this opens up the question of whether decomposition is a homogenous strategy or whether there are features which distinguish the three approaches. For instance, in structural decomposition the emphasis is on decomposing a whole into its parts, in this case a metabolic system into its active and passive parts, its metabolites and enzymes. Whereas,, in temporal decomposition, the emphasis is on distinguishing different levels of a time hierarchy using the criteria of relative speed of change of different variables. Is there a significant difference between the relationship between parts and wholes and the relationship between different levels of a time hierarchy? From one perspective parts and wholes seem to pick out levels of a hierarchy which are clearly discrete from each other whereas the levels of a time hierarchy appear to change continuously. Leading on from this, how does functional decomposition and a hierarchy of functions intersect with both the spatial and temporal organisation of metabolic systems? Biological functions involve objects which occupy space e.g. metabolites, and processes which occur over time, e.g. reactions. Given that functions have spatial and temporal dimensions understanding the relationship between the three forms of decomposition is a challenging topic. This line of questioning raises interesting philosophical questions about the ontological and epistemic status of and relationships between structural, temporal, and functional decomposition. Whilst I do not attempt to provide thorough answers to these questions here, in the following paragraphs, and in section 7.2.2 I provide some relevant reflections for further consideration.

In the introduction to the thesis and this chapter I highlighted how the interpretation of Simon's account of the organisation of nearly decomposable systems by Wimsatt (1972) and Bechtel and Richardson (Bechtel and Richardson, 1993) has focused on features of the spatial organisation of biological parts. However, Simon places equal emphasis on the temporal and spatial organisation of complex systems in his original argument for near-decomposability. His argument is based on the claim that there is overlap in the hierarchical organisation of the spatial and temporal dimensions of systems. Indeed, his technical paper which precursors his more general argument in "The Architecture of Complexity" (Simon, 1962) is entitled "Aggregation of Variables in Dynamic Systems"(Simon and Ando, 1961). In both these papers Simon illustrates his argument with the example of heat diffusion through a perfectly insulated building. He describes a building with rooms containing

cubicle offices. Initially there is a thermodynamic disequilibrium between the cubicles and between the rooms. in which case, the temperature within the rooms would reach thermodynamic equilibrium before the temperature between the rooms equalised.

> 'The main theoretical findings from the approach can be summed up in two propositions: (a) in a nearly decomposable system, the short-run behaviour of each of the component subsystems is approximately independent of the short-run behaviour of the other components; (b) in the long run, the behaviour of any one of the components depends in only an aggregate way on the behaviour of the other components.'(Simon, 1962 p474)

Simon is arguing for a direct correlation between spatial and temporal scales – larger spatial structures will co-occur with processes operating over longer timescales. A system can be decomposed into subsystems because the frequency of interactions is higher within subsystems than between subsystems and this is correlated with the observation that properties of subsystems will undergo changes at a faster pace than properties of the systems which they comprise. In fact in a slightly later account, 'The organisation of Complex systems' (Simon, 1977), Simon explicitly highlights the role of temporal decomposition as a simplifying strategy for investigating nearly-decomposable complex systems:

> '…we can break the sequence of characteristic frequencies into three parts: (1) low frequencies, …. ; (2) middle-range frequencies; and (3) high frequencies … . Motions of the system determined by the low-frequency modes will be so slow that we will not observe them – they will be replaced by constants.' (Simon, 1977 p250)

The equal significance that Simon gives to the temporal and spatial hierarchical organisation of complex systems provides further support for my account of temporal decomposition. However, I want to challenge the exact overlap in spatial and temporal organisation which Simon appeals to. I think the close correlation observed by Simon between the spatial and temporal organisation of systems is an artefact of the particular example he choose to illustrate his position, the perfectly insulated building. I don't think spatial and temporal decompositions of the system need necessarily overlap. In building upon Simon's ideas even Wimsatt observed that the spatial and temporal organisation of systems can change independently (Wimsatt, 1976 p237). Take the example of the relaxation time approach, here

biochemists are dealing with a single level of spatial organisation involving reactions and metabolites. However, within this one level of spatial organisation three levels of a temporal hierarchy are identified – fast reactions, average reactions, and slow reactions. This indicates that the assumption that spatial and temporal organisation are directly correlated needs to be carefully examined when looking at how biologists appeal to hierarchies in order to delimit the focus of their enquiry. In any case, my notion of temporal decomposition is clearly distinct from structural decomposition as defined by Bechtel and colleagues. Bechtel defines structural decomposition as the identification of the relevant parts, and in "Discovering Complexity" appeals to Simon's criterion of "Minimal interactions among parts"(Bechtel and Richardson, 1993 p25) in order to delimit subcomponents. I am defining temporal decomposition as the identification of relevant dynamic variables, and appeal to the criterion of the relative speed of change of systems properties to discriminate between which are relevant and irrelevant.

4.4 Concluding remarks

In this chapter I have argued that we should recognise three decompositional strategies researchers use when making the simplifications necessary to investigate complex metabolic systems. Alongside decomposing the system into the relevant parts and the relevant component operations as argued by Bechtel and Abrahamsen, researchers also decompose systems into the relevant dynamic variables. Accounts of structural decomposition are often connected to an ontological assumption of a spatial hierarchical organisation, according to which levels and modules can be discriminated according to the degree of interactions amongst component parts. Similarly, temporal decomposition is based on the ontological assumption of a time hierarchy where the different levels reflect the relative speed of change of properties of the system. The time hierarchy facilitates idealising assumptions about which variables can be treated as constants and which need to be treated as dynamic during the timescale required to investigate the phenomena of interest. Idealising assumptions which play a key role in temporal decomposition, such as steady states, have a widespread use in the analysis of complex dynamic systems. In addition, there are notable instances of biochemists during this period

segregating much wider ranges of biological processes according to different timescales, for instance evolutionary and biochemical processes (Heinrich et al., 1977 p3; Heinrich and Schuster, 1996 p75). This indicates that the conclusions of this analysis could have a more general relevance outside of philosophical perspectives on mathematical models of metabolism.

5 Experimenters and modellers interactions in biochemical systems biology: Research commitments and the definition of problem spaces

Abstract

In this chapter I consider interactions between experimenters and modellers, and between sub-groups of modellers, in biochemical systems biology. I base my analysis on two of my own case studies of interactions between these groups in early biochemical systems biology, and two case studies from MacLeod and Nersessian's work on current biochemical systems biology (MacLeod and Nersessian, 2013c; MacLeod and Nersessian, 2013a; MacLeod and Nersessian, 2013b). In the case studies from early biochemical systems biology interactions are frequently hostile and characterised by attacking and defensive positions. In contrast, open collaborative interactions are established in the case studies from current biochemical systems biology. I frame my analysis in relation to Peter Galison's work on interactions between modellers and experimenters in high energy physics (Galison, 1995; Galison, 1999). Galison gives an account of what I refer to as closed collaborative interactions. I argue that a crucial factor accounting for the three different types of interaction – hostile interactions, closed collaboration, and open collaboration - is the quality of researcher's commitments and in what sense they function as constraints. Researcher's commitments relate to their assumptions about best practice in research. I propose two significant dimensions of commitments, 1) running from rigid and specific, to flexible and broad commitments, 2) whether researchers perceive differences in commitments between research communities as negative, neutral, or positive. I show how different configuration of these dimensions are associated with different types of interactions, and research dynamics associated with three types of problem spaces; externalised, internalised, and adaptive.

5.1 Introduction

Research involving bringing together the resources of multiple different disciplinary approaches – including differences in methods, data, and explanations - has dramatically increased in the life sciences over the past two decades (Parker et al., 2010; Vermeulen et al., 2013). This trend is particularly noticeable in systems biology where the increasing quantity and diversity of large scale biological data sets has served as an impetus for bringing researchers from a wide range of disciplines together, for example; molecular biologists, ecologists, mathematicians, engineers, informaticians and clinicians. The growing emphasis on collaborative research is partly motivated by changing research methods as mathematical modelling has become increasingly important in the analysis of biological data. It is also motivated by the type and ambition of research questions which are frequently geared towards understanding the dynamics of complex, and often highly specific, biochemical systems with a view towards developing targeted applications.

Philosophers and sociologists have used this shift towards collaborative research to examine what it takes for researchers from different backgrounds to develop collaborative working practices. The interface between experimental biochemists and mathematical modellers has received particular attention as the process of building mathematical models of biochemical systems has come to constitute a major research activity (Calvert, 2010; Calvert and Fujimura, 2011; Penders et al., 2008). This body of work has emphasised the many challenges involved in this process as researchers struggle with communication problems and different standards for research practices and knowledge claims between these two groups. However, despite these difficulties this is a diverse collection of researchers who are committed to having a collaborative dimension to their work. As Calvert points out, ‘With enough desire, commitment, and labour these differences may not only be surmounted, they may be productive.’ (Calvert and Fujimura, 2011 p162). The potential productivity and creativity in bringing diverse perspectives and methods to bear on challenging research problems is a common theme in philosophical analysis (Longino, 2002). The concept of integration, with its explicit rejection of processes of reduction and unification, has been developed and

expanded in the context of examining collaborative research practices in contemporary systems biology (Brigandt, 2013a; O'Malley and Soyer, 2012).

However, instances where groups of experimenters and modellers fail to establish collaborative working practices are also apparent. In this chapter, I use a comparative approach to identify the characteristics which inhibit and facilitate collaborative interactions. I present two of my own case studies from early biochemical systems biology. They recount interactions between modellers and experimenters and sub-groups of modellers which took place from the early 1980's to the early 1990's, the period just before biochemical systems biology began to take off as a major area of research. The case studies illustrate hostile interactions between these groups of researchers. Despite the impetus for collaborative interaction researchers fail to establish these and end up occupying attacking and defensive positions. As a contrast I use Miles MacLeod and Nancy Nersessian's recent ethnographic work on research practices in two current biochemical systems biology laboratories (MacLeod and Nersessian, 2013c; MacLeod and Nersessian, 2013b; MacLeod and Nersessian, 2013a). This work provides examples of successful collaboration between modellers and experimenters who often occupy positions of openness and interest in relation to differences in their methods and explanations.

My analysis highlights the importance of differences in the characteristics of research commitments in understanding why particular types of interactions emerge between experimental and modelling communities. In particular, the way in which those research commitments function as constraints on the practices of a community of researchers. According to Sabina Leonelli research commitments encompass:

> 'items as diverse as the theoretical perspective held by biologists; their research goals and interests; their ways to perform research and interpret protocols; and their assumptions about the representativeness of their research materials and the applicability of their results.' (Leonelli, 2009b p202)

Research commitments provide guidance for a research community in relation to acceptable research questions, methods, and explanations. As such, the commitments of researchers are related to constraints. They constrain the way in which researchers think and act. They also give rise to practical constraints, for

example those arising from limitations in the accessibility of resources - data, finances, computation power, and biological samples - required to implement a particular methodology. By research commitments I am referring to what researchers assume to be best practice in research. They can include commitments to other sorts of assumptions, for instance the use of particular idealising assumptions, and other aspects of research such as materials and instruments.

My analysis of different types of interactions that occur between experimental and modelling communities highlights the significant role that differences in two different characteristics of researchers commitments play. Firstly, sometimes commitments provide only loose guidance and are flexible and malleable, in other cases they can be so rigid and specific that they constitute barriers and standards that researchers must maintain in order to carry out research which is acceptable to the community which they belong. Secondly, I include in my definition of commitments researchers attitudes towards plurality in the commitments of different groups of researchers. Researchers who are committed to the superiority of their approach may view differences negatively, whereas researchers who view their commitments pragmatically may perceive differences in a positive light (This dimension of research commitments will be further examined in the fifth section of this chapter where I describe three different problem spaces). I argue that different combinations of these qualities affect in what sense research commitments constrain research and the types of interactions which occur between groups of modellers and experimenters in biochemical systems biology.

The impetus for mathematical modellers and experimenters to interact, and the difficulties involved in this, are common themes in the natural sciences. Peter Galison's (1995; 1999) extensive work on high energy physics constitutes one of the most detailed philosophical analysis of collaborative interactions between groups of experimenters and theorists (In Galison's account, theorists carry out mathematical modelling). His work has been used in relation to current discussions about collaborative interactions in systems biology (Calvert, 2010; Calvert and Fujimura, 2011; Nersessian and Newstetter, Forthcoming). In the next section of this chapter I give an interpretation of Galion's work emphasising the role that the researcher's commitments play in facilitating a particular type of 'closed' collaboration between these two groups. I will use this to further frame and contrast my analysis of different

types of interactions between modellers and experimenters in biochemical systems biology. In the third section of this chapter I present case studies of hostile interactions in early systems biology and analyse how researchers commitments intersect with inter group interactions. In the fourth section I present two case studies from current systems biology and analyse how researcher's commitments intersect with open collaborative interactions. In the final section I argue that each of the three types of interactions is connected to researchers operating in a different problem space.

5.2 Galison: Commitments and closed collaboration in high energy physics

Galison describes the work of independent groups of experimental and theoretical physicists as being heavily constrained by different sets of research commitments:

> "constraints enter the discussion because they mark the endpoints of scientific enquiry, the boundaries beyond which inquirers within the community find it unreasonable to pass. Such borders may be at the lofty plane of conservation laws and symmetries or buried in the common knowledge of plastics, metals, and computer chips. (Galison, 1995 p14)

Researchers within a given community will have commitments to particular theories and more methodologically oriented beliefs and understandings about how to use the relevant materials. This includes both the physical materials used in experiments, and the mathematical equations and computers used in constructing models. The fact that these commitments constitute clear boundaries defining acceptable and unacceptable research practices suggests that they are rigid and specific. They give researchers firm guidance on what counts as acceptable research practices and knowledge claims for that group. The *for that group* is important. Galison describes a situation where researchers think their commitments constitute the best way of working for their sub-community, not for high energy physicists in general.

Galison aims to show that despite these distinct identities researchers from different communities can engage in successful collaborative interactions. Whilst different communities of researchers are defined by clear sets of commitments, Galison describes their relationship with each other in a way which indicates that they are tolerant of inter-cultural differences:

> "Throughout such exchanges there is no attempt to make experimentalists into theorists or vice versa. On the contrary, the concept of collaboration embraced by the physicists during the war involved reinforcement of these sub-cultures and an emphasis on exchange."(Galison, 1999 p153)

When groups of experimenters and theorists come into contact with each other during the course of their research Galison describes a situation in which the groups are happy to accept differences in the way the two cultures operate. There is no attempt to impose the standards for research, including the interpretation and meaning given the material involved in exchanges, from one group to another. Neither group regards itself as operating under superior research commitments and both are happy to accept that different styles of research will be appropriate in different settings.

I want to label this type of interaction between groups of experimental and theoretical physicists, closed collaboration. I refer to this as closed because even though exchange is taking place imported items are used and interpreted according to the commitments of the community that has received them. The commitments of the gifting community are not part of the exchange process. I argue that this type of interaction is facilitated by a situation where researchers have strong research commitments, but are tolerant of differences in research commitments between themselves and other groups of researchers with which they are interacting. Galison argues that semi-autonomous research subcultures can retain their independent identities whilst not being isolated from each other. They can enter into collaborative exchanges through establishing trading zones. These are zones in which the groups can exchange aspects of their work, for example 'theorists trade experimental predictions for experimentalists results,' (Galison, 1999 p146). They do this through establishing contact languages in order to facilitate communication. Galison stresses that these contact languages do not have to involve any common meaning or value to the things being exchanged:

> 'Whatever their differences, none of the participants in these debates had any illusion that they and their interlocutors were "really speaking" through the same concepts of quarks and jets' (Galison, 1999 p851)

Communication and exchange can take place without any compromise or alteration in the commitments which define the different subcultures. Two subcultures do not have to share values and research standards in order to collaborate with each other.

I wish to refer to this sort of exchange as closed collaboration because, even though collaborative exchange is taking place, the two research cultures involved remain closed to the influence of the theoretical commitments and research practices of the other culture. When something enters into the trading zone, it effectively goes through a process of neutralisation, it becomes detached from the research commitments of one culture and reinterpreted through the research commitments of the other. In closed collaborative interactions the rigid and specific yet tolerant qualities of research commitments mean that inter-cultural differences are not seen in a negative or positive light because a research culture behaves as if it is immune to their influence.

Galison describes these sorts of research commitments and collaboration as being connected to a particular research dynamic. Firstly, because researchers are heavily attached to their commitments and are not easily influenced by the research commitments of other groups, Galison describes research cultures as undergoing extended periods of stability in research (Galison, 1999 p145). Secondly, Galison connects the rigidity of commitments and the heavy constraints they place upon creativity in research:

> ‘Each of these broad classes of constraints helps restrict the laboratory moves and verbal conclusions that appear reasonable to the working experimentalist. Each helps isolate phenomena and divide them into classes. It is the progressive imposition and acceptance of these constraints that constitutes the separation of signal from background.’ (Galison, 1987 p255)

Through providing researchers with clear internal parameters for how to carry out research and what counts as good research, constraints give form and direction to scientific work. However, at the same time they allow novel stable phenomena to emerge which might lead researchers to question their existing assumptions and ways of working and lead to discontinuity and changes in the commitments of a research culture. These creative processes emerge due to the internal dynamic of a research culture practising according to its own research commitments. Galison is

not describing a situation where novelty and innovation primarily result from interactions with external research cultures.

It is worth pointing out that, whilst I have attributed this position on the relationship between constraints, inter-cultural interactions, and research dynamics to Galison, aspects of it resonate strongly with the work of other influential philosophers of science. Firstly, there is some overlap here between Galion's notion of constrained stable research cultures and Ian Hacking's notion of particular styles of research which are self-vindicating (Hacking, 1992a; Hacking, 1992b). Secondly, Galion's account of creativity chimes with Hans Jörg Rheinberger's notion of experimental systems consisting of technical objects which provide a structured context within which more vague epistemic things can start to be discerned (Rheinberger, 1997b)**.** My analysis of four case studies from biochemical systems biology illustrates different sorts of commitments which can be held by communities of experimenters and modellers and how they facilitate two other kinds of interactions between the two groups, hostility and open collaboration.

5.3 Interactions between modellers and experimenters in early biochemical systems biology

5.3.1 Metabolic control analysis and experimental biochemists

> 'The lack of progress in certain areas of biochemistry (and genetics) has, in my opinion, been due to the addiction of its practitioners to false concepts and, like all addicts, they are not very effective workers. … All is not bleak, however. There is a cure. This consists of reading the relevant papers until one has understood them. Once free of the addiction, unlike the ex-alcoholic, one is cured for life.' (Kacser, 1983 p311)

In the third chapter of this thesis I examined the relationship of metabolic control analysis to theories of metabolic control. In this chapter I focus on the difficult relationship those developing metabolic control analysis had with experimental biochemists. Interactions between the two groups are documented in exchanges of letters in Trends in Biochemical Sciences and in the proceedings of a symposium on metabolic control attempting to bring experimenters and modellers together in 1989.

The interactions are frequently hostile with both groups attacking each other's work and responding defensively.

The mathematical modellers were initially expecting and seeking a mutually beneficial exchange with experimenters (Kacser and Burns, 1973; Heinrich and Rapoport, 1974b). The mathematical model they produced was intended to be a useful tool for experimenters to facilitate clearer understanding of metabolic control in order to "marry theory to experiment and observation"(Kacser and Burns, 1973 p91). In exchange they expected experimental biochemists to invent and carry out the experiments necessary to produce the new types of data required by the model. By the beginning of the 1980's there had been virtually no engagement of experimental biochemists with metabolic control analysis, and even in the early 1990's modellers were still lamenting that the approach 'has not been applied widely enough in experimental systems.' (Liao and Delgado, 1993 p221). Modellers were particularly concerned that their systemic perspective on metabolic control would only be validated once the model had been used to analyse actual experimental data sets.

As I explored in chapter three of this thesis, experimental research on metabolic control in the 1970's was dominated by the concept of a rate limiting step - a single step in a metabolic pathway which was responsible for controlling the rate of flux through the entire pathway. Modellers observed that there were frequent disagreements about which step in a particular pathway was rate limiting (Fell, 1997). They created a model which they hoped would clarify these disputes in two ways: first, by providing quantitative definitions for determining which reactions were involved in metabolic control; and second, through providing a systemic perspective on control where multiple reactions in a system could be involved in regulation, and where the controlling properties of reactions weren't intrinsic but context dependant. The model required two types of experimental measurements both of which weren't produced by already established experimental procedures in biochemistry. Firstly, measurements carried out *in vivo* of the impact of changing the concentration of an enzyme on systems flux. These were used to calculate the control coefficient, the contribution of an individual enzyme to metabolic control. Secondly, measurements carried out *in vitro* of the impact of changing the concentration of molecules – substrates, effectors, products – associated with a reaction on its rate, but in conditions which mimicked the *in vivo* state. These were used to calculate the

elasticity coefficients, a measure of how much the activity of individual reactions in the system could be modified (Kacser and Porteous, 1987b). The model involved a connectivity theorem describing the relationship between control coefficients and sensitivity coefficients which was based on the assumption that changing the concentration of a reaction in a system was the equivalent of modifying the reaction rate.

In 1982 the first paper was published which involved the production of the specific data types required and its analysis using metabolic control analysis. This study used enzyme inhibitors to conduct experiments affecting the activity of enzymes *in vivo* to calculate the control coefficients and analyse metabolic control in mitochondrial respiration (Groen et al., 1982). A short article was promptly published highlighting the development in *Trends in Biochemical Sciences* (Gillies, 1983). This article stressed that, whilst the initial publications on metabolic control analysis had been too abstract and theoretical to be of much use to experimentalists, the recent development of relevant experimental techniques 'makes the quantification of pathway control less esoteric and more accessible to biochemists in general' (Gillies, 1983 p3). Instead of welcoming this article, modellers reacted with anger and this sparked a series of hostile exchanges in the same journal. They were annoyed that Gillies thought that the initial publications had been too theoretical and stressed that, throughout its development, metabolic control analysis had been intentionally presented in a manner which had stressed the importance of experimental work and was accessible to all biochemists (Kacser, 1983; Porteous, 1983). They used these responses to express concerns that biochemists had intentionally ignored metabolic control analysis for years because they held a prejudice against the importance of theory (Porteous, 1983 p202). These two response letters additionally highlighted where modellers thought experimenters had gone wrong in their research on metabolic control, and how experimental research should be carried out in light of the development of metabolic control analysis. They chastised biochemists for using 'qualitatively evocative, but quantitatively meaningless and misleading terms' (Porteous, 1983 p201). Instead they stressed that metabolic control analysis provided a general and quantitative framework which could be used to facilitate the clear interpretation of experimental results and cure 'the addiction of its practitioners to false concepts.' (Kacser, 1983 p311).

In 1984 another team of biochemists published a paper using metabolic control analysis to analyse whether phosphofructokinase was the rate limiting step in glycolysis (Bosca and Corredor, 1984). They concluded that being rate limiting was not an intrinsic property of reactions but dependent on their context in a particular biological system. Again this paper was met with a barrage of attacks from modellers, who were keen to point out that Bosca and Corredor were completely wrong to use metabolic control analysis to try and identify rate limiting steps when the model was intended to present an alternative to this perspective on metabolic control (Fell, 1984; Porteous, 1985). Porteous made a plea to the journal editors to 'indicate that your prestigious journal is not incline to publish metaphors as substitutes for measurements … If the authors come back at you quote George Riemann's apt definition: 'Science is an attempt to understand Nature by means of exact concepts.' (Porteous, 1985 p15). Researchers aligned with metabolic control analysis were upset that it had not been used in the exact manner they intended and considered its associated concepts to have been misinterpreted. Bosca and Coredor's paper was soon defended by the well-known biochemist Efram Racker who argued that, on the contrary, whilst much remains unknown about metabolic systems qualitative observations are very useful and warned against 'the danger of applying clean thinking to dirty systems.' (Racker, 1985 p270).

These debates led to a symposium being organised in 1989 in order to bring modellers and experimenters together and to try and establish more productive interactions between them (Cornish-Bowden and Cárdenas, 1990). I am going to divide the areas of contention between these two groups into four categories: 1) mathematical modelling, 2) biochemical experiments, 3) idealising assumptions, 4) theories of metabolic control. Each of these problems had a different significance for the ongoing disagreements and each was resolved to a different extent by the papers at the symposium and subsequent research (Fell, 1992).

Gillies (1983) initial remarks about the technical inaccessibility of the original model for biochemists partly referred to their lack of familiarity with using mathematical analysis in research. Modellers had aimed to create a model which would be mathematically tractable and produce results which could easily be interpreted to determine the distribution of control amongst reactions. In order to increase the mathematical tractability of the model they used a linear approximation of the non-linear behaviour of reactions and metabolic systems. This enabled

analytic solutions for equations describing changes in reaction behaviour over time (Heinrich and Rapoport, 1974b p94). The model involved two theorems, 1) the connectivity theorem for calculating the relationship between elasticity coefficients and control coefficients, and 2) the summation theorem, for calculating the distribution of control amongst reactions. The summation theorem stated that all the control coefficients in a system should add up to one, rendering the implications of the model's analysis for metabolic control easily interpretable. In response to experimentalists reporting confusion about how to implement the model David Fell and colleagues developed the matrix method. This involved an algorithmic procedure for calculating control coefficients without requiring full comprehension of the mathematics involved (Rankin Small and Fell, 1989; Fell and Sauro, 1985).

From the outset, modellers were very keen that the application of their model would be the analysis of actual experimental data sets pertaining to particular biochemical systems. However, as I previously mentioned, the model required that new experimental techniques be developed as the model could not be used in conjunction with existing types of experimental data. First, data needed to be collected from experiments on intact metabolic systems - previously experimental biochemistry had largely focused on investigating the properties of components in isolation. Second, data needed to be collected about the behaviour of reactions in isolation but in conditions which mimicked the *in vivo* state - previously experiments on isolated component had mainly been carried out in conditions which simplified the experimental procedure.

> 'It is, however, important that such determinations are not carried out at some arbitrary or traditional concentrations of substrates and products. Instead they must be held constant at precisely those steady state levels which obtain in the organism.'(Kacser and Burns, 1973 p70)

Additionally, the linear approximations upon which the model was based were only accurate over very small changes in concentrations. This meant that the experimental interventions had to involve very subtle manipulations and accurate recording of results (Fell, 1992 p319). The modellers had an expectation that experimenters would embrace the model and set about developing the new types of experiments required to obtain the relevant data. There were early indications that this would be problematic. Joseph Higgins, who developed an important precursor to the model in the early 1960's, later recalled experimenters chanting 'chain Higgins to

the (lab) bench' in response to the impracticality of the experiments he was suggesting (Higgins, 1990 p48).

The difficulties posed by carrying out the experiments, and the expectation that experimentalists would take the lead in surmounting them, also alienated biochemists from engaging with metabolic control analysis. During the 1980's several experimental techniques had been developed involving small collaborations between modellers and experimental biochemists. The symposium organisers wanted to use the workshop to showcase the range of experimental methods which had by now been developed in conjunction with metabolic control analysis and the metabolic systems to which they had been applied. The organisers thought that the fact that the model had not widely been used in conjunction with experimental data sets had caused biochemists to perceive metabolic control analysis as too general. They hoped that the presentation of experimental work at the symposium would 'dispel the idea that metabolic control analysis was an abstract subject with little relationship to "real" biochemistry.' (Cornish-Bowden and Cárdenas, 1990 pviii).

Discussion over the idealising assumptions upon which the model was based was one of the most significant areas of disagreement between experimentalists and modellers. It was the idealising assumptions, rather than the practical issues with the mathematics and experiments outlined above, which caused experimentalists most concern regarding the use of mathematical modelling in biochemistry. An opinion piece by the experimental biochemist Daniel Atkinson was circulated before the symposium. He aimed to highlight to modellers why experimenters were concerned about metabolic control analysis in the hope that they would directly address these worries in their presentations. His main concern was that metabolic control analysis involved general assumptions about metabolic systems which meant the results had little biological relevance. For instance, he took issue with the relationship between elasticity coefficients and control coefficients because of 'the assumption that all enzyme modulations are equivalent to changes in the amount of enzyme sharply differentiates that model, and also many others, from actual metabolic systems.'(Atkinson, 1990b p8). He felt that these types of idealising assumptions ironed over the details which experimental biochemists had spent years uncovering; instead the models were based on *apriori* assumptions which were of more relevance to mathematics than biochemistry (Atkinson, 1990a).

Modellers wrote brief responses to Atkinson's concerns. They mostly considered Atkinson's opinions to be extreme, pessimistic, and present an artificial dichotomy between the activities of experimentation and theorisation. Some indicated that they thought that 'metabolic control analysis already contains the features which Atkinson seeks' (Fell and Sauro response to Atkinson, 1990b p15). They felt that because metabolic control analysis had been constructed in order to analyse experimental data, and experimental data which was relevant to particular *in vivo* contexts with the specific aim of understanding metabolic control, the model should not be considered an abstract approach. Fell and Sauro argued that the level of detail Atkinson required was unnecessary and that all scientific work, including experimental biochemistry and mathematical modelling, required the use of simplifying assumptions (See Fell & Sauro response to Atkinson, 1990b p12). However, modellers presented metabolic control analysis as having a specific application, but still being a general model which could be applied 'to any linear, branched or cyclic succession of solute translocations and chemical transformations in any sub-cellular organelle, cell, tissue or organisms of any complexity.' (Porteous, 1983 p201). Atkinson was challenging the very idea that it was possible to construct a general mathematical model which was applicable to this diversity of metabolic contexts. The emphasis of modellers on the importance of experiments being carried out at the level of biochemical systems, and on isolated components in *in vivo* like conditions, did not address this fundamental concern. Unsurprisingly, Atkinson (1990a) was not satisfied with these responses and in his concluding remarks to the symposium state retained his initial position that metabolic control analysis had nothing to offer experimental biochemistry.

Experimentalists and modellers also clashed over their theoretical perspective on metabolic control. Experimentalists were reluctant to stop using the notion of a rate limiting step and adopt the systemic perspective offered by metabolic control analysis. These concerns intersected with debates about the assumptions used in the mathematical model, and the role of mathematical models in biochemical research. Experimental work on the rate limiting step had been based on a diverse range of experimental techniques and criteria for rate limiting status, and this diversity had led to confusion in identifying the rate limiting step in particular pathways (Fell, 1997). Those developing metabolic control analysis aimed to introduce a mathematical tool which could provide a general analysis of metabolic

control and overcome the disagreements stemming from this methodological diversity (Kacser and Burns, 1973; Heinrich and Rapoport, 1974b). Modellers wanted to provide a single quantitative definition of a reactions contribution to metabolic control and a method for its analysis. On the other hand, biochemists wanted to hold onto the complexity located in their range of experimental methods and diverse qualitative criteria for analysing the location of control in metabolic pathways.

5.3.2 Metabolic control analysis and biochemical systems theory

> 'The letter by Kacser is symptomatic of the problems that for 20 years have plagued the development of new approaches to the quantitative analysis of complex biochemical systems. Individuals who enter the field reading only Kacser's work receive a very limited perspective, because he refuses to acknowledge the more general framework for understanding complex biochemical systems that has been established by the BST approach. It is a minor tragedy, when one reflects upon on the progress that might have been made had these individuals been working at the forefront of their field or been teaching others the full scope of the subject.' (Savageau, 1992 p135)

At around the same time, hostile interactions also emerged between researchers involved in metabolic control analysis, and what came to be seen as a rival modelling approach, biochemical systems theory. Although initial work on metabolic control analysis and biochemical systems theory had been carried out in the late 1960's and early 1970's, there are no published records relating to direct interactions between the two groups until the late 1980's. In 1987 Kacser and Porteous published an article in *Trends in Biochemical Sciences*, entitled 'Control of metabolism: what do we have to measure?', which addressed the concerns of biochemists about the abstractness of metabolic control analysis and the feasibility of the associated experiments. In this article they state ' … we should not start to set up an algebraic model incorporating *ad hoc* assumptions or a priori assertions.' (Kacser and Porteous, 1987b p7). This remark was not explicitly directed at Savageau and biochemical systems theory but appears to have acted as the spark igniting years of hostile interactions between the two groups. Savageau responded angrily directly attacking Kacser and Porteous; 'Many of their assumptions are of questionable validity, others are at odds with current knowledge in the field.'

(Savageau, 1987 p219). Over the next five years a fiery debate ensued between the two groups over the role of mathematical modelling in biochemistry and the assumptions used in model building.

Michael Savageau (1969a) began developing biochemical systems theory in the late 1960's. Around this time there was growing interest in the behaviour of intact metabolic systems. Savageau's goal was to build a model which could be used to conduct simulated experiments investigating systemic biochemical behaviour which could not be conducted using existing laboratory techniques. In addition, as I explored in chapter three, the model was a response to constraints on model building arising from limited kinetic data availability and mathematical intractability of highly detailed models. Researchers involved in biochemical systems theory thought that mathematical modellers should share the goal of producing a single unified model which could be used in conjunction with multiple different research questions about the systemic properties of metabolism (Irvine, 1990 p140). Savageau's major concern was that the researchers in the field needed to come together and focus their efforts on developing a single general mathematical model which would be a useful tool in multiple research contexts. He regarded the diversity of approaches as leading to 'confusion and error' and inhibiting progress with the contribution which mathematical modelling could make to biochemistry (Savageau, 1987).

> 'Such open and constructive competition also tends to unify the field, because participants are engaged in mutual criticism and learning. The entire discipline advances with a minimum confusion, and, perhaps more importantly, new entrants to the field find it easier to identify important issues and make meaningful contributions. It is regrettable that this healthy pattern has not characterised the scientific exchange in our sub discipline.' (Savageau, 1992 p136)

Savageau was committed to the idea that biochemical systems theory was the best approach for achieving this goal and the mathematical model around which the researchers should focus their efforts. He felt threatened by the higher profile of metabolic control analysis and between 1987 and 1992 was involved in at least ten full length journal articles and numerous shorter pieces trying to convince researchers that biochemical systems theory was a more rigorous and useful approach than metabolic control analysis (e.g. Savageau, 1991; Savageau et al., 1987a).

Kacser and colleagues intention for metabolic control analysis obviously differed significantly from Savageau's intention for biochemical systems theory. They wanted to develop a mathematical tool which experimentalists could use to interpret two particular sets of biochemical data to inform understanding of the distribution of control in metabolic systems. The aim of metabolic control analysis was limited to a single research question, and the interpretation of systems level experimental data, rather than multiple research questions and using simulations to investigate the behaviour of biochemical systems rather than experiments. Early on in the debate Kascer and colleagues pointed out these differences and argued that the two approaches should not be conflated:

> 'Control analysis arose out of experimental necessity to deal with a number of problems and apparent paradoxes rather than from a desire to erect an all-embracing general theory which Savageau aims at.' (Kacser and Porteous, 1987a p223)

Unlike Savageau they were happy for there to be diversity in approaches to mathematical modelling in biochemistry. Researchers involved in metabolic control analysis usually assumed a defensive position responding to attacks from those involved in biochemical systems theory. Remember that, metabolic control analysts were simultaneously trying to convince experimental biochemists that there model was not a general abstract approach but directly informed by experimentally derived properties of metabolic systems. In this context, Savageau's arguments that modellers should coalesce around a single and extremely general modelling approach were perceived as very threatening. Kacser and colleagues sought to distance themselves from the more abstract type of model Savageau was working on (Kacser, 1991).

The articles published by Savageau and colleagues not only argued that biochemical systems theory was a superior strategy for building general mathematical model of metabolic systems, but that metabolic control analysis was in fact an inferior version of the same model and for this reason should be subsumed within the framework of biochemical systems theory (Savageau et al., 1987a). Savageau (1987a) even implied that metabolic control analysis had been developed from his initial publications on biochemical systems theory and was angry that those involved in developing it had not acknowledged this earlier foundation. The articles

presented three main reasons presented for why biochemical systems theory was the superior 'version' of the model focusing on differences in the idealising assumptions the models were based upon.

Firstly, biochemical systems theory used kinetic orders, or rate constants, providing a general description of reaction kinetics over a wide range of changes in substrate and effector concentrations. Metabolic control analysis, on the other hand, was based upon sensitivity coefficients which described the response of reactions to changes in the concentration of effectors and substrates over a very small range of concentrations. Savageau (1987a) considered sensitivity coefficients to be an ad hoc version of the more rigorously structured kinetic orders, as such a model based on kinetic orders would be a better candidate for building general models of metabolic systems which could be used to simulate a wide range of metabolic behaviour.

> '… one of the principle advantages of the Power-Law Formalism recognized very early was that one need not be engulfed by detailed kinetic analysis; one could make a few measurements after small deviations under *in vivo* conditions and determine the key parameters in the formalism.' (Savageau et al., 1987a p135)[10]

Kinetic orders were based on the assumption that reaction behaviour could be accurately estimated from a small number of experimental data points, whereas sensitivity coefficients were based on the assumption that a detailed experimental investigation of reaction behaviour in a particular biological context was required:

Secondly, biochemical systems theory and metabolic control analysis were based on different idealising assumptions about the temporal organisation of biochemical processes. Both models involved a description of how reaction rate is affected by changes in the concentration of associated molecules, which differed from standard Michaelis-Menten kinetics (See chapters 2 and 3 of this thesis). This aspect of reaction behaviour is non-linear but metabolic control analysis was based on a linear approximation of this behaviour which meant the approximation would only be relevant across a limited range of changes in substrate or effector concentration. In contrast, the power law which Savageau had created and based

[10] In the quote Savageau also mentions the importance of collecting data in *in vivo* like conditions. However, this point was of minor concern for biochemical systems theorists who focused on using assumptions which helped them minimise the amount of experimental data required by the model. Whereas a corresponding experimental program played a central part in metabolic control analysis.

biochemical systems theory upon was a non-linear approximation. He argued that is approximation method had been experimentally validated to be accurate over a wide range of conditions. Again, this supported the case that biochemical systems theory was the superior approach for building general models of metabolic systems which could be used to explore a diverse range of research questions (Voit and Savageau, 1987).

Thirdly, Savageau and colleagues also queried the central role which the connectivity theorem and the summation theorem played in metabolic control analysis. They argued that whilst they can also be shown to be present in biochemical systems theory they are irrelevant and '… have no operational role in BST. They are only corollaries in BST, and need never be made manifest during the development of the theory for the purposes of application.' (Savageau et al., 1987b p165).

Athel Cornish-Bowden (1989), who acted as a mediator, responded to Savageau's papers by explaining the significance of the differences between the two approaches. His main point was that metabolic control analysis and biochemical systems theory were intended to serve very different roles in research and as such the independence of each model should be retained:

> "The primary objective in metabolic control theory has been to assign clear meanings to some of the concepts used vaguely and inconsistently in earlier discussions of metabolic control, to analyse the theoretical properties of the quantities thus defined, and to use the analysis to understand why real systems behave in the way they do. … The objectives of biochemical systems theory, to develop mathematically tractable models of metabolic systems that mimic the behaviour of real systems and allow one to predict how they will behave in new circumstances, appear to be quite different.' (Cornish-Bowden, 1989 p367)

Cornish-Bowden pointed out that metabolic control analysis was intended to analyse the distribution of control amongst reactions in particular metabolic systems and biochemical systems theory was intended to simulate a variety of different behaviours in a general model. If this was appreciated, then the differences in assumptions and content of the models could be seen as appropriate. For instance, given the desire to simulate general behaviour, kinetic order was the appropriate basis of biochemical systems theory, and given the desire to analyse the distribution

of control in particular conditions, sensitivity coefficients were the appropriate basis for metabolic control analysis. Also whereas, within the general aims of biochemical systems theory the central role of the summation and connectivity theorems made no sense, within the limited aims of metabolic control analysis to calculate the distribution of metabolic control they made complete sense.

Cornish-Bowden's arguments apparently did not get through to Savageau and colleagues, who continued to publish articles reiterating the same opinions (Sorribas and Savageau, 1989a; Sorribas and Savageau, 1989c; Sorribas and Savageau, 1989b; Savageau and Sorribas, 1989; Savageau, 1990; Savageau, 1991; Savageau, 1992). These articles largely repeated arguments put forward by Savageau and colleagues in three papers published in 1987 (Savageau et al., 1987a; Savageau et al., 1987b; Voit and Savageau, 1987). Many of these papers were highly technical, making comparisons between the capacity of biochemical systems theory and metabolic control analysis to constitute extremely general models of biochemical systems. Researchers developing metabolic control analysis were less active in this debate, but they published one paper taking a similar comparative approach. They compared the usability of metabolic control analysis and biochemical systems theory for experimental biochemists. This paper concluded that biochemical systems theory was inaccessible to experimentalists '…due to the complexity of the mathematical formulation…' (Groen and Westerhoff, 1990 p117).

5.3.3 Research commitments and Hostile interactions

In these two case studies from early biochemical systems biology, the commitments of all three groups, experimenters, modellers working on metabolic control analysis, and modellers working on biochemical systems theory, are so rigid and specific that they provide firm guidance on acceptable research practices. Modellers working on metabolic control analysis maintained four commitments; 1) developing a systemic perspective on metabolic control, 2) developing a general quantitative approach to interpreting experimental data relating to metabolic control in any metabolic system, 3) using the model in conjunction with experimental data which were collected in conditions which replicated the *in vivo* context and 4) that metabolic control analysis was the appropriate modelling approach to achieve these aims. Experimental biochemists were committed to: 1) work which paid detailed

attention to the properties of particular biochemical systems, 2) the value of qualitative descriptions, 3) maintaining methodological and explanatory diversity to reflect the complexity of metabolic systems, 4) the theory of the rate limiting step. Modellers involved in biochemical systems theory upheld commitments to: 1) developing a general approach to modelling biochemical systems, 2) the field of biochemical mathematical modelling should unify around one approach, 3) The use of kinetic orders and non-linear approximations in order to achieve this generality and reduce the amount of experimental data required by the model, and 4) that biochemical systems theory was the general approach around which the field should unify. In each case, these sets of commitments are rigidly held by the different groups during the period examined and include commitments to specific methods and explanations. There are stark differences between research groups in commitments to, qualitative vs. quantitative analysis, mathematical vs. experimental approaches, specific vs. general models, and diversity or unity in research methods.

As the epigraphs at the start of each of the previous two sections illustrate, the groups are intolerant of these differences. Each of the groups is committed to the superiority of their commitments not only for researchers within their group, but also as the best set of research practices for other groups of researchers. Modellers involved in metabolic control analysis are of the opinion that experimentalists should adopt their systemic perspective on metabolic control and metabolic control analysis as a quantitative research tool. They also believe that their approach is superior to biochemical systems theory in relation to the criterion of being a mathematical model which is more accessible to experimental biochemists. Savageau displayed extreme intolerance of differences between biochemical systems theory and metabolic control analysis. He attempted to completely unify the two models and eliminate any differences in meaning between the two sets of terminology; for example, his attempt to argue that the sensitivity coefficients of metabolic control analysis are in fact inferior versions of the kinetic orders used in biochemical systems theory. This is a very different situation than Galison describes in high energy physics, where different cultures are happy to work together and develop modes of collaborative interactions without insisting upon standardised terminology and shared meaning.

These instances where different groups of modellers and experimenters hold rigid and specific commitments which heavily constrain research practices, and are

intolerant of inter-cultural differences, intersect with the occurrence of hostile interactions. The groups perceive intercultural-differences in research commitments as a potential threat to their set of research practices. The hostility often appears to be driven by disputes over access to economic and social resources, including the material and financial resources and social status required to maintain and expand research programs. Given that these resources for mathematical approaches to biochemistry during this historical period the different groups are compelled to enter into often heated competition. This resulted in groups attacking other styles of working and defending their own approach. Metabolic control analysts regarded the inferior commitments of experimentalists as the reason for lack of engagement of the community with their model. They found this threatening, as they believed that achieving their aim of providing a quantitative approach for analysing biochemical data on metabolic control required the involvement of experimental biochemists. Savageau considered metabolic control analysis to be undermining interest in biochemical systems theory. He responded by arguing that biochemical systems theory was the superior modelling approach and attacking metabolic control analysis for not living up to the same set of standards. In the section 4.5 of this chapter I will discuss the role that this particular configuration of rigid and intolerant commitments has on the problem space which researchers are operating in and the corresponding research dynamic.

Galison uses his work on collaboration in high-energy physics as an alternative position to Thomas Kuhn's notion of the structure and dynamics of scientific research. Kuhn proposes that science consists of internally coherent paradigms which have clear boundaries with other paradigms - existing either synchronically or diachronically - across which nothing can pass. This organisation and interactions are associated with paradigm shifts, clear breaks between one paradigm and another. Galison argues that in the case of high energy physics research has a heterogeneous structure, it is carried out by groups of experimenters, instrumentalists, and theorists each characterised by their own sets of commitments, methods, research questions. These groups can exchange things which each other (See section 5.2). As a consequence of this structure and interactions the groups can undergo non-synchronous changes meaning there aren't clear cut boundaries between research paradigms. The work of Kuhn on paradigms, and Lakatos on

research programs, already provides accounts of competitive or hostile interactions between groups of researchers who do not engage in closed or open collaborations. However, I have chosen not to use this work to frame my account of hostile interactions because of the commitments it would entail to a particular type of explanation for this failure of collaboration.

In this chapter I have looked at how differences in the content and quality of researchers commitments intersect with the way in which they interact with other groups of researchers. However, I have not looked at why researchers hold rigid or flexible commitments and why they view different groups in a negative or positive light. Framing hostile interactions in a Lakatosian light would involve constructing an explanation in terms of their 'logic of discovery', the relationship between internal aspects of the research program - the core, negative heuristics, and positive heuristics.

> 'But rational reconstruction or internal history is primary, external history only secondary, since the most important problems of external history are defined by internal history. … the rational aspect of scientific growth is fully accounted for by one's logic of scientific discovery.'(Lakatos, 1978 p118)

However, the question of the relevance of internal and external factors in accounting for the style of interactions is left open by this chapter as an area for extension. Recent work on the increase in collaboration in data intensive science (Rajan and Leonelli, 2013; Davies et al., 2013) has highlighted the significance of political and social factors for contemporary epistemic developments. This makes it seem likely that, in the cases examined here, the wider social and cultural context influencing things such as resource availability would play a significant role in accounting for researchers disposition towards particular modes in interacting with other groups.

5.4 Interactions between experimenters and modellers in current biochemical systems biology

In a recent series of papers, Miles MacLeod and Nancy Nersessian (2013b; 2013a; 2013c) analysed collaborative interactions between modellers and experimenters in current biochemical systems biology. This work comes out of a much larger four year ethnographic investigation into two 'pioneering' research labs.

The investigation involved attending lab meetings, journal clubs etc. and carrying out semi-structured and unstructured interviews, mainly with modellers from the labs but also with their experimental collaborators. In these papers MacLeod and Nersessian explore a range of philosophical issues, from the factors which account for on-going cognitive innovation by researchers in the two labs, to the role of theory in model building in biochemical systems biology. Researchers based in both labs primarily have a background in engineering and mathematical model building but have different modes of interacting with experimental biochemical research. Researchers in one of the labs, Lab G, are only involved in developing models and work on modelling problems stemming from collaborations with experimental labs – these range from biofuels to Parkinsons disease. MacLeod and Nersessian refer to this as the unimodal strategy. Researches in the other laboratory, Lab C, are involved in developing mathematical models and carrying out experiments and are working on the specific problem of 'understanding cell signalling dynamics in a reduction-oxidation (redox) environment' (MacLeod and Nersessian, 2013c p36). Additionally they collaborate with external experimental biochemists. This is referred to as the bimodal strategy. In two papers MacLeod and Nersessian give a detailed case study of an individual PhD researcher from each of the labs, and I will use an abridged version of these to explore the relationship between researcher's commitments and collaborative interactions. MacLeod and Nersessian have focused on these individuals partly because their research experience is representative of many of the individuals in the laboratories.

5.4.1 The unimodal strategy

The case from LabG gives an account of the researcher G12's work on 'the relation between oxidative stress and the generation of monocytes implicated in the generation of plaques on the vascular smooth muscle of blood vessel walls.' (MacLeod and Nersessian, 2013a p540). G12's research involved three main interrelated modelling tasks. Firstly, the construction of the biochemical pathway. This task began with the pathway provided to G12 by her experimental collaborator. G12's aim was to produce a pathway capable of accounting for the major dynamics of the network and it was immediately apparent that the pathway reconstruction

provided by her collaborator was insufficient. This was attributed to the different research goals of modellers and experimenters:

> ‘The focus of her collaborator on this project, as a molecular biologist, had been on assembling the main direct chain of causal influence, not on what other contributing factors would be affecting the systems dynamic behaviour, which was necessary for G12’s task.’(MacLeod and Nersessian, 2013a p541)

In reconstructing the pathway G12’s work was shaped by trying to work with the available data and attempting to build a model which would eventually be mathematically simple enough once all the parameters had been added. During her reconstruction, G12 used biochemical literature and her experimental collaborator for further input and verification of the model and adjusted her model according to her collaborators advice (MacLeod and Nersessian, 2013a p543). However, in order to fulfil her aim of reconstructing a pathway which would account for the dominant dynamics of the network she occasionally inferred interaction for which there was no direct experimental evidence.

G12’s second modelling task involved the construction of a mathematical representation of the pathway. This involved a range of strategies for dealing with lack of relevant experimental data and making appropriate simplifications. She black boxed components if insufficient data was available about their individual dynamics. G12 also off-lined components, excluding them from the dynamics of the system, either because of lack of information, or because including it was deemed too complex. These decisions were again both informed by experimental literature and collaborators, but also through G12 carrying out guess work and experimentation with the model. Sometimes G12 would even directly contradict the literature in her modelling assumptions. In addition to deciding which components and interactions, and what level of detail, to include in the mathematical model, G12 also had to decide what type of mathematical framework to use to represent the system. She chose ordinary differential equations to build a non-spatial representation of the system. This is the most common type of mathematical framework used by modellers in the lab, due to its relative conceptual simplicity and ability to account for the dynamics of the system. G12 used a mix of two different approaches to constructing ODE’s depending on her objective and the available data. Firstly, the power law formalism developed by Savageau in biochemical systems theory. This

enabled dynamic behaviour to be approximated from limited biochemical data. MacLeod and Nersessian emphasise that:

> 'BST and its templates are not starting points of the modelling process … They are chosen, if suitable, during the model-building process (...) and modified as required. The framework, in other words, is added to the nest in the model- building process, if it serves that process given the available data and the aims of the modeller.' (MacLeod and Nersessian, 2013a p552)

Secondly, the PhD student usedrate laws, such as Michaelis Menten kinetics, which require larger biochemical data sets and give a more detailed representation of biochemical dynamics.

The third modelling task involved setting the parameter values for the systems dynamics. MacLeod and Nersessian highlight this as being a model building task involving a high degree of ingenuity. Time series or rate law data from which to determine the parameter values was not available. G12 used a variety of different strategies including the use of time series data referring to the same molecules but from different cell lines. Simplifying assumptions were made in order to reduce and simplify the parameter values required - steady state assumptions were applied to fast reactions in the system, and linearization assumptions were applied to some non-linear relationships. The most significant tool used by G12 in order to determine the missing parameter values was to run Monte Carlo simulations in order to determine which parameter values would give rise to the dynamic behaviour observed in the system. In order to run these simulations the parameter space of the model needed to be significantly reduced by making further simplifying assumptions about the significance and behaviour of systems components and interactions.

5.4.2 The bimodal strategy

MacLeod and Nersessian's case study from Lab C describes the PhD research of C9 who is investigating the 'different sensitivities in cancer cell lines to chemotherapy drugs, specifically to what we call drug X.' (MacLeod and Nersessian, 2013b p4)[11]. A clinical researcher at a medical school highlighted this issue to the director of Lab C who hypothesised that the differential sensitivity was related to

[11] MacLeod and Nersessian's refer to the specific molecules involved in C9's project by alternative names in order to keep her identity anonymous.

differences in the cell signalling functions of Reactive Oxygen Species – a group of reactive molecules containing oxygen. C9's PhD involved constructing four models to investigate this. Whilst C9 carried out all the modelling and biochemical experiments for the project herself she regularly discussed her project with the director of Lab C and with a senior biochemist from a different institution.

The first model involved C9 building a mathematical model of a single pathway, pathway Y, and the dynamics of its interactions. She did this by searching through existing literature to reconstruct the topology of the pathway and determine the parameters for the dynamics of the interactions. The ODE model seemed to accurately predict pathway behaviour when compared against experimental results found in the literature. At this point she was encouraged by the senior biochemist who was advising her to 'shift their attention from the small Y model to the whole system of redox regulation itself.' (MacLeod and Nersessian, 2013b p4), in order to have a richer understanding, in particular of the environmental factors, affecting the Y pathway. C9 followed up on this suggestion and constructed a much larger ODE model again through searching the available literature. Again this model successfully simulated the behaviour of the system in accordance with experimental results found in the literature. For both of these models the structure of the Y pathway and the redox regulation system was well known and C9's major task was compiling the relevant parameter values.

C9 then moved on to investigating the impact of drug X on the system. However, there was relatively little existing experimental information about X and C9 decided to carry out the necessary experimental work herself. She obtained two cell lines from the clinical researcher at the medical school, one from a patient who was insensitive to drug X and one from a patient who was sensitive to drug X. C9 was expecting the cell which was insensitive to drug X to have a high activation of the Y pathway. However, her experimental results indicated that in fact the opposite was happening - this contradicted her working hypothesis and the pre-existing experimental literature. Further experimentation led the Lab C director to suggest they build a third model looking in more detail at the production of redox oxygen species by drug X in the cell. Previously, this had been black boxed in the second model and the same estimates of a series of reactions had been used to model both the insensitive and sensitive cell types. Now C9 focused on looking at differences in this particular part of the pathway between the two cell types. During this process C9

iteratively performed experiments and modelled and simulated hypothetical mechanisms. Once the model was constructed C9 carried out simulated perturbations and experimental perturbations and compared the results in order to verify the model. This was successful and enabled her to claim that she had discovered the mechanism responsible for the different sensitivity of cells to drug X. What's more it had implications for a clinical application capable of predicting which patients cancer would respond to drug X and which would not. Reviewers of the corresponding article criticised the levels of drug X they had been using as they were much higher than those used clinically. C9 ran laboratory experiments with lower level of drug X which produced remarkably different behaviour than the higher levels initially used. Amazingly enough this different behaviour was replicated in model simulations based upon lower levels of the drug providing further validation for her proposed mechanism.

In reflecting upon her motivation for using this approach C9 highlights two major benefits. Firstly, it overcomes the problem of creating a theoretical model and then waiting to see if is taken up by biochemists and applied in an experimental context:

> 'In comparison with these "theoretical modellers", she explained: "we don't just come up with ideas and then just shoot them out there and wait for people to do them" (MacLeod and Nersessian, 2013b p9)

Secondly, that the bimodal strategy is useful for circumnavigating communication issues that may arise if she was trying to explain to experimental biochemists the experiments she wanted them to carry out. However, the article also highlights downsides of pursuing this approach, mainly that the level of expertise developed in modelling and experimentation will be lower than by a researcher who is only involved in using one set of techniques.

5.4.3 Research commitments and open collaboration

In the two case studies from current biochemical systems biology, I argue that researcher's commitments still act as constraints on research but in a different way than in Galison's account. The commitments of G12 and C9 have the characteristic of being flexible and sometimes broad. This results in a situation where researchers encounter flexible constraints. Both G12 and C9 display a sustained commitment to

a particular research question. However, they do not display rigid and specific commitments to any methodological approach for addressing it. G12 has a broad methodological commitment to developing a dynamic mathematical model and C9 has a broad commitment to iteratively using modelling and experimental techniques.

> 'The most important feature of practice in our ISB labs is that the well-structured task environments that characterize established sciences like molecular biology or bioinformatics do not exist. … Little can be outsourced to routines or protocols.' (MacLeod and Nersessian, 2013c p39)

G12 and C9 do not adhere to a consistent set of commitments defining acceptable methodological practices. Instead MacLeod and Nersessian describe a situation where researchers respond to the multiple constraints experienced in research by being flexible in their approaches to working around and within these constraints. For instance, during G12's third modelling task she experimented with a wide range of different strategies to deal with insufficiencies in available parameter value data - from using data sets pertaining to different biochemical systems to running monte carlo simulations. The commitments of researchers are flexible, and this in turn makes the constraints the researchers are working within flexible as well:

> 'Adapting a problem to one that can be solved from one that cannot is the central function of an adaptive problem solving environment. Not only are methods and data transformed but also how the problem is understood and represented, until a coherent solution can be reached.' (MacLeod and Nersessian, 2013c p47)

The constraints encountered by a researcher will partially arise from their commitments to particular methods. For instance, certain modelling approaches will demand specific biochemical data sets - the availability of which may constitute a limitation on research. If researcher's methodological commitments are flexible the constraints they encounter will also be adjusted until a path of least resistance is found and a solution achieved.

In addition to working with flexible and broad constraints, G12 and C9 also regard many inter-cultural differences between modellers and experimenters in a positive light as a potential source of valuable resources for their own research. G12 commented that 'she always felt like she started "from zero" with every new modelling task on a different biological system.' (MacLeod and Nersessian, 2013a

p541) and was happy to seek biochemists advice whilst reconstructing the biochemical pathway. In addition G12 was not tied to a particular mathematical modelling strategy such as biochemical systems theory, but used aspects of multiple modelling approaches as and when she thought appropriate. This is remarkably different from the situation in the case study from early biochemical systems biology when Savageau was aggressively committed to the notion that biochemical systems theory should be the only strategy used to model biochemical systems. C9 was happy to accept suggestions from the senior experimental biochemist about how she should move her research project. His suggestion that she should now focus on the small model within the context of the whole redox system was motivated by his stronger commitment to the importance of understanding the role of the pathway within the broader physiological context. Her willingness to accept suggestions from a senior team member may reflect her junior status alongside the influence of a wider research culture which values collaboration and flexibility.

This context of flexible and broad commitments, and perception of inter-cultural differences in methodological and theoretical expertise as valuable, intersects with what I term open collaborative interactions. What I mean by this is that during interactions between groups of modellers and experimenters the modellers are open to allowing their research practices to be influenced by the commitments of the other group. This is most noticeable in C9's bimodal research strategy. In C9's research there is a significant emphasis on being open to learning from the methods and knowledge of experimental biochemists to the extent that she performs many biochemical experiments herself. What is noticeable about C9's research process is that it involves being open to new theories for the action of drug X, constructing different scales and types of mathematical models, and conducting biochemical experiments as and when appropriate. MacLeod and Nersessian argue that 'one set of skills that are universally important for ISB researchers to cultivate are the requirements of *cognitive flexibility* and *epistemic pragmatism*.' (MacLeod and Nersessian, 2013c p44). By this they mean that in order to be successful researchers are open to the possibility of new ways of doing things which may often challenge their established practices and research commitments. This situation is remarkably different from my interpretation of Galison's account where although researchers are tolerant of intercultural differences they remain closed to letting

those differences significantly affect their internal research practices. It is also remarkably different from early biochemical systems biology where differences are regarded as threatening.

5.5 Three different problem spaces

The idea that scientists manage their task of investigating complex systems through operating in constrained problem spaces is a frequent theme in philosophy of science (Simon, 1962; Rheinberger, 1997b). I use the comparative work carried out in this chapter to highlight three different types of problem spaces, each of which is related to a distinct research dynamic. MacLeod and Nersessian describe researchers as occupying an adaptive problem space which facilitates ongoing creativity and innovation. Galison describes a dynamic of extended periods of continuity interspersed with shorter bursts of change which I claim intersects with researchers operating in a closed problem space. I will argue that research in my case studies from early biochemical systems biology involves the externalisation of aspects of the problem space. My account brings to light how the quality of the problem spaces that researchers are operating are not just affected by the internal commitments of the group, but also how those internal commitments affect their interactions with external groups of researchers.

Galison describes researchers as operating under rigid and specific commitments which create a research environment which is highly structured by constraints. I argue that these continuous commitments constitute a closed problem space. Galison highlights that these constraints are a valuable asset to researchers, ‘constraints are constitutive of the positive research program. They create a problem domain, giving it shape, structure and direction.’ (Galison, 1995 p22). Through delimiting acceptable research practices and explanations constraints provide researchers with important boundaries for their work giving it shape and meaning. Galison describes these problem spaces as existing largely unchanged over extended periods of time. Part of what enables these spaces to persist unchanged is the fact that they remain immune to the influence of the commitments of external research groups even during interactions and exchanges with those groups. The problem spaces are only defined by the internal commitments of the group, they are closed to any external influence. For Galison, the extended duration of these

problem spaces and their internal definition is what enables creativity to take place. As researchers operate within the same set of constraints, it allows them to become familiar with them, develop new ways of working within them, and explore them to their limits. This period of continuity allows them to distinguish the significant from the insignificant and allows novelty in methods and explanations to emerge as they stand out against this clearly defined background.

This situation is remarkably different from that described by MacLeod and Nersessian for groups of experimenters and modellers in current systems biology. MacLeod and Nersessian describe researchers as operating in *'adaptive problem spaces* defined by multidimensional problem–solving tasks and emergent approaches.' (MacLeod and Nersessian, 2013c p36). Research is still guided by commitments particular to a group of modellers or experimenters, but they view these constraints pragmatically, if altering and updating the commitments helps move research on a particular question forward then they are open to doing so. As research commitments change the constraints encountered during research exhibit a corresponding degree of flexibility allowing new methods and explanations to emerge. An important process shaping this shifting landscape is interactions with research groups with different research practices and explanatory knowledge. As groups of experimenters and modellers come into contact with each other they are open to these interactions changing their commitments and integrate aspects of other group's expertise as and when appropriate. They view these as potential resources they can use to address constraints and work on their research question. These processes contribute to a high level of ongoing innovation in research practices and explanations. In an adaptive problem space it is not the continuity of commitments but the flexibility of commitments which is associated with creativity and innovation. Directly comparing her work to Galison's, Nersessian states that, 'Although the central metaphor of a trading zone is exchange, the central metaphor of an adaptive space is *emergence*.' (Nersessian and Newstetter, Forthcoming p716)

The situation in early biochemical systems biology illustrates yet another relationship between commitments, interactions between experimenters and modellers, and the type of problem space researchers occupy. To some extent modellers in early biochemical systems biology are also occupying the type of

internally defined closed problem space which Galison describes. I examined these aspects of metabolic control analysis and biochemical systems theory in the third and fourth chapters of this thesis. For instance, biochemical systems theory was a creative response to the constraints of limited availability of kinetic data and building tractable mathematical models. However, both groups of researchers also went through a process of externalising aspects of their problem space onto other groups of researchers. Metabolic control analysts encountered problems with how to produce the experimental data required by the model and expand the use of their model in experimental research. Michael Savageau encountered the problem of how to intensify the development of biochemical systems theory. Both groups held an external group of researchers responsible for these problems, metabolic control analysts blamed experimental biochemists, and biochemical systems theorists blamed metabolic control analysts. The researchers did not respond to these problems by changing their commitments such that the problem is rephrased and the constraints altered. For instance metabolic control analysts could have reframed their commitment to providing experimental biochemists with a quantitative tool, to a commitment to developing a bimodal approach to research and producing the experimental data themselves. In Macleod and Nersessian's case study C9 openly acknowledges that this bimodal strategy is a way around lack of uptake of mathematical models by experimental researchers. Researchers also did not respond to these problems by engaging with the concerns of the other group and being open to amending their research practices. In MacLeod and Nersessian's account of G12's work we saw how the commitment to a single general model has been relaxed, biochemical systems theory is now used as part of a repertoire of mathematical tools.

I argue that in the two case studies from early biochemical systems biology, researchers respond to these problems by externalising the problem space. What I mean by this is that they focus on the commitments of an external group of researchers as the constraint or problem they are encountering. This leads them to see the solution to the problem as being dependant on the external group of researchers changing their research commitments. This often involves a group of researchers believing that their commitments are superior and an external group of researcher's commitments inferior. For instance, Savageau believed that

biochemical systems theory was a superior approach to building general mathematical models of biochemical systems. He set about attacking modellers who were committed to metabolic control analysis as a quantitative tool for understanding the behaviour of biochemical systems. Solving the problem of getting more modellers interested involved in biochemical systems theory involved getting this external group of researchers to change their commitments to be in line with his. This process of externalising aspects of the problem space is associated with a period of stagnation in research. It means that group a researchers becomes dependant on the movements of an external group for addressing the problem and moving research forward.

5.6 Concluding Remarks

It appears that experimenters and modellers in current biochemical systems biology have managed to develop more epistemically productive modes of interacting with each other than their counterparts in early biochemical systems biology[12]. The impetus for engagement between experimental biochemists and different groups of mathematical modellers was present in the earlier period of biochemical systems biology. However, different groups of researchers struggled to engage collaboratively and interactions were often hostile. Researchers clung to rigid and specific commitments to particular methods and explanations and frequently exhibited intolerance of other ways of doing things. This led to researchers occupying attacking and defensive positions as they externalised their problems stagnating research. Different groups of researchers in current biochemical systems biology have found ways of avoiding debilitating hostility. They often manage to establish productive open collaborative interactions which can bring multiple sets of knowledge to bear on complex research problems. Researchers operate in adaptive problem spaces in which researchers commitments, and correspondingly the constraints they encounter, are modified until an optimal solution is found. This is connected to a dynamic of ongoing innovation in methods and explanations. However, this is not a uniform picture across current biochemical systems biology.

[12] An interesting extension to this chapters focus on the relationship between research commitments, social interactions, and problem spaces, would be to look at the three different formations analysed in this chapter through Helen Longino's account of 'a normative theory of social knowledge' (Longino, 2002 p129).

Distinct modes of research can be identified across which little collaborative interaction occurs. For instance, researchers engaged in stoichiometric and kinetic approaches to metabolic modelling, sometimes referred to as top-down and bottom up approaches, largely operate in distinct research spaces with notable opposition in the research commitments of each group (Krohs and Callebaut, 2007; Jamshidi and Palsson, 2010; Heinemann and Sauer, 2011).

Much of the interest in the increase in collaboration in the life sciences is motivated by trying to understand if and how research practices have changed over the past few decades. Indeed, MacLeod and Nersessian (2013c) use their work to argue that the dynamic style of research they observe in current systems biology is not just a moment of flux in a period of transition, but a mode of research which will be sustained over time. There are striking continuities in some aspects of the major general constraints biochemical systems biologist's face. Researchers in both periods struggle with the availability of biochemical data, particularly kinetic data. [13] They also both come up against the problem of how to appropriately simplify mathematical models in order to make them tractable and interpretable. However, the comparative work in this chapter indicates that researchers respond to these constraints in different ways. In the examples from early biochemical systems biology, due to the rigidity of researchers' commitments and their intolerance of differences with other groups, constraints are more continuous. In the case studies from current biochemical systems biology, due to the flexibility of researchers' commitments and their positive perception of differences with other groups, the exact nature of constraints is continually shifting. What I have not done is look at the wider contextual reasons for why researchers in different circumstances may ascribe to different kinds of commitments and what causes the qualities of those commitments to change. The wider cultural context in which biochemical systems biologists are operating has shifted significantly over the last thirty years, in terms of funding, technology, available data, institutional organisation (See Rajan and Leonelli, 2013; Davies et al., 2013). The impact of these shifts in the cultural status of biochemical systems biology on researchers' inclination towards hostility and intolerance, or

[13] The systems biologists Hans Westerhoff and Berhand Palsson state that 'Unknown to many, the 'pre-online PDF' era contains answers to many of the challenges and pitfalls facing the field.' (Westerhoff and Palsson, 2004 p1252). However, as this and the next chapter in this thesis illustrate, in the current context researchers also have to develop new approaches to effectively addressing the challenges of interdisciplinary research.

openness and interest, regarding inter-group differences in research commitments is an interesting area for further research.

6 Reconstructing metabolic networks: Mathematical modelling meets big data

Abstract

Over the past two decades mathematical modelling has become a major tool for analysing large scale omics data sets. In this chapter I analyse the impact of changes in biological data on the epistemic role of mathematical modelling using a detailed historical case study of the development of constraint based optimisation in systems biology. Central to my analysis is the observation that the data landscape modellers are incorporating into their research involves two types of broader changes within biology: 1) Changes in the data available to researchers, 2) Changes in researchers perceptions of the evidential value of that data. Initially genomics data was received into a wider epistemic context which facilitated its role in stimulating a dramatic increase in the scale of research surrounding constraint based optimisation. However, after an initial expansion in the volume of research, significant alterations in the ontological assumptions of modellers and the level of analysis they aspired to took place. In this new epistemic context the significance and impact of genomics data was revised. Metabolic network reconstruction and analysis moved from assumptions of physico-chemical determination and analysis of the general properties of metabolism, and towards attention to biological context, and analysing context specific spaces of metabolic possibility. These shifts were a response to new challenges associated with big data, and the important role of open collaborative interactions between mathematical modellers and researchers from a variety of disciplines in addressing them. These changes in the local epistemology of researchers significantly altered their assessment of the evidential value of data sets constraining the initial expansion of research. The case of metabolic network reconstruction illustrates that open collaborative interactions facilitate researchers overcoming constraints, however they are also involved in the negotiation of new standards imposing different boundaries on research.

6.1 Introduction

Mathematical modelling and large biological data sets are two methodological features frequently used to characterise research practices in contemporary systems

biology. The production of increasingly diverse high throughput biological data sets is associated with an increase in the use and development of mathematical modelling techniques in order to facilitate its analysis and contribution to biological knowledge (Kitano, 2001). The widespread motivation behind bringing modelling and data together is not only that the increased size and organismal diversity of data resources will lead to a high quantity of larger mathematical models, but that building mathematical models based on large biological data sets will lead to significant epistemic developments in knowledge about living systems. Biologists now have access to data that can potentially be used to investigate the complexity of biological systems at multiple scales in much high detail than ever before. It is anticipated that this will lead to advances in more applied areas such as clinical research (Butcher et al., 2004). Indeed, the association between integrative research, bringing together diverse resources, expertise, and techniques, and the development of novel research methods and perspectives on living systems is often pointed to as another key feature characterising research in systems biology (O'Malley and Soyer, 2012; Brigandt, 2013a; MacLeod and Nersessian, 2013c). This chapter examines how the relationship between omics data and metabolic network reconstruction and analysis unfolded over time. It primarily looks at the development of constraint based optimisation in the context of genomics data, but also comments on the impact of the availability of more diverse omics data in the later stages of model development.

Scientific, historical, and philosophical accounts have pointed to particular relationships between broad classes of mathematical modelling techniques, biological data sets, and perspectives on biological systems (Bruggeman and Westerhoff, 2007; Krohs and Callebaut, 2007; Krohs, 2010). Top down modelling approaches have been associated with the use of large high through put data resources and a more 'systems theoretic perspective' (O'Malley and Dupré, 2005) on biological systems. Bottom up modelling approaches have been associated with the use of small biochemical data sets, in particular kinetic data, and a more 'pragmatic perspective' (O'Malley and Dupré, 2005) on biological systems which may verge on a reductionist understanding of biochemical systems. These broader categorisations of the relationship between mathematical models, biological data, and perspectives on systems do point out general interesting clusters in systems biology. This chapter

examines the relationship between these factors at a much finer resolution by closely following the impact of changes in biological data upon one approach to mathematical modelling – constraint based optimisation. From this finer grained perspective the relationship between mathematical model, biological data, ontological assumptions and research questions is constantly shifting and dependent upon the historical moment from which it is viewed.

In analysing the status and use of data within systems biology, I build on and extend Sabina Leonelli's (2009a; 2010b) work on model organism communities which examines the role of biochemical databases in facilitating the re-use of biological data resources in multiple different contexts. Leonelli argues that the evidential scope of data isn't limited to the particular claims it was initially produced in relation to – instead, the evidential value of data isn't implicit but connected to its context of use [14]. She shows how the way in which data is packaged in biological data bases facilitates the reuse of data in multiple different research contexts. Data needs to be decoupled from information about the particular claims it was initially produced in relation to, but it needs to be attached to meta-data about its provenance, e.g. information about the experimental conditions in which it was produced. Packaging data with information about its provenance allows different research teams to assess the value of that data for a variety of claims about phenomena according to their local research standards. Recently, Leonelli pointed out the significance of 'familiarity with research *in vivo*' for assessing the evidential value of data found online (Leonelli, Forthcoming). She argues against the possibility of ever fully automating the production, processing, and interpretation of biological data. This is because research teams need to include members who are familiar with laboratory research practices in order to make use of the meta-data in their assessments of the evidential value of data resources.

The claim that the value of biological data resources is not implicit but dependant on the context of use is central to my analysis[15]. Helen Longino (1979; 1990) provides an excellent philosophical analysis of the relationship between

14 Leonelli's argument is part of an expansive philosophical discussion of the relationship between 'Data and Phenomenon' stemming from the work of Bogen and Woodward (1988; Machamer, 2011).

15 This perspective provides a different angle to work in philosophy of science, including work on data intensive biology, which has looked at the value of data in terms of its context of production focusing on the quality of the instruments and procedures used for its generation (Bechtel, 1990; Bechtel, 2000; Wimsatt, 2007a; Krohs, 2011).

background substantive assumptions, in particular those about causal relationships, and the evaluation of the evidential value of data in relation to hypothesis. This chapter examines the impact of changes in the content of the biological data on the local epistemologies of mathematical modellers. It also highlights how amendments in local research practices and ontological assumptions about metabolic systems affect the way that researchers perceive the data landscape in which they are operating. It builds on the previous chapter's analysis of relationships between mathematical modellers and experimental biochemists by examining the implications of large scale biological data resources on a community of modellers where detailed attention to biochemical data had previously played only a small role in research.

In the next section of this chapter I give a historical case study of the development of Berhard Palssons work on metabolic network reconstruction and analysis in association with large biological data sets. I begin just before Palssons use of the first whole genome sequences as a basis for network reconstruction and illustrate the importance of the wider epistemic context for the subsequent expansion in research which took place. I then go on to explore significant qualitative changes which then occurred in relation to biological datasets, ontological assumptions, research questions, and the social dimensions of research practices, and how these changed the relationship of metabolic network reconstructions and analysis to genomics and wider omics data sets within Palssons constraint based optimisation approach. Finally in the discussion, I reflect more on the role of social interactions in relation to researcher's commitments and the development of novel ontological assumptions and epistemic goals, and the research dynamic of data intensive mathematical modelling.

6.2 Developing genome scale metabolic networks

Constraint based optimisation models were simultaneously developed in a variety of different contexts in the mid 1980's - from bioengineering to educational tools for undergraduate biochemists (Watson, 1984; Papoutsakis, 1984; Fell and Small, 1986). The approach was picked up by Bernhard Palsson in the early 1990's (Savinell and Palsson, 1992). Palsson went on to become the main developer and protagonist for the use of the model in the context of data intensive biology. He was interested in developing large scale mathematical models of metabolic systems

which could be used to analyse systemic metabolic function. As I discussed in chapter four, the development of large scale dynamic models of metabolism was seriously constrained by insufficient high quality kinetic data. However, researchers believed that adequate data about the components of metabolic systems – the metabolic reactions and metabolites, and the structure of metabolic systems – the stoichiometric relations between reactions and metabolites, was available. Compositional and structural data were used as the basis for network reconstruction, building a two dimensional map of the metabolic network. Palsson valued constraint based optimisation because it provided a means of analysing potential functional properties of large scale metabolic networks without requiring kinetic data (Varma and Palsson, 1994, See chapter four of this thesis for a detailed analysis of the early development of this approach). As I explore in the next section, these properties of only requiring compositional and structural data, and being able to analyse systems level properties of metabolic networks, were key features facilitating the initial expansion of the approach in the context of whole genome sequence data sets.

6.2.1 How can genomics data be used to understand metabolism?

The first whole genome data sets for *Haemophilus influenza* and *Escherichia* coli were published in the mid 1990's (Blattner et al., 1997; Fleischmann et al., 1995). These publications also contained functional analysis of genome sequence data, including classification of coding regions to particular types of metabolic reactions. Soon after these were published Palsson and colleagues wrote a series of articles advocating the use of the constraint based optimisation technique for their analysis (Palsson, 1997; Edwards and Palsson, 1998; Schilling and Palsson, 1998; Schilling et al., 1999a). Following this, along with a PhD student, he published the first whole genome scale metabolic models using the sequences of *Haemophilus influenzae* and *Escherichia coli* (Edwards and Palsson, 1999; Edwards and Palsson, 2000a). He proposed that the modelling approach could provide a simple way of analysing large genomic datasets and developing a systemic perspective on functional genomics.

The availability of genome sequences in the data landscape in which modellers were operating initially appeared to further exaggerate the existing discrepancy in the availability of the three different data types; compositional,

structural, and kinetic. Palsson argued that functional annotations of gene sequences would be able to provide the "complete "spare parts catalogue"" (Schilling and Palsson, 1998 p4193) giving researchers comprehensive lists of the metabolic reactions a particular organism was capable of. What's more these comprehensive compositional data sets would not only be available for organisms such as *Escherichia coli* which had been previously well studied, but would be available for any organism with an annotated genome sequence. On the other hand, whole genome sequencing had no impact on the quality or quantity of kinetic data available to model builders. Palsson had developed constraint based optimisation models specifically because they did not require kinetic data. He argued that in this new data context, where model builders were inundated with compositional data, this modelling approach was an ideal mathematical tool for researchers to start the much needed task of analysing large genomics data sets. The change in the availability of different sources of data seemed to enhance the benefit of constraint based optimisation even further over dynamic approaches to modelling metabolism. Constraint based optimisation was promoted as being unconstrained by the availability of appropriate data resources. Looking at the process of reconstructing metabolic networks from genome sequence data and examining the ontological assumptions modellers held about metabolic components and systems is important for understanding why Palsson regarded annotated gene sequences as such a high quality and prolific source of data for reconstructing metabolic networks.

Palsson argued that because metabolism was a well-studied area of biochemistry the analysis of open reading frames into functionally annotated gene sequences would readily provide comprehensive information on the metabolic reactions for many organisms:

> 'Given the long history of metabolic research, the assignment of metabolic genes to ORFs has been particularly successful. About 91% of the known metabolic enzymes found in *Escherichia coli* K-12 had ORF assignments in the initial publication of its DNA sequence.' (Schilling and Palsson, 1998 p4193).

He emphasised that genomics data would provide organism specific lists of metabolic reactions. However, what he meant by organism specific was that the list of open reading frames and corresponding list of metabolic reactions would be organism specific, not that the relationship between open reading frames and

metabolic reactions was organism specific. In other words, the same open reading frame was assumed to have the same relationships with metabolic reactions regardless of the specific organism the gene sequence came from. In the initial reconstruction of the metabolic network based on the genome sequence of *Escherichia. coli* non organism specific databases, such as KEGG, were used to derive the functional annotation of a gene sequence (Edwards and Palsson, 2000a; Ogata et al., 1999). Biochemical data was also used as a source of compositional data about the metabolic reactions an organism was capable of, and this did come from organism specific data resources, such as EcoCyc, which were available for the highly studied *Escherichia coli* (Karp et al., 2000). However, at this time, neither genetic or biochemical data was regarded as stronger evidence for the presence or absence of a metabolic reaction.

Once the organism specific list of metabolic reaction was compiled from a variety of biochemical and genetic data resources, the next stage in network reconstruction was to determine the structural relationships between the different reactions. Reaction stoichiometry refers to the types and numbers of substrates and products associated with a particular reaction. The stoichiometry of individual reactions was used as the basis to determine the structural relationships between metabolic reactions in the network. Genomics data did not directly increase the quantity or quality of stoichiometric data available to researchers. However, Palsson and colleagues argued that the availability of stoichiometric data also did not pose a constraint on network reconstruction. They proposed that:

> ‘‘A universal stoichiometric matrix (U-Stoma) can be constructed as a database of metabolic reactions from which organism-specific metabolic reactions can be selected. Therefore, individual metabolic genotypes of all organisms are comprised of a subset of columns from the U-Stoma.” (Schilling et al., 1999a p290)

The term ‘universal stoichiometric matrix’ is important. Palsson assumed that the stoichiometry of reactions to be universal, i.e. the same reaction would have the same substrate and product relationships regardless of the type of organism, and context within the organism, it was in (Palsson, 2006 Chapter 2). In the quote, the term organism specific only refers to the list of component reactions specific to the organism’s gene sequence, not the stoichiometry of reactions.

The stoichiometry of reactions was understood as an invariant physico-chemical property in comparison to variable biological features of network organisation such as gene regulation, and as such was thought to be consistent regardless of the environment the organism was in:

> "Cells are subject to both invariant (i.e. non-adjustable) and adjustable constraints. The former are physico-chemical in origin and include stoichiometric, capacity and thermodynamic constraints. … Adjustable constraints are biological in origin, and they can be used to further limit allowable behaviour. These constraints will change in a condition-dependent manner. Regulatory events impose temporary, adjustable constraints on the solution space as shown" (Covert et al., 2001 p76)

There were some early developments in trying to integrate genetic regulation into metabolic network reconstruction (Covert et al., 2001). However, even now most network reconstructions do not involve this dimension. In specific research projects gene regulation is sometimes added as an additional level of information to the basic network structure (I will discuss this development further towards the end of section 6.2.6).

These ontological assumptions about reaction stoichiometry meant that the corresponding data appeared to be readily accessible as it could be used irrespective of the organismal and experimental conditions under which it was obtained. As *Escherichia. coli* was well studied by biochemists, organism specific databases such as EcoCyc, and non-organism specific databases such as LIGAND, were used in the first two network reconstructions to determine stoichiometric relationships (Edwards and Palsson, 2000a; Reed et al., 2003; Goto et al., 2002; Karp et al., 2000). In the 2003 model there started to be some recognition that biological context affected reaction stoichiometry and the model was constructed assuming a Ph. of 7.2 (Reed et al., 2003)**.** However, neither organism specific or general biochemical data was regarded as being a more reliable source of evidence for network structure.

Palsson was advocating the use of constraint based optimisation on the basis that it was unconstrained by the availability of compositional and stoichiometric data resources, and that these two data types could easily be combined to build genome scale reconstructions of metabolic networks (Schilling et al., 1999a p290). Explaining the process of network reconstruction only occupied a couple of paragraphs in the

publications of the first two reconstructions of *Escherichia coli* network (Edwards and Palsson, 2000a; Reed et al., 2003). Much more attention in the papers was given to how the model could be used to analyse network function. Reconstructing network structure was considered to be a relatively straightforward aspect of model implementation.

6.2.2 Metabolic network reconstruction expands

The availability of whole genome sequence data sets and the relatively small number of available mathematical models for its analysis led to the expansion of research involving genome scale reconstructions of metabolic networks in several different dimensions. After the initial use of genome sequence data to reconstruct the *Haemophilus influenza* metabolic network in 1999 research began to focus on the development and analysis of the *Escherichia coli* metabolic network (Edwards and Palsson, 1999).

Firstly, the size of the reconstructed networks increased. Prior to the availability of genomics data the *Escherichia coli* network reconstruction contained 30 metabolites and 53 metabolic reactions (Varma and Palsson, 1993). The first network reconstruction of *Escherichia coli* based on genome sequence data contained 438 metabolites and 627 metabolic fluxes (Edwards and Palsson, 2000a). The network reconstruction in 2003 based on an updated functional annotation of the *Escherichia* coli gene sequence contained 625 metabolites and 931 reactions (Reed et al., 2003; Serres et al., 2001). The updated version in 2003 not only expanded the network but also involved updating and removing tens of reactions in the model.

Secondly, the numbers of organisms modelled using the constraint based optimisation approach increased dramatically. Palsson argued that because the network reconstructions could be based on annotated gene sequences it was possible to apply the model to any organism with a functionally annotated gene sequence not just model organisms which had additional extensive biochemical data resources (Reed and Palsson, 2003 p2696). Palsson illustrated this claim in 2002, producing a reconstruction of the metabolic network for *Helicobacter pylori,* a less well studied human pathogen found in the stomach (Schilling et al., 2002). By 2003 whole genome scale metabolic networks had been reconstructed for *Escherichia coli, Haemophilus influenzae*, *Helicobacter pylori*, *Saccharomyces cerevisiae*, and

Methylobacterium extorquens (Reed and Palsson, 2003 p2692). Only six years later more than fifty genome scale metabolic network reconstructions had been published, including organisms from the eukaryota, bacteria, and archaea (Oberhardt et al., 2009).

Thirdly, the production of genome scale metabolic reconstructions also rapidly expanded beyond research groups directly involving or connected to Palsson. Palsson began working on genome scale metabolic network reconstructions alongside a couple of PhD students in the bioengineering department at the University of California. Over the following decade the research teams had not only become increasingly larger in size, but also more international, and interdisciplinary now involving experts from engineering, biochemistry, and bioinformatics. Compared to the speed of the expansion there was relatively little organisation amongst research groups working with the constraint based optimisation method. The number of network reconstructions significantly exceeded the number of organisms which had been modelled (See Figure 2). This led to a situation where there were numerous inconsistent models available for the same species. Between 2003 and 2005 a different group of researchers published a new reconstruction of the yeast metabolic network every year (Förster et al., 2003; Duarte et al., 2004; Kuepfer et al., 2005; Herrgård et al., 2008).

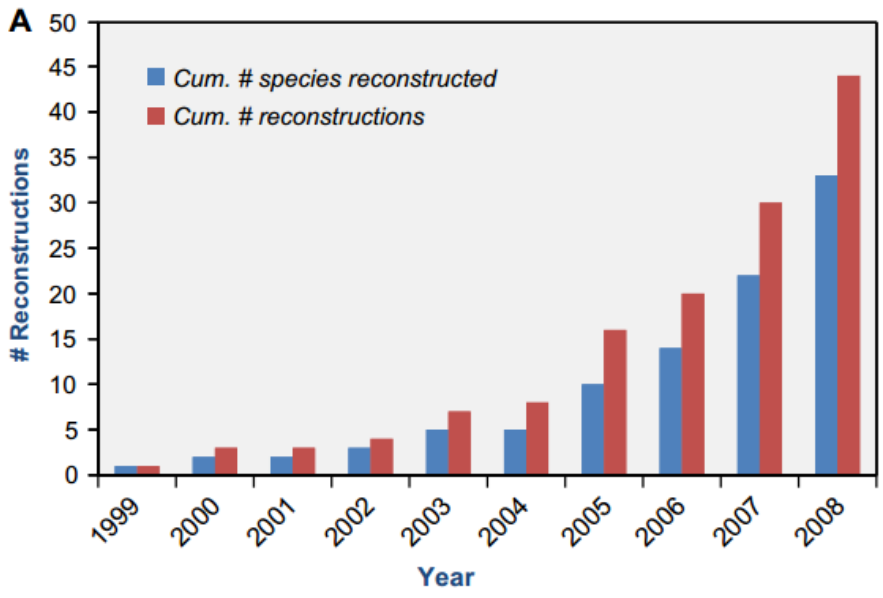


Figure 6-1: Graph illustrating the exponential growth in the number of reconstructions, and the significant number of species with multiple reconstructed networks (Oberhardt et al., 2009).

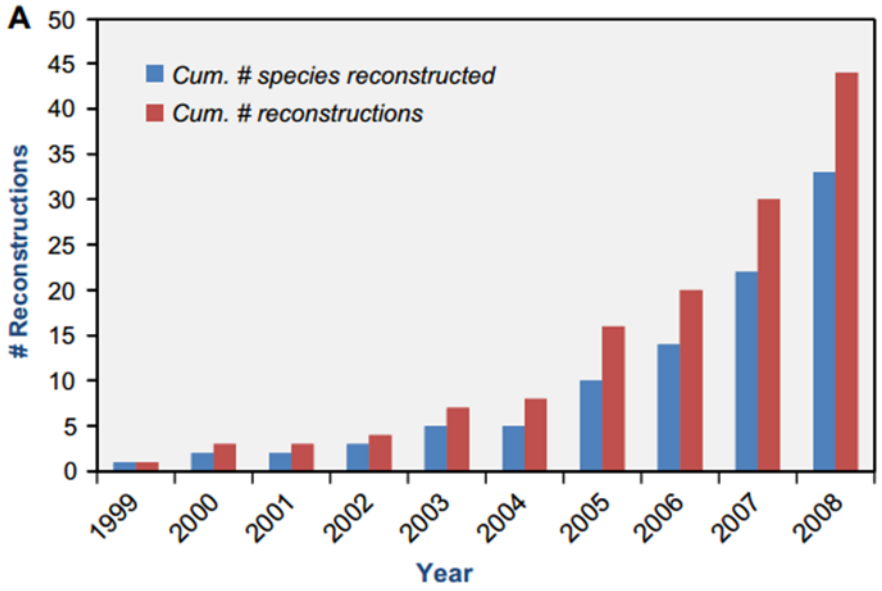


Figure 6-1: Graph illustrating the exponential growth in the number of reconstructions, and the significant number of species with multiple reconstructed networks (Oberhardt et al., 2009)

6.2.3 Shifting research questions: From metabolism to phenomics

Alongside these immediate expansions in metabolic network reconstruction the basis of the model on genome sequence data also led to a more radical shift in the type of research questions which it was used to address. Previously it had been used to address questions about the systemic properties of metabolic networks. Palsson argued that the basis of the model on genomics data meant that the model could also be used to address questions examining the systemic relationships between two levels, the metabolic genotype and the metabolic phenotype.

Palsson promoted the use of the constraint based optimisation approach in the context of analysing genome sequence data because it offered a means of analysing metabolism and metabolic genetics from a systems perspective. He pointed out that in light of whole genome sequence data it was becoming increasingly apparent that there was not a one to one relationship between genotypes and phenotypes (Schilling et al., 1999a). This meant that the analysis of whole genome sequence data sets through the identification of single genes and their association with particular functional proteins was limited. Instead, genome

sequence data needed to be analysed in a way which recognised that functional biochemistry arose from genes and gene products interacting in complex and non-linear ways. However, the constraint based optimisation approach had not been developed in the context of producing systemic perspectives on functional genetics. An outline of the process of network analysis is needed to understand Palsson's case for its application to this new area of research.

Constraint based optimisation stood out as an approach because of its ability to analyse the functional behaviour of metabolic networks without requiring kinetic data about the dynamics of individual reactions (Varma and Palsson, 1994). The analysis of the reconstructed metabolic network involved two stages, firstly, placing constraints on how metabolic flux could be distributed around the network, secondly, using an optimisation function to determine a particular flux distribution under specific conditions. A steady state assumption was the main constraint imposed on flux distribution through the network. (See chapter four of this thesis for a detailed analysis of the use of steady state assumptions in mathematical models of metabolism). In this case the assumption was that the concentration of intermediary metabolites was in a steady state i.e. they were not accumulating or depleting. This meant that the stoichiometric equations for reactions in the network needed to be balanced. This constraint significantly reduced the number of possible flux distributions around the network. However, metabolic flux distribution was still underdetermined by network stoichiometry and the steady state assumption. A second significant constraint was the allowable type and quantity of network inputs and outputs, these parameter values would be informed by biochemical data about the consumption and production of metabolites in particular organisms. In the second stage, an optimisation function was used to determine a single distribution of metabolic flux. An optimisation criterion would be imposed on the network, e.g. maximise the output of a particular metabolite given a certain input. A linear optimisation algorithm was then used to search through the possible flux distributions to determine a single flux distribution which would best achieve this optimisation function. The use of the optimisation criteria was justified by the assumption that functional metabolism arose through optimising evolutionary processes (Schilling et al., 1999b p301).

Palsson emphasised that this approach did not assume an elevated status for canonical biochemical pathways such as the Krebs cycle, i.e. these were not used to inform constraints on the distribution of metabolic flux:

> 'Through the approach developed here, we move away from the traditional definitions of biochemical pathways to a new classification of pathways - classification based on systemic function as opposed to historical discovery.' (Schilling and Palsson, 1998 p4190)

These pathways had been established through experimental techniques starting with the analysis of the characteristics of isolated components before establishing connections between them. In constraint based optimisation, the optimal distributions of flux, or metabolic pathways, obtained through analysing the network were free to result from any combination of interactions between the components in the network within a given set of constraints.

Another difference between the description of pathways in more classical biochemical approaches and constraint based optimisation was the perspective they could provide on metabolic function. Dynamic models of particular metabolic pathways were built up from detailed descriptions of individual components including kinetic parameters; as such they could provide potential explanations of exactly how a series of reactions could regulate itself to achieve a particular flux distribution. In constraint based optimisation, the lack of kinetic parameters included in the reconstruction of the network was one of the main reasons that metabolic flux distribution was underdetermined. The analysis of the network using the optimisation criteria suggested flux distributions which the network might be able to achieve.

> 'in the absence of detailed knowledge of enzyme kinetics, we can estimate the metabolic distribution if we postulate the "objectives" that underlie the cell's behaviour. An underdetermined set of equations can be solved uniquely, given an objective function using linear optimisation techniques.' (Savinell and Palsson, 1992 p424)

However, the lack of kinetic parameters from the model meant it could not be used to give a possible account of how the network regulated flux through reactions so as to actually achieve particular flux distributions. Constraint based optimisation provided a general account of possible functional metabolic states the network might be able to achieve, not a specific causal mechanical account of actual metabolic states which small scale dynamic models of particular metabolic pathways aimed to

provide. As it did not include any dynamic information it also could not provide any detail on the continuous transition of the network between steady states, only discrete snapshots of potential metabolic flux distributions given certain conditions

Palsson argued that because the metabolic reactions in the network had been derived from genome sequence data the model effectively allowed researchers to investigate how functional metabolic states arose from systemic interactions between gene products (Schilling et al., 1999a). He introduced two new concepts associated with the model: Firstly: the metabolic genotype, the metabolic reactions included in the network, Secondly: the metabolic phenotype, the distribution of metabolic flux through the network which best achieved the optimisation function. Palsson presented constraint based optimisation as a model which could allow researchers to investigate phenomics, the relationships between metabolic genotypes and metabolic phenotypes from a systems perspective.

The major application of network analysis was to investigate general features of the organisation of metabolic networks such as the robustness of genotype phenotype relationships. The model included multiple parameters which could be altered during the analysis of the network: network composition, network structure, network constraints, optimisation criteria, input metabolites etc. Palsson carried out different *in silico* investigations looking at how altering these parameters affected the optimal metabolic flux distribution (Edwards and Palsson, 1999; Edwards and Palsson, 2000a). One of the major applications of the model was to remove reactions from the metabolic network and observe the impact on metabolic flux distribution. Using the framework of the metabolic genotype and the metabolic phenotype Palsson presented this as an investigation of the robustness of metabolic phenotypes to changes in metabolic genotype, or *in silico* gene deletion studies. He found that deleting individual reactions from a network usually had an insignificant impact on the ability of the network to achieve the optimisation criteria. These *in silico* experiments were carried out on reconstructed networks for particular organisms and the findings were presented in an organism specific manner (Edwards and Palsson, 2000b p937). However, the finding that metabolic phenotypes were frequently robust to changes in metabolic genotype was common to the reconstructed networks of different organisms. Constraint based optimisation was used, most famously by Andreas Wagner (2005a chapter 9), to substantiate a

general hypothesis about how biological networks were structured to achieve robustness, rather than an account of how organism specific metabolic networks achieved robustness in particular conditions. Although Palsson also hoped that the model could eventually be used to inform bioengineering projects involving targeted modifications to particular organisms, in the earlier period the predictive capabilities of the model were not strong enough to pursue this ambition (Edwards and Palsson, 1998).

6.2.4 A new view of the data landscape

As these seemingly unrestrained expansions in genome scale metabolic network reconstruction were going on there were simultaneous changes in the research commitments of the communities carrying out these reconstructions. These created new standards and practices for research which significantly changed researcher's perceptions of available data resources and their impact on mathematical modelling. Palsson published two methodology papers in 2010 formalising these developments in constraint based optimisation. The first one gave a detailed account of best practice for network reconstruction (Thiele and Palsson, 2010a). The second provided guidance on how the research community should organise itself so as to maximise the quality and efficient production of network reconstructions (Thiele and Palsson, 2010b).

Reconstruction of a new network currently involves two stages. Firstly, an automated reconstruction is generated that constitutes a draft reconstruction of the network. Secondly, the draft reconstruction is amended through a manual reconstruction process resulting in a high quality curated network. The exponential increase in the quantity and diversity of available high throughput data led to concerns about the amount and efficiency of work involved in managing and utilising these data flows manually. This resulted in attempts to develop automated programs for integrating and transforming data into computational models. In 2002 a set of Pathway Tools was published, including the PathoLogic program for the automated reconstruction of metabolic networks from annotated genome sequences and general biochemical databases (Karp et al., 2000; Karp et al., 2002). In 2005 this

program was used for the first time in conjunction with the constraint based approach to build a reconstruction of the *Streptomyces coelicolor* metabolic network (Borodina et al., 2005). The authors found that:

> "Automatically created models need to be manually curated using books, literature, and other available information sources. The most problematic aspects are wrongly or insufficiently defined substrate specificity, reaction reversibility, protein complexes, cofactor specificity, and the missing enzymes." (Borodina et al., 2005 p821)

These problems were attributed to the lack of organism specificity in many of the major genomic and metabolic databases. Achieving high quality automatically produced network reconstructions has continued to be problematic and only constitutes the first four relatively simple steps of the 96 step 2010 protocol for the constraint based optimisation approach (Thiele and Palsson, 2010a), although attempts to achieve high quality automated metabolic network reconstructions are ongoing (DeJongh et al., 2007). To create the automated draft reconstruction, the most updated version of the annotated genome sequence, which provides an initial set of component metabolic reactions, is integrated with data from general biochemical organism non-specific databases, such as KEGG and BRENDA, providing a rough guide to the structural relationships between components (Ogata et al., 1999; Schomburg et al., 2002; Thiele and Palsson, 2010a). As 'the quality and wealth of organism-specific information will directly affect the quality and coverage of the metabolic reconstruction' (Thiele and Palsson, 2010a p95) this version is treated as a low quality rough draft of the network.

The second stage of network reconstruction, the manual curation process, involves researchers consulting organism specific databases, journal publications, books and target organism experts, verifying and updating the network component by component and interaction by interaction. Researchers even carried out biochemical experiments which facilitated improvements to the 2011 draft reconstruction of the *E. coli* metabolic network (Orth et al., 2011 p2). Firstly, all the compositional data for the organism derived from the annotated genome must now be verified by organism specific biochemical data. Verifying functional gene annotations with biochemical data is an impressive task given that the latest reconstruction of *E. coli* contains 1366 genes (Orth et al., 2011). The reconstruction process is now based on the ontological assumption that the relationship between

genes and functional products, in this case metabolic reactions, is organism specific - the same gene can be associated with different reactions in different target organisms. In a critique of annotations provide in the Gene Ontology (Ashburner et al., 2000), Palsson and colleagues argued that:

> 'The GO annotation does not adequately reflect this basic condition dependency of gene function as it strives to maintain uniformity in the experimental conditions underlying the annotation.' (Shlomi et al., 2007 p1626)

The association of genes with gene products is now regarded as organism specific and condition dependant, as such, care is taken to evaluate genomic evidence for the composition of the network using biochemical data resources. All the components included in the model are given a 1-4 confidence score depending on the data type they are based upon. Biochemical data is regarded as far more reliable than genomic data, and if the presence of a component is only indicated by the genomic data then it is to be assigned the lowest confidence score (Thiele and Palsson, 2010a p97). In 2007 this process led to significant changes in 289 of the reactions included in the earlier 2003 model of *E. coli* (Feist et al., 2007p3).

Secondly, the structural links, the type of substrates and cofactors, associated with different reactions need to be verified through organism specific data resources. This is because the metabolites associated with the same reaction are now assumed to differ between target organisms, and so this information cannot reliably come from a general or universal information source. At best, if organism specific biochemical data is unavailable, Palsson advises that data should be used from phylogenetically close organisms (Thiele and Palsson, 2010a p97). After this the metabolites are given a pH dependant charged formulae. Whilst the pH is frequently assumed to be 7.2 it is recognised that this varies depending on environmental conditions and location within different organelles (Thiele and Palsson, 2010a p97). Only then is the reaction stoichiometry completed giving the number of metabolites and products on either side of a reaction. The stoichiometry of reactions is now assumed to be something which is specific to a particular organism and specific to the conditions the organism is in - 'Organism specific features, such as substrate and cofactor utilization of enzymes, intracellular pH and reaction directionality remain problematic and thus require manual evaluation.' (Thiele and Palsson, 2010a p93). Additional

organism specific information is added, such as the directionality of reactions, and the location of reactions in the cell. Aspects of the reconstruction are assigned a confidence score which then assists with further iterative rounds of network development and evaluation. Describing the process of network reconstruction, as well as being itself the subject of entire articles, now takes up significant space in the updated 2007 and 2011 publications of the *E. coli* metabolic network reconstruction and its analysis.

The publication of an updated protocol for genome scale metabolic network reconstruction involves significant shifts in researcher's ontological assumptions about the relationship between genomes and metabolic components and the structure of metabolic networks. Rather than gene sequences being associated with the same metabolic reactions regardless of the organism, the relationship is now assumed to be organism specific. Reaction stoichiometry is no longer assumed to be a universal physico-chemical property, but something which varies between organisms and with a variety of properties of the intracellular environment. These shifts in assumptions have had major implications for the impact of whole genome sequence data on mathematical reconstructions of metabolic networks. Initially whole annotated genome sequences were regarded as a high quality source of data about the composition of metabolic networks facilitating the rapid reconstruction of metabolic networks for a wide variety of organisms. The evidential value of these data resources has radically shifted as changes in researchers assumptions are coupled with changes in their perception of the data landscape in which they are operating. Annotated genome sequences are now regarded as providing the least reliable source of evidence for network composition. This data needs to be painstakingly manually evaluated and supplemented using a variety of organism specific published data resources and the advice of subject specialists.

The stringent requirements for high quality organism specific data on different aspects of cellular metabolism for every component and connection in the network now constitutes a major constraint on model building. The task of network reconstruction has now become:

> '... very labour and time intensive, spanning from 6 months for well-studied, medium sized bacterial genomes, to 2 years (and six people)

> for the metabolic reconstruction of human metabolism...' (Thiele and Palsson, 2010a p93)

For well-studied model organisms like *E. coli* this data is more readily available, but must be laboriously obtained through manually exhausting disparate potential resources. For organisms which are less well studied, the genome sequence is no longer enough to base the network reconstruction upon and the dearth of biochemical resources limits the implementation of the constraint based optimisation approach. These high quality curated networks are stored in 'BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions.' (Schellenberger et al., 2010), so they can be accessed by research teams in order to carry out network analysis. Upon its publication in 2010, BiGG contained curated networks for seven organisms, by July 2014 this number had only risen to ten (http://bigg.ucsd.edu/bigg/main.pl accessed 30/07/2014). The process of metabolic network reconstruction has become heavily constrained by the availability and accessibility of appropriate data resources limiting the quantity and diversity of organisms for which high quality network reconstructions can be produced.

6.2.5 A community response to the dynamics of data

The data landscape in which mathematical modellers are operating is constantly shifting along two different axes. Firstly, as has been widely documented the quantity and diversity of large biological data repositories is continually changing as technologies for data production, processes of data sharing and data curation, develop. This shifts the potential data resources which mathematical modellers have available to them. Secondly, the data landscape changes because as the research commitments of mathematical modellers change so too does their assessment of the evidential value of the data available for them to use. The case of constraint based optimisation illustrates that changes in this second dimension can be as fast paced and as significant for research practices as developments in the actual contents and availability of data resources. Increasingly, intentional engineering of the social organisation of research communities involved in network reconstruction is being used to maximise effective research in this dynamic environment.

Data intensive biology poses multiple challenges for mathematical modelling. As I pointed out in section 6.2.2, when data availability led to unrestricted growth in mathematical model building the disorganisation of the modelling community resulted in multiple inconsistent reconstructions being produced for the same organisms. In section 6.2.3 I showed how model building has become a labour and time intensive process. The community of researchers around constraint based optimisation recognises that in this dynamic data context it needs to develop practices which help it make mathematical model building as streamlined and efficient as possible. Initially it looked towards automated model reconstruction as a tool for reducing the labour required for model building and introducing consistency across research groups (Borodina et al., 2005). However, Palsson's emphasis is now on developing the social organisation of the research community as the best approach to supporting high quality and efficient mathematical modelling (Thiele and Palsson, 2010b). Far from being a process which can remain relatively quick and straightforward through using automated processes to manage the changing data landscape, network reconstruction has become an elongated and social aspect of scientific practice (See also Kitano et al., 2011 and ; Leonelli, Forthcoming).

Significant changes in the social composition and organisation of the research community involved in constraint based organisation took place over the first decade of genome scale metabolic network reconstruction. In the late 1990's Palsson worked on genome scale network reconstructions along with a PhD, Jeremy Edwards, in the bioengineering department at the University of California (Edwards and Palsson, 1999; Edwards and Palsson, 2000a). By 2002 Palsson began to take part in projects involving 'the integration of diverse interdisciplinary quantitative, experimental and computation approaches' (Kolker et al., 2002 p342), which involved collaboration between individuals from multiple different disciplinary backgrounds and research organisation. At this point in time such interdisciplinary collaboration was still considered a 'unique feature' (Kolker et al., 2002 p342) of research rather than standard practice. Over the next few years the number of research groups working on network reconstruction and analysis, and the interdisciplinary character of those groups, expanded. At this stage interdisciplinary collaboration was occurring between small groups of researchers who independently reconstructed their own network reconstruction and were working on a specific

research question. As I previously mentioned, this gave rise to a situation in which multiple inconsistent metabolic networks were being reconstructed for the same organism. This was perceived as causing issues for evaluating and comparing results across different research groups. Palsson and others suggested that different research groups working on the same organism needed to find ways of collaborating in order to achieve a consensus reconstruction of the network of particular organisms upon which they could base their independent research projects. This idea was first discussed at the 'The 25th International Specialised Symposium on Yeasts' in 2006 (http://issy25.vtt.fi/index.htm, accessed 30/06/2014) (Mo and Palsson, 2009). The proposal was promptly put into practice and a community consensus version of the yeast network reconstruction was published in 2008 (Herrgård et al., 2008).

The community approach to building metabolic network reconstructions of yeast, and soon after *Saccharomyces cerevisiae* was considered successful (Mo and Palsson, 2009). Based on these experiences Palsson published guidelines on how to organise an effective 'Reconstruction annotation jamboree' (Thiele and Palsson, 2010b). The ''jamboree' approach'' involves 'a large, focused work meeting, where we defined the protocol for the curation process as well as resolving the majority of discrepancies between the existing reconstructions.' (Herrgård et al., 2008 p1156). It is a meeting in which researchers agree the evidential standards and procedures involved in network reconstruction and then implement in their analysis of currently available data to achieve a consensus version of the network. The meetings involve 'experts in systems biology (for modelling); chemistry and metabolomics (for metabolite information); biochemistry, molecular and cell biology (for reaction and genetic information); and bioinformatics (for gene annotation and database structure).' (Thiele and Palsson, 2010b p1). They involve a large group of interdisciplinary researchers working together over two to five days with a smaller group preparing for the jamboree and finalising the consensus reconstruction after the jamboree has taken place. Significantly, it is recommended that follow up jamborees are held every two years. This mode of community organisation is based on the understanding that in a sense the reconstructed metabolic network will never be complete but will continually shift as data resources are updated, researchers assumptions about metabolic systems change, new

modelling approaches are developed, and different hypothesis are formulated. Even though the 'reconstruction jamborees' intend to arrive at a coherent community consensus, it is recognised that this consensus is temporary and will soon change.

6.2.6 Contextualising network analysis

The process of network analysis has also undergone major developments over the fourteen years that constraint based optimisation had been used in data intensive biology. Initially, metabolic network analysis was used to support general claims about metabolic function. Experimental validation of model predictions showed consistency rates of between 70%-86% (Gianchandani et al., 2010). It was hoped that network analysis could eventually play a more significant role in targeted metabolic engineering, for instance facilitating the development of biofuels, and understanding metabolic disease states to facilitate drug discovery (Raman and Chandra, 2009). In order to play a significant role in guiding discovery the accuracy of predictions made using constraint based optimisation needed to be improved. The developments in network reconstruction outlined above, and further refinements of network analysis examined below, aimed at achieving increasingly accurate predictions of organism and context specific metabolic behaviour(Oberhardt et al., 2009). Effectively many of these developments involve trying to mitigate the lack of dynamic regulatory information included in the model by imposing further constraints on metabolic flux distribution. These developments were coupled with a diversification in the applications of constraint based optimisation, alongside the ongoing refined investigation of genotype – phenotype relationships, and network properties such as robustness. Although these models still don't produce causal mechanistic perspective in the same way as small scale dynamic models, they do provide much more refined context specific predictions of the metabolic behaviour a particular system might be capable of and the impact of perturbations on that system.

The developments previously documented in enhancing of network reconstructions played a significant role in improving the accuracy of network predictions. A study of *Pseudomonas putida* in 2008 indicated that the quality of the network structure had more impact on the accuracy of predictions than the quality of the objective functions used during network analysis (Puchałka et al., 2008). A

combination of genomics and biochemical data is still the predominant basis for basic metabolic network reconstruction. However, over the past fifteen years there has been an increase in the diversity of different large scale omics data sets available to mathematical modellers. Researchers have used these data sets in a later stage of network refinement in order to tailor the structure of the networks towards their particular research agenda (Feist et al., 2008). The contextualisation and interpretation of a wider range of omics data types is now considered one of the major uses of the model. For instance, complex eukaryote organisms consist of different cell types exhibiting their own metabolic behaviour. Gene expression profiling data has been used to build reconstructions of metabolic networks which are specific to particular human tissue types, for instance skeletal muscle cells (Becker and Palsson, 2008; Shlomi et al., 2008). Metabolomics data has been used to tailor the *Escherichia coli* metabolic network to represent important differences in anaerobic and aerobic metabolism (McCloskey et al., 2013). In some cases, new omics data sets are collected from biological samples which are geared towards providing the particular information required to build the context specific network reconstructions. Another development in the structure of the networks analysed which also represents a significant new use of the model is the analysis of interspecies interactions. Researchers have coupled the reconstructed metabolic networks of multiple organisms to analyse metabolism in the context of eco-systems, this has included work on biofilms and host pathogen interactions (Bordbar et al., 2010; Klitgord and Segrè, 2010).

The use of an objective function during network analysis, and the assumption that metabolic networks will always be optimising, has undergone scrutiny and development. An influential paper was published in 2007 in which researchers used the *Escherichia coli* reconstruction to evaluate the predictions made using various objective functions against experimental data collected in particular conditions. This study highlighted that no single objective function could produce accurate predictions for all possible conditions and that objective functions needed to be carefully selected depending on the conditions, for instance the type and quantity of inputs, of the metabolic simulation (Schuetz et al., 2007). Another significant development was the development of a non-optimising function, the minimisation of metabolic adjustment (Segre et al., 2002). This assumes that networks will respond to

perturbations by minimising the redistribution of their metabolic fluxes, even if this results in a sub-optimal growth rate. As such, it allows for the possibility that metabolic networks may not have undergone evolutionary processes establishing optimisation strategies in response to all possible perturbations. This development again increased the ability of researchers to tailor network analysis to more organism and context specific research questions.

6.3 Data, social interactions, and research standards

The impact of omics data on constraint based optimisation mathematical modelling is affected by two different dimensions 1) the contents of available data sets, 2) researchers perceptions of data sets. In this section I discuss the crucial role that the social organisation of research communities has played, on the one hand in managing and responding to challenges associated with changes in the contents of available data, and on the other hand in leading to changes in researchers perceptions of that data and establishing new standards for its evaluation. Analysing these relationships highlights that in the context of genome scale metabolic network reconstruction open collaborative interactions support effective research in multiple ways; they enable researchers to work around constraints, but they also facilitate the construction of research standards imposing new constraints on network reconstruction.

In the case of constraint based optimisation, as research involving the reconstruction of metabolic networks on the basis of genomics data sets expanded unanticipated problems and challenges began to emerge. Firstly, it became clear that different research teams were building reconstructions of the metabolic network of the same organism which were inconsistent with each other. Secondly, the process of network reconstruction was becoming increasingly labour and time intensive. These issues partially resulted from changes in the contents of available data sets. Increasingly large and diverse omics data sets were being produced, and the functional annotations of those data sets were often updated (Palsson and Zengler, 2010). Differences between research groups network reconstructions partly resulted from differences in the contents of the data sets researchers were basing

their network reconstructions upon, and the increasing length of time partially reflected an increase in the complexity of available data sets.

However, as Palsson later pointed out issues also arose due to 'the irregularities in omics data that are caused by underlying intricate molecular mechanisms' (Palsson and Zengler, 2010 p788). Not all the irregularities in the data researchers were using resulted from technical issues with the way data was being produced and curated. Differences in the contents of omics data sets relating to the same organism also stemmed from the organism specificity and condition dependency of molecular features which were not necessarily reflected in data production and curation practices (Shlomi et al., 2007). Different research teams would interpret omics data sets differently depending on their ontological assumptions about metabolic components and their interactions. Whereas one team might accept genomics data as evidence for the presence of a reaction in a metabolic network another might not depending on whether they assumed that the relationship between genome sequence and metabolic reactions was a static physico-chemical property or a fluctuating condition-dependant relationship. As different teams held different ontological assumptions, and different standards for the evidential value of data and processes of network reconstruction, inconsistencies arose between multiple network reconstructions for the same organism. Additionally the work required to reconstruct a network increased significantly for teams requiring biochemical as well as genetic evidence for the presence of absence of a reaction.

Initially the research community consisted of multiple distinct groups, with different levels of expertise and disciplinary diversity, each evaluating data and reconstructing networks according to their own standards. This situation led Palsson to publish his 2010 'Protocol for generating a high-quality genome-scale metabolic reconstruction' (Thiele and Palsson, 2010a) in order to try and standardise research practices. However, alongside these technical guidelines Palsson also thought that due to the speed and complexity of developments in data production and network reconstruction the social organisation of the research community would play an important role in overcoming the challenges being posed by network reconstruction. This led to his 2010 publication giving guidance on how to effectively coordinate collaboration between the community of researchers involved in network reconstruction (Thiele and Palsson, 2010b). The social organisation of the research

community was seen as playing an integral role in maintaining consistency in the metabolic network reconstructions being used in conjunction with constraint based optimisation, and in ensuring that the time and effort involved in network reconstruction is streamlined. On the one hand, the role of 'Reconstruction annotation jamborees' was to update network reconstructions in line with changes in the contents of data sets available to researchers. On the other hand, before any changes could be made to network reconstructions, 'Evaluation and decision criteria need to be established, e.g., how current knowledge is evaluated, which reactions/genes should be kept and based on which evidence.' (Thiele and Palsson, 2010b p2). Jamborees are also a location in which researchers from a diversity of disciplinary backgrounds negotiate their standards for the evidential value of data and the assumptions which lay behind them. These collaborative interactions play a crucial role in affecting how researchers perceived the evidential value of the data used as the basis for network reconstruction. Significantly Palsson suggests that jamborees be held every couple of years, allowing researchers to keep a pace with rapid changes in the contents of the data available for network reconstruction, and to renegotiate their standards for evaluating that data based on changes in their knowledge and assumptions. Whilst the community jamborees aim to achieve a consensus reconstruction it is recognised that the consensus is transient and will soon change; they provide researchers with a temporary focus around which they can periodically gather and depart.

In chapter five I discussed MacLeod and Nersessian's recent work on the characteristics of interdisciplinary interactions in systems biology which facilitate creativity and innovation in research (MacLeod and Nersessian, 2013c; MacLeod and Nersessian, 2013b; MacLeod and Nersessian, 2013a). They argued that interactions which are characterised by openness and epistemic pragmatism allow researchers to rapidly overcome constraints by being flexible in their research commitments, enabling them to inhabit 'largely unstructured task environments' termed adaptive problem spaces (MacLeod and Nersessian, 2013c p1). Researchers from different disciplinary backgrounds involved in the reconstruction of metabolic networks appear to be engaging in open collaborative interactions. The interdisciplinary research groups, and the community jamboree's, constitute spaces where researchers commitments are flexible and they are open to amending their

ontological assumptions and evaluation of data sets in response to alternate research commitments held by other members of the community. As MacLeod and Nersessian argue, this process helps researchers overcome constraints on research, in this case those arising from inconsistencies between different network constructions, the quantity and complexity of the data they are attempting to evaluate, and the amount of time it takes to produce a high quality network reconstruction.

However, the case of reconstructing metabolic networks illustrates that it is important to emphasise that open collaborative interactions also play a role in constructing new constraints. These constraints provide the structure and boundaries facilitating research. Open collaborations don't lead to a situation where anything goes. Longino has contributed some crucial arguments supporting the position that the evidential value of data is not intrinsic but dependant on its context of use (Longino, 1979; Longino, 1990). She argues that given that knowledge is relational to its social context, ideally that social context needs to be such that it supports the critical and transparent development of research standards and standards for the evaluation of knowledge claims (Longino, 2002). Longino gives four criteria describing an 'idealized epistemic community' in which this takes place (2002 p134). 1) Venues: there need to be locations, including academic journals, and meetings, in which critical debate about researchers commitments can take place. 2) Uptake: the community must be open to integrating criticism into its practices, not just tolerate its expression. 3) Public standards: the community must be explicit in its standards, importantly the 'standards are not a static set but may themselves be criticized and transformed, in reference to other standards, goals, or values held temporarily constant.' (Longino, 2002 p131). 4) Tempered equality: the community should pay equal attention to a diversity of perspectives, whilst being aware of when expertise and experience may appropriately bias attention to some opinions over others. Longino's 'idealized epistemic community' is one in which open collaborative interactions take place, different members of a research community are open to each other's perspectives and integrate or uptake these into their research practices. However, it is also one in which open collaborative interactions take place in particular locations and aim to publically articulate, at least temporary, standards for research.

In the case of metabolic network reconstruction open collaborative interactions are not only about the removal of constraints, they are also about constructing constraints which provide a consensus framework within which a research community can operate and coordinate its research. Interdisciplinary interactions in the research community play a role in changing the ontological assumptions which underpin researcher's assessments of the evidential value of data. For instance, from assuming that the relationship between metabolic reactions and metabolites is a general physico-chemically determined, to the assumption that it is a relationship which is organism specific and context dependant. As I illustrated this change in ontological assumption led to researchers encountering novel constraints including an increase in the amount of effort required to reconstruct a metabolic network and a decrease in the amount of what researchers consider to be high quality data available for network reconstruction.

It is interesting that the more recent social interactions involved in network reconstruction appear to have much in common with Longino's description of an 'idealized epistemic community'. They involve particular locations for 'critical discursive interactions' (Longino, 2002 p129), e.g. jamborees, the uptake of critical opinions into practices of network reconstruction, public standards for network reconstruction along with the public recognition that these standards are temporary, and tempered equality, jamboree's involve many different voices each of which are listened to more closely at different points in the reconstruction process. However, the jamboree proposal of collaborative interactions is relatively recent and has only been implemented in association with a handful of network reconstructions (Herrgård et al., 2008; Thiele et al., 2011; Thiele et al., 2013). Evaluating the characteristics of social interactions at jamborees would benefit from a sociological analysis of a jamboree in practice alongside their representation in journal articles. The complex social organisation and interactions which are increasingly taking place in relation to data intensive systems biology provide an interesting basis for further research evaluating and developing philosophical work in social epistemology.

6.4 Conclusion

In this chapter I reflected upon the impact of large biological data sets on mathematical modelling in biochemistry through examining a detailed historical case study of the constraint based optimisation approach. I have argued that understanding that impact requires not just looking at how high throughput data production is changing the amount and diversity of data available, but also the local epistemic context of researchers which affects their assessment of the evidential value of that data. Researcher's ontological assumptions about the components of metabolic systems, their relationships to genome sequences, and their stoichiometric relationships underwent a significant shift during the transition to data intensive genome scale metabolic network reconstruction. Initially they were assumed to be general, even physico-chemical, properties common to all metabolic networks and then shifted to being features of metabolic networks which researchers regarded as being dependent on the particular organismal and environmental context in which they were located. Researchers perspective on metabolic components and their interactions has shifted from being reductionist, in the sense that the properties of components are determined in isolation from their wider systemic context to what Dupré terms relational, 'the capacities of a thing,… , can be seen to depend on the relationship between the thing and the environment in which it exists.' (Dupré, 2009 p37).

As I have shown throughout the chapter this altered researchers perception of the epistemic value of particular data resources and the constraints which data availability imposed on research. Initially, genome sequence data and general biochemical data was perceived as a high quality and ubiquitous source of information for network reconstruction. Within the relational perspective on metabolic networks, genomic and general biochemical data became seen as a poor quality data resource and network reconstruction needed to be based on context specific biochemical data the availability and complexities of which constrained network reconstruction. Alongside this there have been significant changes in the level of analysis researchers aspire to. Initially, networks which were constructed on the basis on reductionist assumptions were used to make very general claims about the relationship between metabolic genotypes and metabolic phenotypes, and the

distribution of metabolic flux with minimal certainty over the capacity of networks to achieve these distributions. Later networks based on a relational perspective on the composition and interactions in metabolic networks were used to analyse condition dependent genotype-phenotype relationships and metabolic flux distributions within far more constrained spaces of metabolic possibility.

These changes coincided with alterations in the composition and organisation of the research community involved in network reconstruction. Initially research teams consisted of two researchers from a single discipline and department, and then research teams began to involve interdisciplinary collaboration between researchers from multiple disciplines and departments, currently collaboration takes place between different research teams working on reconstructed metabolic networks of particular organisms in 'reconstruction annotation jamborees'. Open collaborative social interactions play a major role in researchers managing and responding to changes in the contents of data sets which are available to them. At the same time, they are forums where researchers negotiate their ontological assumptions and standards for data evaluation influencing how they perceive that data. As MacLeod and Nersessian argue, epistemic openness allows the community to find ways of overcoming constraints which inhibit the efficient reconstruction of high quality metabolic networks. However, the case examined also illustrates that open collaborative interactions are also involved in researchers reaching a public consensus on new standards which impose different constraints and boundaries on network reconstruction.

Sabina Leonelli and Rachel Ankeny (2011) have examined the impact of biological data bases and other cyber infrastructures on the researcher practices of communities of model organism biologists. These communities differ significantly from the groups of mathematical modellers I examined in this chapter. Leonelli and Ankeny argue that model organism biologists already constituted communities with a degree of formal organisation in which biological data was a significant feature of research and some data sharing practices were already in use. They found that one of the major impacts of biological data bases was to increase the visibility and standardisation of existing research practices within the different communities. More fundamental changes were also observed, including the facilitation of interdisciplinary and international collaboration, and research based on comparisons

between different model organisms. In contrast, twenty years ago researchers working on mathematical models of metabolic systems tended to form smaller clusters around particular models with little or no overarching organisation. Interactions between modellers and experimental biochemists producing biochemical data were often minimal and incredibly problematic due to stark differences in the research commitments of the two communities (Cornish-Bowden and Cárdenas, 1990; Green and Wolkenhauer, 2013, See also chapter five of this thesis). Mathematical modelling has a history of being associated with abstract perspectives on biochemical systems and general accounts of functionality whereas experimental biochemists strove for systems specific detailed causal explanations. The impact of changes in the data landscape and social organisation on constraint based optimisation led to fundamental shifts in researchers ontological assumptions and epistemic goals as these different communities began to engage in more epistemically open interactions.

This fine grained perspective contrasts with broader accounts which often posit a less dynamic relationship and stricter categories defining rigid structural divisions between type of research in systems biology (O'Malley and Dupré, 2005; Krohs and Callebaut, 2007). The detailed nature of this case study highlights the continual transience of the relationship between a mathematical modelling technique, ontological assumptions, level of analysis, biological data, and research community. The configuration of these aspects of research depends on the particular time at which the community is observed and analysed. Research practices appear to be driven by a momentum which involves simultaneous pathways of re-stabilisation and destabilisation in different areas. In the case examined researchers worked to integrate changes in the content of biological data sets into their research, this in turn led to changes in the wider epistemic context of network reconstruction and analysis which needed to be managed through developing the standards for network reconstruction, community organisation, and integrating more diverse sources of omics data. As one feature of the local epistemology became established and standardised, this often mirrored the destabilisation of another feature. As one set of research problems is solved another is created, sustaining the drive to find innovative solutions leading to the development of new standards for practice in a different aspect of research.

7 Discussion: Philosophical and historical implications

This thesis has examined the significant role that a wide range of assumptions played in the development of mathematical models of metabolism from the 1960's to the present day. In this chapter I discuss how it expands on the current philosophical perspective on assumptions in biology along several dimensions. In chapter two I highlighted the philosophical emphasis on assumptions about the spatial organisation of parts and wholes in biological systems. These assumptions constitute ontological claims about what biological systems are composed of and how they are organised. They also act as the basis for idealising assumptions facilitating simplifying research strategies, often reductionist, used during the investigation of complex biological systems. Much of the philosophical work building on this has focused on how these assumptions are connected to explanatory practices in biology, and particularly mechanistic explanation. Philosophical literature on mathematical modelling across the sciences has paid attention to the wider variety and role of assumptions in scientific research. For instance, within the models as mediators framework assumptions can be related to a variety of different sources - theory, data, instruments, and are seen as playing a transformative role enabling learning and novelty to arise through the use of models in research.

The thesis began by asking an open ended question about what assumptions researchers make and how they play a role in the development of mathematical models of metabolism (See Thesis Introduction p11). The four categories of assumptions bought to light through this research contain a wide variety of things which influence model building in different ways. What is it that connects these categories under the common label of assumptions? All of the different types of assumptions refer to beliefs researchers hold, either about the metabolic systems they are investigating or the research methods and practices they are using in their investigations. Assumptions do not constitute theoretical aspects of research which are subject to confirmation or direct investigation during the research process. Instead they provide the background knowledge and expectations which constitute the structure within which research questions are formulated and investigations carried out. However, whether something counts as an assumption at all varies over the course of model development. For instance, in chapter three I examined how

Joseph Higgins constructed the pre-cursor model to metabolic control analysis on the basis of an ontological assumption, or a belief, about the existence of a rate limiting step. This belief was not widely discussed or justified as it was a common piece of knowledge taken for granted by a large community of researchers involved in the construction of metabolic control. The construction of the mathematical model led to the rate limiting step ceasing to be an assumption and becoming a theoretical position which was being tested and investigated.

The assumptions researchers make can be implicit or explicit. Implicit assumptions are those which are not clearly stated or discussed but nevertheless underpin the work which is being carried out. For instance, Savageau's account of descriptive rate laws discussed in section 4.2.1.2. does not mention their basis on the assumption that the concentration of substrate exceeds the concentration of enzyme enabling the concentration of enzyme substrate complex to be treated as a constant. When these types of descriptive rate laws were developed in the 1920's this assumption was made explicit by researchers (Briggs and Haldane, 1925), but by the 1970's it is taken for granted to the extent that it is implicit in Savageau's work. Explicit assumptions are those which are clearly stated in the publications examined. Explicit assumptions fall into two categories. Firstly, the content of the assumption is made explicit. Ontological assumptions are often explicit in the sense that whilst they are clearly stated they are taken for granted as common knowledge amongst the relevant research community and their use does not have to be subject to further discussion. During the early development of the control based optimisation method Palsson and colleagues clearly stated that they understood reaction stoichiometry to be universal but took this to something which was obvious and did not require further explanation. Secondly, the content of the assumption is made explicit and researchers also explicitly highlight that it is an assumption and attempt to justify and explain its use. For instance, researchers developing metabolic control analysis made it clear that the model was based on an idealising assumption of a linear relationship between two variables where the actual relationship between those variables was known to be non-linear. This assumption was justified on practical grounds and the limitations it imposed on the model explicitly discussed. It is interesting that most of the assumptions examined throughout this thesis's analysis of mathematical modelling fall into the explicit category. There are two

reasons for this: Firstly, a perceived need to justify and explain novel model building processes in order to increase their acceptability in the wider community. Secondly, the frequency of interactions between researchers from different backgrounds. Often interactions between research cultures which hold different assumptions will lead to the explicit discussion of and attempt to justify these assumptions. In Chapter six I examined how the growth and diversification of the research community around constraint based optimisation bought the assumption of universal stoichiometry to the forefront leading to its revision through explicit discussion and re-formulation.

In this chapter I bring together material from across chapters three to six relating to the four different categories of assumptions and the impact they have on the development of mathematical models of metabolism. I examine four different categories which emerged in the thesis: ontological assumptions, idealising assumptions, assumptions about biological data, and researcher's commitments. I expand the category of ontological assumptions to include assumptions about the temporal organisation alongside those about the structural and functional organisation of biological systems. I also highlight what I term 'connecting assumptions' as another sub-category of idealising assumption alongside simplifying assumptions. Connecting assumptions are types of idealising assumptions which facilitate the construction and use of a model through facilitating its connection to particular resources for model building, or to particular explanatory claims. The analysis in this thesis also brings to light the significant role that assumptions about data play in the development of mathematical models. These include assumptions which connect models to particular data resources, and also modeller's assumptions about the contribution that biological data makes to the mathematical representation of the system. I also discuss the role of researcher's commitments, their assumptions about how research should be carried out. An important dimension of these assumptions, which influences the characteristics of social interactions between research communities, is the quality of researcher's commitments, the way in which they are committed to certain aspects of research.

After giving an overview of these four categories I discuss their philosophical implications. First, I look at the role of assumptions about data in reassessing the role of reductionism in biological mathematical models. Secondly, I explore the contribution which this thesis makes to emerging philosophical perspectives on the

temporal organisation of biological systems. Thirdly, I discuss the relationship between research dynamics and assumptions, specifically the role that discrepancies between assumptions play in eliciting changes in research. In the final section I reflect on the historical implications of this thesis for philosophical perspectives on contemporary biochemical systems biology, and also return to the methodological concerns for research in the history and philosophy of science tradition raised in chapter two section 2.1.1.

7.1 Types of assumptions

7.1.1 Ontological assumptions

Ontological assumptions involve knowledge about metabolic systems which isn't being investigated by the model but is assumed to be correct and used to guide model construction. They can range from general ontological assumptions which are found across different research groups working in an area to assumptions which are specific to particular research groups and questions. This thesis illustrates the importance of ontological assumptions about the temporal organisation of metabolic systems alongside assumptions about their structural and functional organisation. Ontological assumptions influence what goes into the model and what comes out of it, their influence is methodological and explanatory. They can play a role in guiding the construction of models and also underpin the knowledge claims made on the basis of model analysis.

First I will contrast more specific with more general ontological assumptions. Specific ontological assumptions are relevant when a model is constructed in conjunction with a particular theory. In chapter three, I illustrated how metabolic control analysis was constructed in relation to the theory of the rate limiting step. This theory involves an ontological assumption about the functional organisation of metabolic systems - metabolic flux is controlled by a single reaction in a metabolic pathway. This constitutes an assumption as the model was not initially constructed in order to test whether this theory of functional metabolic organisation was correct, it was constructed to determine which reaction in a metabolic pathway constituted the rate limiting step. It is conceivable that a mathematical model of metabolism could have been constructed in relation to a different ontological assumption about

metabolic control. For instance, constructing a model on the basis of end-product inhibition would involve the functional organisational assumption that the end product of the last reaction influenced the reaction rate of the first reaction in a pathway.

More general ontological assumptions tend to be found across multiple different models of the same phenomena. In section 2.3 I cited historical work which identified the defining feature of biochemistry as 'enzyme theory'. Enzyme theory recognises that biochemical transitions from inputs to end products were not carried by a homogeneous protoplasm but involved a series of changes in state mediated by substrate specific enzymes. Initially this would have constituted a specific ontological assumption but over time has become a general assumption found across research on metabolism. The functional and spatial ontological assumptions that biochemical transformations involve intermediary metabolites and enzyme substrate specificity are found across the mathematical models examined in this thesis.

Specific ontological assumptions are more likely to be related to a particular epistemic role for the model than are general ontological assumptions. Out of the models examined in this thesis, metabolic control analysis is the only model constructed with a specific application in mind - determining the distribution of control amongst reactions. Biochemical systems theory and constraint based optimisation, in contrast, were constructed as general models which could potentially play a role in multiple different research questions. It seems likely that all mathematical models will involve general ontological assumptions, but those with a particular epistemic role are more likely to involve specific ontological assumptions.

I have already mentioned that all the models examined in this thesis are based on the ontological assumption that metabolism involves transitions from one biochemical molecule to another via reactions mediated by substrate specific enzymes. This involves the compositional assumption that metabolic systems are constituted by metabolites, enzymes, and other biochemical molecules which play a regulatory role. They are based on the assumption of a spatial hierarchy decomposing systems into their constituent parts. It also involves the functional assumption that enzyme mediated reactions transform one metabolite into another. This knowledge of interactions between metabolites and reactions is used to construct structural metabolic maps. As I explored in chapter four section 4.2.2, the particular ontological assumptions made about stoichiometric relationships between reactions facilitate their representation in mathematical models using simple

algebraic equations. In section 2.2.2 I examined how discussions about ontological assumptions about the relationships between parts and wholes are a common theme in philosophy of biology. As I illustrated in chapter four, William Bechtel used historical cases of biochemists determining the composition and functional relation relationships in biochemical pathways in order to develop his concepts of structural and functional decomposition.

Ontological assumptions relating to whether biological context affects the composition and functional interactions in metabolic systems also play an important role in mathematical models of metabolism. Even though all the models are based on the organisation of metabolic components into metabolic systems, they do not all involve reductionist ontological assumptions about the direction of causality. As I explored in the section 2.3 and in chapter four up until the 1960's research largely focused on determining the relationships between metabolites and enzymes allowing the construction of canonical metabolic pathways. This research was largely carried out through experiments on components *in vitro*, i.e. in isolation from their context in a biological system. During the 1960's to the 1980's it was common for modellers to make the reductionist assumption that the composition and structural organisation of metabolism was independent of biological context. Chapter six explored how, in the case of constraint based optimisation, this changed from the 1990's onwards towards modellers assuming that biological context also affected the composition of the metabolic systems and the functional interactions between components. In contrast, as I highlighted in chapter four, the kinetic behaviour was always considered a context dependant property of metabolic reactions. Whether or not biological context is assumed to influence the behaviour of components has implications for the strategies used to investigate biological complexity which I will explore further in the following section on idealising assumptions and assumptions about data.

Another ontological assumption frequently used in the models is that the rate of metabolic reactions, the number of metabolic transformations carried out over a set time period, is a variable property. As I explored in chapter four, compared to metabolic composition and structure, reaction rate was always treated as something which was affected by the wider biological context. Classical Michaelis-Menten enzyme kinetics involves the exploration of changes in the rate of a reaction in response to changes in substrate concentration. The extent to which modellers

assumed that reaction rate was influenced by biological context varied. In chapter five I highlighted how the more general kinetic orders used in biochemical systems theory contrasted with the very specific elasticity coefficients used in metabolic control analysis. The modellers working on metabolic control analysis assumed that not only was the concentration of molecules directly associated with a reaction important for affecting reaction rate, but that reaction rate was affected by a wider variety of properties in its specific biological context.

There is current philosophical debate about the relationship between biological research involving the analysis of part whole relationships and biological research involving mathematical modelling of dynamic behaviour (See section 2.2.2). In chapter four I argued that constructing mathematical models of metabolism often involves making ontological assumptions about the temporal organisation of metabolic processes alongside those about the spatial organisation of components. Firstly, as I discussed in section 4.2.2 of chapter four the ontological assumptions that researchers make about reaction rates necessitate the use of differential equations to represent this behaviour in mathematical models. Secondly, modellers assume that metabolic processes constitute a time hierarchy ordered according to the speed with which properties of the system – reaction rates, and metabolite concentrations - are changing. Highlighting the significance that ontological assumptions about the temporal organisation of metabolic systems play in the construction of mathematical models is a significant extension to philosophical perspectives on the role of assumptions in biology. Philosophical work on the role of assumptions in mathematical models reviewed in chapter two (2.2.1) tends to emphasise the idealising role of assumptions about dynamic behaviour, however my analysis indicates that these assumptions also have an ontological dimension. I will return to this in the discussion in section 7.2.2 where I focus on the relationship between ontological and idealising assumptions about the temporal organisation of metabolic systems.

The chapters in this thesis predominantly focused on the development of mathematical models, rather than the interpretation of results and their use in substantiating biological explanations. The ontological assumptions which I have explored so far largely play a role in the construction of the mathematical models examined in this thesis and I will pick up on these themes in the following sections on idealising assumptions and assumptions about the data resources used during

model building. However, the case studies also illustrate several occasions when ontological assumptions associated with models also play a significant role in the interpretation of the outcomes of model analysis. For instance, metabolic control analyst's assumptions that the elasticity rate of reactions was context dependent supported the systemic theoretical perspective on metabolic control which emerged out of the model (see chapter three). Also, as I explored in chapter six, as constraint based optimisation started to be based on the ontological assumption that the composition and structure of metabolic systems was context dependant the model began to be used to explore context specific, rather than general, spaces of metabolic possibility. I will discuss these cases further in section 7.2.3 on dynamic relationships between assumptions. Another set of ontological assumption which affected the interpretation of results from the mathematical models are assumptions about the relationships between genes and metabolic reactions. In chapter three, on metabolic control analysis, and chapter six, on constraint based optimisation, researchers often assumed a direct relationship between genetic composition and metabolic composition. This assumption allowed them to use models which had been constructed to analyse metabolic systems to draw conclusions about the relationships between genotype and metabolic behaviour.

7.1.2 Idealising assumptions

In chapter two of this thesis I discussed idealising assumptions as those which play a role in making a working version of the model possible. By working version, I am referring to the construction, calculation, and interpretation stages of mathematical modelling. Philosophical attention on mathematical modelling has focused on simplifying assumptions, those which involve a reduction in the complexity of the phenomena being modelled facilitating model construction, mathematical tractability, and the interpretation of outputs. The cases in this thesis also illustrates that connecting assumptions are another significant type of idealising assumption which makes models of metabolism possible. These are assumptions which connect the model to particular inputs and particular outputs, for instance enabling modellers to use certain data resources during the construction of the model and make certain epistemic claims during the interpretation of modelling

results. Ontological assumptions frequently play a role in motivating the use of, and justifying the use of, particular idealising assumptions. However, modellers also make idealising assumptions which do not correlate with their ontological assumptions in order to get the model to work.

Bechtel and Richardson's notions of structural decomposition and functional decomposition involve the use of assumptions which facilitate the investigation of complex biological systems. As I explored in chapter four, they are based on the idealising assumption that biological systems are nearly decomposable, there are more significant connections within components than between components. This assumption allows researchers to delimit the horizontal boundaries of the phenomena, those at the same spatial scale of organisation, by performing structural decompositions identifying the relevant parts, and functional decompositions, identifying the relevant operations. This facilitates a reduction in the number of components, and the number of functional operations (connections between components), which they need to account for in their investigations. Most of the modelling approaches examined in this thesis were attempts to build large scale models of metabolic systems, rather than small scale models of metabolic pathways. As such, the use of idealising assumptions to reduce the number of components parts, metabolites and enzymes, or the number of connections between component parts, does not play a significant role. However, the assumption of near decomposability also plays a role in delimiting the vertical connections which need to be included in researcher's investigations, as it reduces the significance of the wider biological context in which the component parts and functional operations are located. All the models involve this assumption to some extent. Firstly, they involve the descriptions of the properties of components in the calculation of systems level behaviour. Secondly, they are all models of metabolic systems which do not involve detailed mathematical representation of the wider cellular context in which they are located.

In chapter four I argued that temporal decomposition is a third type of simplifying strategy which should be recognised in addition to structural and functional decomposition. I also offered an interpretation of Herbert Simon's work on nearly-decomposable systems illustrating the neglected importance of assumptions about temporal as well as spatial organisation in previous work on simplification and complex systems. Researchers modelling metabolic systems are often interested in

investigating their dynamic properties rather than just developing static representations of metabolic components and their operational relationships. However, detailed mathematical models of metabolic dynamics ran into two problems. Firstly, the equations were often intractable, and secondly they required large amounts of kinetic data which wasn't readily available. Using three different modelling approaches I illustrated how modellers reduced the number of dynamic variables they needed to include, simplifying the models in order to get around these problems. The importance of these types of simplifying strategies in mathematical models of dynamic behaviour has already been highlighted in philosophical literature (see section 2.2.1). In chapter four I argued that researchers assumed that metabolic process constituted a time hierarchy according to the speed with which properties of the system - reaction rates and concentrations of components - changed. They then assumed that they could remove variables at either extreme of the time hierarchy, those which changed extremely slowly or extremely quickly, from their model. This allowed them to retain the dynamic variables which were relevant for the phenomena of interest. The reduction in the dynamic variables enabled working versions of the model to be produced by reducing the complexity of the mathematical representation of the system and the amount of kinetic data required. I will discuss the relationship of this to ontological assumptions about the temporal organisation of metabolic systems in section 7.2.2.

Researchers may make simplifications on the basis of temporal decomposition by appealing to ontological assumptions. However, sometimes modellers make assumptions which allow them to reduce the complexity of the mathematical representation of systems dynamics which go against their background ontology. For example,metabolic control analysis involves the assumption that the dynamic behaviour of individual reactions and systems is linear, i.e. there is a direct correlation, say, between the rate of a reaction and the concentration of a substrate. However, this goes against their background knowledge that actually dynamic behaviour is non-linear. Researchers have to acknowledge that this assumption is, and correspondingly the results of the model are, only valid over very small changes in conditions. This distinction between simplifications based on ontological assumptions and model building processes involving deviations from ontological assumptions has been referred to as a distinction between abstraction and idealisation in the philosophical literature.

Idealised models have been defined as those which involve a distortion which is known to be inaccurate, whereas abstraction involves a reduction in the complexity of the model whilst maintaining accuracy in what is included in the representation (Cartwright, 1994; Morrison, 1999). Whilst my definition of idealising assumptions as assumptions which facilitate the construction of working versions of models refers to both these processes, as the previous example illustrates, I still appreciate that they are distinct strategies used during model construction.

The idealising assumptions made in temporal decomposition are simplifying and connecting. They simplify the mathematical representation of systems dynamics facilitating mathematical tractability, at the same time as reducing the amount of data required by the model allowing it to be connected to kinetic data resources as a material input for model construction. Other idealising assumptions invoked by modellers play a primary role of connecting the mathematical model to particular data resources. These assumptions facilitate the production of working mathematical models through framing relevant data resources as available for, rather than constraining, model production. As I will explore further in the next section 7.1.3, they play an important role when data is assumed to be a key ingredient in transforming abstract mathematical models into models of actual metabolic systems. For instance, in chapter four and chapter six, I explored researcher's assumptions that the composition and structure of metabolic systems were context independent, even physico-chemical properties. Palsson in particular invoked this assumption to support the use of existing stoichiometric data to model the metabolic networks of not just of established model organisms, but also organisms on the periphery of existing research. Palsson's assumptions about a straightforward relationship between genetic sequence and metabolic reactions similarly facilitated his arguments that genome sequence data could be used as the basis for network reconstruction. Palsson's assumption not only facilitated the use of genome sequence data to build the model, but they also allowed the interpretation of the model's output to be connected to claims about the relationship between metabolic genotype and metabolic phenotype. Other idealising connecting assumptions are used to mitigate deficiencies in the data used to build the model and the mathematical representation of the system and connect its outputs to claims about the world which would otherwise be difficult to substantiate. For instance, in chapter four I examined how constraint based optimisation invoked an optimisation

assumption in order to make up for the lack of kinetic data and mathematical representation of dynamic behaviour and still be used to analyse the distribution of metabolic flux under particular conditions.

7.1.3 Assumptions about data

All the mathematical models examined in this thesis involve the use of biochemical data to determine the values of some of the model parameters. Whilst these assumptions are often tightly connected to ontological and idealising assumptions they also involve features which do not easily fit in these categories. The chapters in this thesis illustrate how assumptions about data have played a major role in the historical development of mathematical models of metabolic systems and will no doubt continue to do so in the context of data intensive science. The use of biochemical data is affected by assumptions in two ways. Firstly, assumptions have implications for the evaluation and use of data resources. I have touched on this relationship several times in the previous discussion of ontological and idealising assumptions. Secondly, mathematical modellers make assumptions about the value which biochemical data adds to the credibility of a model and its outputs. More specifically, they often assume that the use of biochemical data to inform the values of model parameters is what makes the model a model of a specific metabolic system rather than just a general mathematical representation. My examination of the role of data in the construction of mathematical models echo's Marcel Boumans' argument that often models aren't evaluated after their construction, but that justification is built in through the application of standards for the ingredients used in model building (Boumans, 1999).

Researchers make assumptions which have implications for the use of particular data resources and these are often related to the ontological and idealising assumptions introduced above. Sometimes assumptions help researchers work around a deficit in a particular data type required to build a model. In chapter four I explored how researchers used assumptions about the temporal organisation of metabolic systems in order to reduce the number of dynamic variables included in the mathematical model and accommodate the lack of available kinetic data. Other times, researchers' assumptions affect their perception of the value of a particular

type of data for the mathematical reconstruction of the metabolic network. In chapter six I examined how Palsson's initial assumptions that the relationship between gene sequence and metabolic reactions was context independent facilitated the reconstruction of metabolic networks on the basis of genomics data sets. Over time the relationship between gene sequence and metabolic reactions came to be seen as something which was affected by the wider organismal and environmental context and genomics data was no longer regarded as a reliable source of evidence for metabolic network reconstruction. I will explore this transition further in the following section 7.2.2 on dynamic relationships between assumptions.

To some extent the modellers all assume that using biological data transforms the model into being a model of a particular metabolic system rather than just an abstract mathematical representation of metabolic systems in general. In chapter five I explored the relationship between metabolic control analysts and experimental biochemists. Metabolic control analysts assumed that if the mathematical model was based on biochemical data sets which had been collected *in vivo* or in *in vivo* like conditions then it was not an abstract model but something which was relevant for analysing actual metabolic systems. However, some experimental biochemists disagreed, they thought that the general mathematical form of the model ironed over biochemical details and differences between metabolic networks and this generality was not mitigated by the biological specificity of the data the model was used to analyse. Chapter six provided another illustration of modeller's making the assumption that biological data, rather than the mathematical form of the model, is where biological specificity is located and introduced. In this case, as modeller's assumptions about the context specificity of the composition and stoichiometry of metabolic networks changed they sought to accommodate this change in perspective into the model through modifying the types of biochemical data used during model construction rather than the underlying mathematical formulation of the model.

Some mathematical models involve the assumption that biochemical data can be used to represent the influence of the wider biological context of the metabolic system in the model when this is not incorporated in the mathematical representation of the metabolic system. As RheinbergerRheinberger (1997b p274) argues in relation to experimental systems, even though the mathematical models are simplified versions of metabolic systems, the epistemic context in which they are

located retains a connection to complexity of the systems they represent. For instance, even though constraint based optimisation does not involve a mathematical representation of the wider biological context of the metabolic system in the later stages of model development modellers assume that by selecting data resources which have been collected in specific contexts the model can represent context specific metabolic behaviour. There are some instances of changes in researcher's assumptions leading to changes in the mathematical framework. For instance, in chapter six I also discuss how researchers responded to challenges to their assumption that all metabolic systems were optimising by developing new algorithms for searching through possible metabolic flux distributions which would select sub-optimal solutions.

Modellers' ontological and idealising assumptions also play a role in researchers' assumptions about what kind of biochemical data is required to make the model a numerical reconstruction of a particular biochemical system. In chapter five I examined the conflict between biochemical systems theorists and metabolic control analysts over this point. Biochemical systems theory constructed general models on the basis of a few experimental data points informing the description of kinetic behaviour. Metabolic control analysts accused biochemical systems theorists of building abstract models with little relationship to actual biochemical systems. They argued that their model, which required detailed experimental data sets obtained under *in vivo* or *in vivo* like conditions, bore a far stronger relationship to real metabolic systems. Chapter six also illustrates this theme, as researchers' ontological assumptions changes so too did their assumptions about what kind of data was required to build high quality metabolic network reconstructions which bore a close resemblance to actual metabolic networks.

I want to emphasise the significance of assumptions about data and their correlations with ontological and idealising assumptions to signal the agency of mathematical modelling in relation to biochemical data sets. In chapter six I argued for the significance of the context of use of biochemical data for understanding its epistemic role in research. Yet mathematical modellers may emphasise particular ontological assumptions, or utilise idealisations in order to accommodate particular types of data into a model. However, mathematical modellers also want something from the data; they want it to meet their standards for building a sufficiently accurate model of a metabolic system. In the case of metabolic control analysis, and the later

stages of constraint based optimisation, the demands placed upon the role of biochemical data are extremely high. The biochemical data is used to provide a relationship with a much wider biological context than is actually represented in the mathematical reconstruction of the metabolic system and its components. Mathematical model building of biochemical models is a process which is responsive to available biochemical data, and also selective and discerning about the use of biochemical data even when theories and hypothesis are not part of the model building process. In all of the cases examined in this thesis the availability of particular data types and researchers' assumptions and corresponding evaluations of that data plays a central role in the trajectory of model development. Current debates about data intensive science have sometimes polarised around distinctions between the dominant role of hypothesis or data in research. Paying attention to assumptions related to the data used in mathematical modelling shows that mathematical models are a powerful influence mediating between the role of data and hypothesis in the analysis of biochemical systems.

7.1.4 Researchers Commitments

Researchers' commitments were another category of assumptions that had a significant impact on the development of the models of metabolism examined in this thesis. By researchers commitments I am referring to what they assume to be best practice in a particular area of research. This research found that researcher's commitments have two significant aspects. Firstly, what they are committed to - methods, theories, standards etc. This aspect of commitments is often intertwined with researcher's ontological and idealising assumptions and assumptions about data. Secondly, the characteristics of their commitments to these things - are the commitments rigid or flexible, are researchers tolerant or intolerant of inter-cultural differences? Examining researchers' commitments as a particular type of assumption is of particular importance in areas of research involving different research cultures coming into contact with each other. Chapters five and six focused on the important role played by researcher's assumptions about how best to do research. The content and characteristics of researcher commitments were shown to play an important role when two groups of researchers with different commitments

came into contact with each other. These encounters laid bare the commitments, which otherwise often remained hidden in the background of research, as researchers had to acknowledge contrasting ways of doing things. As such, I review them briefly here, but they have a prominent position in my discussion of the role of assumptions in relation to research dynamics in section 7.2.

The first aspects of commitments are often related to the ontological and idealising, and assumptions about biochemical data explored above. Researchers held commitments to particular theories, for instance the theory of the rate limiting step, or a systemic perspective on metabolic control. Commitments to particular research methods were another prominent category. Experimental biochemists were committed to organism specific expertise gained through experimental interventions, metabolic control analysts to mathematical models which had a specific epistemic role, and biochemical systems theorists to general models of metabolic systems which could be used in the context of a variety of research questions. Methodological commitments often involved attachments to particular styles of research involving qualitative or quantitative approaches and systems specific or general analysis. These intersect with particular research standards, for instance those relating to the type of biochemical chemical data considered necessary to build an appropriate reconstruction of a metabolic network. Whereas biochemical systems theorists assumed that general descriptions of the dynamic behaviour of reactions could be based on a few experimental data points, metabolic control analysts were committed to developing novel experimental methodologies in order to collect data about the dynamic behaviour of metabolic systems and reactions which were specific to particular metabolic contexts.

Chapter five introduced a conceptual framework highlighting the pertinent characteristics of researcher's commitments. A first feature was whether researchers held commitments which were rigid and specific or flexible and broad. This is in a sense a reflection of the strength or intensity of researcher's assumptions about best practice. Do they have very fixed ideas about how research should be carried out, or do commitments act more as a loose guide giving researchers a minimal structure within which to carry out research. The second characteristic was whether researchers perceived differences with the commitments of other research cultures as negative, neutral, or positive. In other words, did they perceive alternative ways of doing things as threatening to their mode of research, did they think of themselves

as being immune to the influence of other cultures, or did they actually perceive these differences as being a potentially useful source of inspiration and innovation.

7.2 Philosophical implications

In the previous section I highlighted the expanded taxonomy that this thesis offers on the types of assumptions which play an important role in developing mathematical models of metabolic systems. Alongside ontological and idealising assumptions about the relationship between biological parts and wholes, this thesis also highlighted assumptions related to the temporal organisation of biological systems, biological data resources, and researchers' commitments to certain research practices. In this section I focus on the philosophical implications of this perspective on assumptions. Firstly, I offer a re-evaluation of the role of reductionism in mathematical modelling. Secondly, I focus on the contribution of this thesis to emerging philosophical perspectives on biology and temporal organisation. Finally, I examine the relationship between assumptions and research dynamics.

7.2.1 Biological data: Reassessing reductionism in modelling

The significant role played by assumptions about data has implications for assessing whether a mathematical model involves reductionist assumptions. In section 7.1.3 I gave an overview of the different roles that assumptions relating to data played. I argued that ontological and idealising assumptions could be used to mitigate a deficit in particular data resources, or to build connections with particular data resources which could be used during model construction. What is notable is that all the different groups of modellers to some extent want to establish a role for biochemical data in the construction process. This leads to the second important assumption; that the use of biochemical data to inform model construction adds something. Biochemical data are treated as powerful ingredients for mathematical models, it is what makes the model a model of something in particular, and it is what ties the model to a real world system. Ontological assumptions come into play again; they affect the standards for the kind of data which is required to support claims that the model is a valuable reconstruction of a real world system. These assumptions

and standards can play a crucial role in affecting whether researchers consider the model to involve a reductionist representation and analysis of the metabolic network.

In section 2.2.2, I discussed reductionist assumptions about the relationship between parts and wholes and their role in mathematical modelling. Reductionist approaches often involve the assumption that the properties of the system can be understood in terms of the properties of its component parts. Mathematical models of biochemical systems based on the bottom up approach, calculating the properties of systems from data about their components, are often considered to involve a continuation of reductionist research strategies (O'Malley and Dupré, 2005). Some of the case studies examined in this thesis show how differences in researchers' assumptions about data affect the accuracy of this assessment.

All the case studies examined in this thesis are based on a bottom up strategy; they all involve data about the properties of components in their calculation of systems level behaviour, frequently based on the reconstruction of a metabolic network. Some of them involve a much smaller quantity of systems level data. In the case of constraint based optimisation this is used to mitigate the lack of data about the kinetic properties of individual reactions, i.e. it assumed that if all the components could be described kinetically then data about systems level behaviour would not be required to investigate the distribution of metabolic flux (chapter six section 6.2.3). Even so, none of the models involve extensive mathematical modelling of the wider biological context in which the mathematical model is located. This approach reflects necessary idealising assumptions, building detailed multi-level mathematical models is complex and would encounter multiple constraints. In some cases, for instance, biochemical systems theory, it also reflects the ontological assumptions that the properties of components are context independent and the properties of biochemical systems are caused by the properties of their component parts. This additional ontological assumption tends to be associated with the construction of very general mathematical models as the properties of systems components are assumed to be consistent across different investigative contexts. Correspondingly, a model based on data about component parts can be used as a general tool to investigate multiple different features of metabolic systems (See my discussion of biochemical systems theory in chapter five, and the early stages of constraint based optimisation in chapter six).

In other cases, researchers hold very different ontological assumptions about the relationship between the properties of component parts and the properties of metabolic systems. This was particular apparent in chapter three in my analysis of metabolic control analysis and chapter six in my analysis of the later stages of constraint based optimisation. In these instances mathematical modellers assumed that the properties of component parts were related to the wider biological context in which they were located. In order to reflect this ontological assumption in the mathematical representation of the system they had stringent requirements for their evaluation of biochemical datasets. It was assumed that the biological data could be used to reflect this relational perspective in the model if the model was constructed from data sets which related to properties of those components in particular conditions. Even though the models still involve reductionist idealising assumptions, they still avoid technical constraints such as mathematical intractability through simplifying the model of the metabolic system, biological data is used to ensure that the model still reflects researchers' ontological assumptions that the properties of component parts are context dependant. Models constructed under these sets of assumptions tend to be more limited in their scope and are used to investigate the properties of metabolic systems within particular conditions. As I discussed in chapter six, as researchers developing constraint based optimisation began to assume that the properties of metabolic components were context dependant, the models began to be used to explore more limited spaces of metabolic possibility under specific conditions.

7.2.2 Temporal organisation: Ontology or idealisation?

In chapter four I argued that temporal decomposition should be recognised as an additional strategy, alongside structural and functional decomposition, facilitating the investigation of complex biological systems. The role of temporal organisation in biological systems and biological investigations has frequently been overlooked as a significant feature of methods and explanations in biology. This thesis contributes to current literature in philosophy of biology attempting to explore its importance. Bechtel and Richardson's structural and functional decomposition are widely regarded as related to ontological assumptions about the organisation of biological systems, and idealising assumptions facilitating the investigation of biological

complexity. They are based on the ontological assumption that systems are partially-decomposable, there are more significant relationships within sub-components than between sub-components, and this facilitates idealising assumptions allowing sub-components to be studied in isolation for their wider context. In response to the submission of a version of chapter four to Studies in the History and Philosophy of Biological and Biomedical sciences part C, an anonymous reviewer commented that my proposal:

> 'should not really be thought of as a third variety of decomposition, analogous to B & R's structural and functional decomposition. … My reason for this is that I think that the first two strategies have a clear ontological reading while the third does not. Using Bechtel and Abrahamsens's terms, structural decomposition identifies components, while functional decomposition identifies operations. But what does temporal decomposition identify? … My concern here is that variables aren't like components or operations, features of the world, but are instead features of our models. "temporal decomposition" is from an ontological point of view, simply a refinement of the characterisation of operations (and their organization). There is no third kind of thing which is being decomposed.'

The reviewer is firstly questioning whether temporal decomposition is based on an ontological assumption. Whereas the reviewer seems happy to accept that structural and functional decomposition relate to the organisation of components and operations in the world, they are unsure that temporal decomposition relates to a feature of the world. Secondly, leading on from this, they suggest that temporal decomposition, is in fact a process of refining the characterisation of operations in models. It is an idealising assumption facilitating the construction of a working version of a mathematical model not to the organisation of biological systems. As the reviewer considers what I have termed temporal decomposition, a means of refining the mathematical representation of operations, they question whether it should be classified as a decomposition of a biological system.

I am not going to attempt to give an exhaustive answer to the question of whether temporal decomposition involves ontological and idealising assumptions about the temporal organisation of metabolic systems. Doing so would involve an extended metaphysical analysis of the relationship between biological components, operations, and processes and that is not what this thesis is about. I will however

point out a few reasons why I want to leave open the possibility that temporal decomposition involves correlating ontological and idealising assumptions in the same way that structural and functional decomposition do. Firstly, metaphysical positions exist in which processes are regarded as a feature of the world, and sometimes a more primary feature of the world than components (See my discussion of Dupré's and Griesemer's work on processual temporal organisation of biological systems in section 2.2.3). Secondly, as I discussed in chapter four, Simon's initial work on the partially decomposable organisation of biological systems, temporal, spatial, and functional organisation play overlapping roles in his account of near decomposability. Simon does not give a different epistemic status to these three features. Thirdly, the mathematical modellers examined in chapter four frequently discuss temporal decomposition as an idealising strategy based on ontological assumptions, they claim that 'Time hierarchies are a general feature of nature.' (Heinrich et al., 1977 p20).

I am going to defend why what I describe in chapter four is 'decomposition' not 'refinement', regardless of its basis on ontological or idealising assumptions. My argument involves returning to the distinction between philosophical attention to causal temporal organisation versus processual temporal organisation which I made in the thesis introduction. Firstly, I will make a distinction between decomposition and refinement. Decomposition involves assumptions which reduce the complexity of the system being researched through facilitating decisions about isolating and removing things which do not need to be taken into account to understand the phenomena in question. Refinement involves making assumptions which facilitate adjustments to the aspects of the system which researchers have already decided they need to take into account in order to understand the phenomena in question. For instance, researchers might refine aspects of a model in order to get the outputs of the model to match relevant experimental data sets. I propose that 1) refinement of the variables included in a mathematical model is aligned with an interest in causal temporal organisation, whereas 2) decomposition of variables in relation to a mathematical model is aligned with an interest in processual temporal organisation. I will argue that what I propose in chapter four sits more comfortably with the second grouping, decomposition and processual temporal organisation, than the first.

As I defined in section 2.2.3, philosophers who are interested in causal temporal organisation, are interested in the temporal ordering of operations taking

place between parts and wholes. Bechtel and Abrahamsen (2010) explicitly focus on temporal organisation from the perspective of recomposing rather than decomposing biological systems. They describe the recomposition of a system as involving the refinement of the spatial and temporal organisation of components and operations during the process of putting the system back together. Temporal organisation comes into play in relation to the causal temporal ordering of operational relationships between component parts. What I describe in chapter four is an assumption about the temporal organisation of metabolic processes which facilitates the decomposition, the isolation and removal of aspects of the model, not the refinement of components, operations, and variables that researchers have already decided need to be included in the model. In chapter two section 2.2.3 I described philosophical attention on processual temporal organisation as that which involves a perspective on the temporal organisation of processes which in some way involves rethinking the dominant part-whole perspective on biological systems. Dupré suggested that there is a hierarchy of biological processes occurring at different speeds. His concept of process is clearly distinct from Bechtel and Abrahamsens concept of an operation. Bechtel and Abrahamsen refer to the functional relationships between parts and between parts and wholes as operations. Dupré develops an idea of process which is used to challenge the clear distinction between objects, such as parts and wholes, and processes, and even between structure and function itself. For Dupré whether something constitutes a process or a static entity depends on the scale of the temporal hierarchy which is being taken into account (Dupré, 2013; Bapteste and Dupré, 2013). My account of temporal decomposition doesn't exactly overlap with Dupré's distinction between processes and objects, but it is comparable. In chapter four I described how researchers appeal to a time hierarchy in order to decompose the system by isolating variables which can be removed from the system and treated as constants within the particular timeframe that system is being examined.

7.2.3 Assumptions and research dynamics

This section explores the relationship between assumptions and research dynamics, specifically the role that inconsistencies in the different types of assumptions in affecting the tempo of research. Sometimes inconsistencies between

assumptions go unnoticed or are considered unproblematic by researchers. For instance in section 7.2.1 I argued that the idealising assumptions mathematical modellers make about the relationships between parts and wholes may be different from their ontological assumptions. The chapters of this thesis show that when these inconsistencies become visible and significant they affect research dynamics, either because they prevent research being carried out, or because they lead to transformations in the research assumptions associated with the model. As such it explores different tempos in the development of mathematical models ranging from stagnation to innovation. Even though assumptions constitute what is taken for granted, or what is already known, by modellers, even though they are not the focus of investigation and discovery, the modelling process often leads to their re-evaluation and re-formulation.

The case studies examined in this thesis all investigate the process of developing mathematical models, rather than the use and application of established modelling procedures. They cover a period of time in which novel mathematical methods and methods of data generation and collection are being developed, and groups of mathematical modellers and laboratory based biochemists are being bought into contact with each other. As such the chapters in this thesis frequently involve occasions in which disjunctures between research assumptions occur, and research assumptions are being exposed, challenged, held on to, or reformulated. Two distinct modes of dynamics involving relationships between research assumptions are highlighted by these case studies which I review below. Firstly, as illustrated in chapter three, researchers may unintentionally bring conflicting assumptions together in their research approach. Secondly, as illustrated in chapters five and six, the process of developing mathematical models often brings groups of researchers together who adhere to distinct sets of assumptions and their responses to this elicits particular types of research dynamics.

Chapter three developed a perspective on what it means for a model to have a partially independent and autonomous relationship with theory by analysing the development of metabolic control analysis. In this case Joseph Higgins built a mathematical model on taking the theory of the rate limiting step as a theoretical ontological assumption about the functional organisation of metabolic control. He made the theoretical assumption that control would be located in one reaction in a metabolic system, and produced a mathematical model in order to facilitate the

process of interpreting biochemical data and identifying this reaction. He also introduced several idealising assumptions in order to simplify the model making it mathematically tractable and connect it to biochemical data resources. These led to a partially independent relationship between model and theory. The model was based on reflection coefficients. These assumed that the influence of a reaction on a system was accounted for by changing the concentration of that reaction in the system and recording the impact on systems flux. They also assumed that the relationship between the change in concentration and the reaction of systems level behaviour would be linear. These idealising assumptions did not exactly overlap with the theoretical assumption as they left open the possibility that more than one reaction could have a role in metabolic control.

This had unanticipated consequences for the theory of metabolic control which is why I argued the model can also be described as having an autonomous relationship with emerging theoretical perspectives. The possibilities open up through the disjuncture between the theoretical and idealising assumptions involved in the model were exploited by researchers interested in developing a systemic perspective on metabolic control. They theory of the rate limiting step was transformed from being a theoretical assumption to being a theory which wasn't assumed but was being challenged and investigated using the model. Henrik Kacser and colleagues used the model to investigate whether multiple reactions had a role in metabolic control and developed the model further, introducing the summation theorem, which assumed that the contribution of a reaction to metabolic control was dependent on the particular systemic context which the reaction was in. In this case researchers unintentionally introduce conflicting assumptions during the model building process. Once these differences are acknowledged attempts to resolve them can lead to fundamental changes in the assumptions associated with model. This illustrates a case of a conflict where the differences between assumptions are located within those of a particular community of researchers.

Researchers commitments, their assumptions about how research should be done, have a major role in the development of mathematical models when that process leads research groups with different sets of commitments to interact with one another. I developed a framework for understanding the different dynamics which could emerge out of these interactions in chapter five where I looked at disagreements which took place between researchers associated with different

modelling approaches and between modellers and biochemists in the 1980's. This theme also played a significant role in chapter six where I argued that social interactions also played a crucial role in facilitating the development of mathematical models and their expanded use in the context of large biological data sets. As I footnoted in chapter five, whilst I chose to situate this analysis in relation to work by Galison and Nersessian in order to explore the relationship between constraints, problem spaces, and social interactions, and interesting extension would be to explore this material in relation to Helen Longino's (2002 p129) account of the importance of 'critical discursive' social interactions in knowledge production.

In the 1980's groups of modellers became frustrated with the lack of uptake of their models. This frustration bought them into contact with different groups of researchers as they tried to promote the use of the particular technique they had developed. These interactions between different groups exposed disagreements in their research assumptions. I examined two instance of this. Firstly, metabolic control analysts assumed that basing a model on data which had been collected under conditions which were sensitive to biological context was sufficient to make the model a model of an actual metabolic system. However, experimental biochemists did not accept the assumption that sufficient attention to biological particularity could be achieved through placing systems specific biochemical data in the context of a general mathematical framework. Secondly, biochemical systems theorists were committed to building a general mathematical model of metabolic systems which could be used in the context of multiple research programs. This clashed with the commitment of metabolic control analysts to the development of mathematical model in relation to the investigation of the distribution of metabolic control. These differences led to years of hostility and aggressive attacks on each other's work and did not achieve the initial aim of seeking these points of contact in order to increase the number of researchers using and developing these approaches. I argued that the crucial factor was the quality of the researchers' commitments. The groups all held very rigid and specific commitments and they felt threatened by the different commitments of other groups of researchers. This led to situations where researchers externalised the problems they were encountering blaming other groups of researchers for not moving research forward. This contributed to a relatively stagnant period of research and development surrounding mathematical modelling of metabolic systems.

As I explored in chapter six, and using McLeod and Nersessian's work in chapter five, the characteristics of interactions between different groups of researchers are radically different in the current context of data intensive systems biology. Researchers developing the constraint based optimisation technique started with certain overlapping ontological and idealising assumptions about the context independent relationship between genome sequence data and metabolic reactions and reaction stoichiometry. These assumptions initially facilitated the expansion of the model as a technique which could be used to investigate the functional systemic properties of metabolic networks on the basis of genome sequence data. However, this expansion bought researchers into contact with constraint based optimisation from multiple different research backgrounds bringing different sets of assumptions about metabolic systems and biological data sets. This is connected to a shift to assumptions that the relationship between genome sequence and metabolic reactions and reaction stoichiometry is context specific and simultaneous changes in the type of data considered to constitute adequate evidence for network reconstruction. The importance of social interactions between diverse groups of researchers in this dynamic area of research has been recognised through the formalisation of interactions in community jamborees to streamline and establish temporary coherence across diverse research programs associated with the modelling technique. In these instance researchers commitments have the qualities of being loose and flexible and having a positive regard of differences in the commitments of research communities. This allows the development of constraint based optimisation to exhibit ongoing creativity as researchers show ‘epistemic pragmatism’ (MacLeod and Nersessian, 2013c) and are willing to change their commitments and listen and accommodate the concerns of other interested groups of researchers engaging with the modelling approach.

7.3 Historical implications

The presentation of case studies exploring the history of mathematical models of metabolism, how the development of models unfolded in time, was integral to all the chapters examined in this thesis. In this section I will bring together how this thesis illustrates the particular contribution which historical case studies make to philosophy of science, and how it contributes to the history of data intensive systems

biology. I also highlight how the research carried out in this thesis contributes to methodological discussions about the role of historical case studies in philosophy of science discussed in section 2.1.1.

7.3.1 A close up history of mathematical models of metabolism

As I have already mentioned, more general philosophical perspectives on the history of biochemical models have often been framed around identifying two or three distinct roots, including a top-down and a bottom up approach. The models examined in this thesis often do not fit exactly into either of these categories, but sit relatively close to one or the other. The approach of relaxation times lies close to pathway analysis in that it stems from models which involve a full kinetic description of reactions. Constraint based optimisation lies closer to network analysis in that it is primarily based on structural data, but it attempts to use this to analyse the dynamic property of metabolic flux distribution.

The perspective in this thesis suggests two alternative vantage points on the history of mathematical biochemistry which exposes the way in which the roots of systems biology can sometimes appear blurred. Firstly, in chapter two, 2.1.1 I highlighted Schikore's comment that, in the current context of diverse case study approaches to philosophy of science in practice, a case for the particular value of historical case studies for philosophical analysis needs to be made. The analysis in chapter six showed how the relationship between modelling approaches, ontological assumption and the use of particular data sets can be continually shifting and depends on the point in time from which it is viewed. This kind of insight is brought to light by a historical methodology which examines the details of how scientific practices unfold in time. Secondly, the proposal in chapter four of temporal decomposition provides another alternative to categorising models into the distinct groups of top-down/structural decomposition, and bottom up/functional decomposition. Temporal decomposition involves making assumptions which facilitate the removal of dynamic variables from the model on the basis of a time hierarchy of metabolic reactions. As I illustrated in chapter four section 4.1 this can be coupled with a top-down or a bottom-up approach. Bernhard Palsson’s research trajectory in particular illustrates an intertwined relationship between detailed kinetic

modelling and network analysis. He began by building detailed kinetic models of red-blood cells using the analysis of relaxation times to reduce the dynamic variables included in the model, he then shifted to developing the constraint based optimisation technique in response to constraints imposed by the lack of available kinetic data, in his most recent work he is now attempting to integrate these two approaches to build dynamic models of genome scale metabolic networks (Joshi and Palsson, 1989; Varma and Palsson, 1994; Jamshidi and Palsson, 2010).

Rather than trying to capture all the philosophical and historical potential diversity of mathematical modelling in biochemistry, focusing on the details of the case studies has facilitated one its significant historical implications. History and philosophy of biology has tended to focus on the 'main events'. There has been ample analysis of areas in the spotlight, firstly molecular genetics and now data intensive systems biology. However, scientific research does not develop in a linear series of theoretical and methodological breakthroughs. Following on from James Griesemer's work which I explored in the introduction to my methodology (Section 2.1.2), this thesis has examined the dynamics of an area of research going on in the background, before particular set of circumstances arose in which some of this work became mainstream. It has explored the development of mathematical models by relatively small groups of researchers before they became high-profile research tools in biochemical systems biology. Prior to the advent of data intensive systems biology Palsson retrospectively commented that 'systems analysis in biology in the early 1980's was seen as a "dead-end career" and "professional suicide."' (Palsson, 2006 pX). It is important to note that the popularisation of mathematical modelling in biochemistry in the context of data intensive science might not have happened. Much of the work going on in the background of science may eventually peter out, the circumstances may never arise which increase the focus on it.

The period of background work examined in the case studies, from the 1960's to the early 1990's, is an era of creativity and innovation. Alongside the development of novel mathematical modelling techniques, the case studies also touched on the development of new experimental methods, and a new theory of metabolic control. Chapter three provided another example of how philosophical analysis grounded in historical case studies is important when philosophical concepts, in this case partial independence and autonomous agency, pick out transient characteristics and relationships which exist at particular stages of research. The researchers examined

in this thesis were coming up with innovative uses for mathematical models of metabolism, creative responses to constraints including the availability of biological data and the mathematical tractability of the models. This thesis has only examined the history of four different types of modelling approach, the number of mathematical models of metabolism being developed at this time is far greater (Garfinkel, 1969; Heinrich et al., 1977; Wright and Kelly, 1981). This is a period of research where multiple small groups of researchers are pioneering a wide diversity of approaches to modelling metabolic systems.

However, this is also a period when none of these models are 'taking off', none of the approaches are attracting the critical mass of researchers and attention needed to propel them into the mainstream (Powell et al., 2007). In chapter five I examined one of the important factors which maintained the isolation of the different modelling communities, the characteristics of researchers' commitments. Modellers often held very rigid research commitments to 'their way' of doing things and remained closed to adapting their methods to accommodate the contributions and concerns of different communities of researchers. In fact they were often threatened by the different research commitments of these different communities and blamed them for a lack of progress in research. In chapter six I examined what happened when one of these models did become a widely used approach in the context of data intensive biology. This increase in use of this model occurred for two reasons. Firstly, because the ontological and idealising assumptions involved in the model meant that metabolic network reconstruction could be based on the increasingly available genome sequence data. Secondly, because the community of modellers exhibited epistemic pragmatism. They were happy to accommodate the perspectives of different communities of researchers and revise their commitments to particular ontological and idealizing assumptions. These changes in assumptions eventually led to genome sequence data being regarded as a poor source of evidence for network reconstruction, but by this time the momentum of researchers behind the model was so strong that it retained its position as a popular research tool.

8 Conclusion: Philosophical and historical contributions, limits, and potential

Researchers' assumptions influenced the development of the mathematical models of metabolism in many ways. Researchers' ontological and idealising assumptions frequently affected their evaluation of model building ingredients. Chapter five explored how research communities with commitments to different assumptions valued different mathematical formulations for describing metabolic systems. Much of the thesis, but in particular chapter six, brought to light how researchers' ontological and idealising assumptions affected their evaluation of available data resources. Assumptions which affected the connections between data and models were particularly significant given the important contribution that researchers often assumed that data made to the quality of the mathematical representation of metabolic systems. Researchers' assumptions also played a role in how they addressed the practical constraints they encountered whilst trying to obtain and integrate the ingredients required for model construction. As chapter six argued, the constraints researchers encountered were partially affected by the ontological assumptions they held. Following on from this, as researchers ontological assumptions changed so too did the constraints they encountered, regardless of any changes in the resources and techniques available for model building. The three major groups of constraints which affected model development related to data availability, mathematical tractability and social interactions. Chapters three, four and six showed how researchers overcame the first of these by utilising idealising assumptions to facilitate simplifications in the data required by the model or to connect the model to available data resources. Chapter four additionally emphasised the important role that researchers' ontological assumptions about the temporal organisation of metabolic systems played in facilitating idealising assumptions. Assumptions about the temporal organisation of metabolic processes facilitated simplifications of the mathematical representation of the system. These simplifications addressed constraints that resulted from mathematical intractability. In early biochemical systems biology research groups frequently encountered problems with the social interactions between the diverse communities required to bring together the material resources and expertise needed to move research forward.

Chapter five explored how the rigidity of researchers' commitments and their negative attitude towards other research cultures stifled research. Chapters five and six analysed how the flexible research commitments and open attitude of researchers in current biochemical systems biology facilitated innovation and problem solving in a fast paced research environment.

Looking at the complexity of the relationships between researchers' assumptions and mathematical models of metabolism has facilitated contributions to a diverse range of philosophical and historical issues. Firstly, the close association between bottom up modelling approaches and reductionist perspectives on biochemical systems has been re-evaluated using material from chapters three and six. Researchers often assumed that biochemical data played a major role in allowing a model to be treated as a representation of a specific metabolic system. They also frequently evaluated the data they used in model building according to ontological assumptions about the context dependency of the properties of component parts. When bottom up models were built using biochemical data sets relevant to particular conditions, there was a discrepancy between the reductionist bottom up idealising assumptions and the systemic ontological assumptions researchers used to evaluate biochemical data sets. Secondly, chapter four illustrated that whilst philosophical attention on the role of ontological assumptions in biology has focused on the relationship of spatial parts and wholes, assumptions about the temporal organisation of metabolic systems are of equal significance in understanding research strategies involving idealisations. Thirdly, the chapters in this thesis have contributed an analysis of small scale research going on in the background of mainstream molecular biology and experimental biology. Chapters three and four illustrated that research in early biochemical systems biology involved a high degree of innovation and creativity. Chapter five explored the role of hostile social interactions in stifling the growth of these research projects and chapter six analysed the conditions that allowed a modelling approach to be propelled into foreground of current data intensive systems biology. The philosophical analysis in this thesis has benefitted significantly from being informed by historical case studies. Chapters three and six in particular illustrated the role of a historicist approach in allowing the appreciation of the transient relationships which occur between different aspects of mathematical modelling: ontological assumptions; idealising assumptions;

theory; data and; social interactions. Furthermore they support need to reflect these historically contingent relationships in our philosophical abstractions.

The methodological approach of this thesis delimited a particular area for investigation. The thesis has mostly examined mathematical models which were developed to analyse the behaviour of entire metabolic systems rather than particular pathways. The cluster of models were analysed because of the actual relationships between them and the main researchers involved in their development. For instance, researchers involved in metabolic control analysis interacted directly with those involved in biochemical systems theory (see chapters three, four and five) and Bernhard Palsson started out by working on mathematical models based on relaxation times before moving on to develop the constraint based optimisation approach (see chapters four and six). The models were not selected for analysis because they easily fitted into categories such as top down and bottom up, or pragmatic and systems theoretic, around which many historical perspectives on contemporary systems biology are based. As I discussed in chapter two's methodological context, the overarching philosophical theme of the thesis, namely researchers' assumptions and its analysis in the chapters in relation to data, temporal organisation, social interactions and research dynamics, was also something which emerged out of the case studies, not something that the case studies were selected in order to contribute to. Methodologies inform decisions about the parameters of research, providing the boundaries which facilitate productive investigation, but, as such, research carried out is always limited in its perspective. Additionally, philosophical research through providing new perspectives also highlights potential directions for further investigation.

Within the remit of researchers' assumptions and model development there is potential to expand the philosophical work carried out in this thesis. For instance, model development often involves comparing outputs of a model to experimental data sets. Comparing how the mathematical model behaves in relation to how the actual system behaves in an experimental context can facilitate the evaluation of the mathematical model. This process involves important assumptions about the relationship between the mathematical reconstruction of the system and the system itself. This aspect of model development has been addressed in philosophical literature examining mathematical models across different sciences (Morgan, 1988; Boumans, 1999; Morgan, 2005; Winsberg, 2006; Krohs, 2008). However, as I stated

in my methodological context, I focused my attention on what seemed to be playing the most important and influential role in model development. It is interesting that the evaluation of model outputs against experimental data did not appear to be a major factor affecting model development, even though it was often part of this process. The research in this thesis also provides a platform to move beyond the philosophical focus on assumptions and mathematical model development. An important issue in philosophy of biology has been the intersection of methods, assumptions and type of explanations. Specifically the relationships between mathematical models, assumptions about the organisation of parts and wholes and mechanistic explanations (Bechtel and Abrahamsen, 2010; Bechtel, 2011; Kaplan and Craver, 2011; Brigandt, 2013b; Boogerd et al., 2013). In light of this, the metaphysical and explanatory implications of temporal decomposition would be an interesting area for further philosophical research. Another area for further exploration is the intersection of my analysis of researchers' commitments and how they shape the problem spaces researchers occupy in relation to literature on social epistemology and the role of values in scientific research (Longino, 1990; Kincaid et al., 2007).

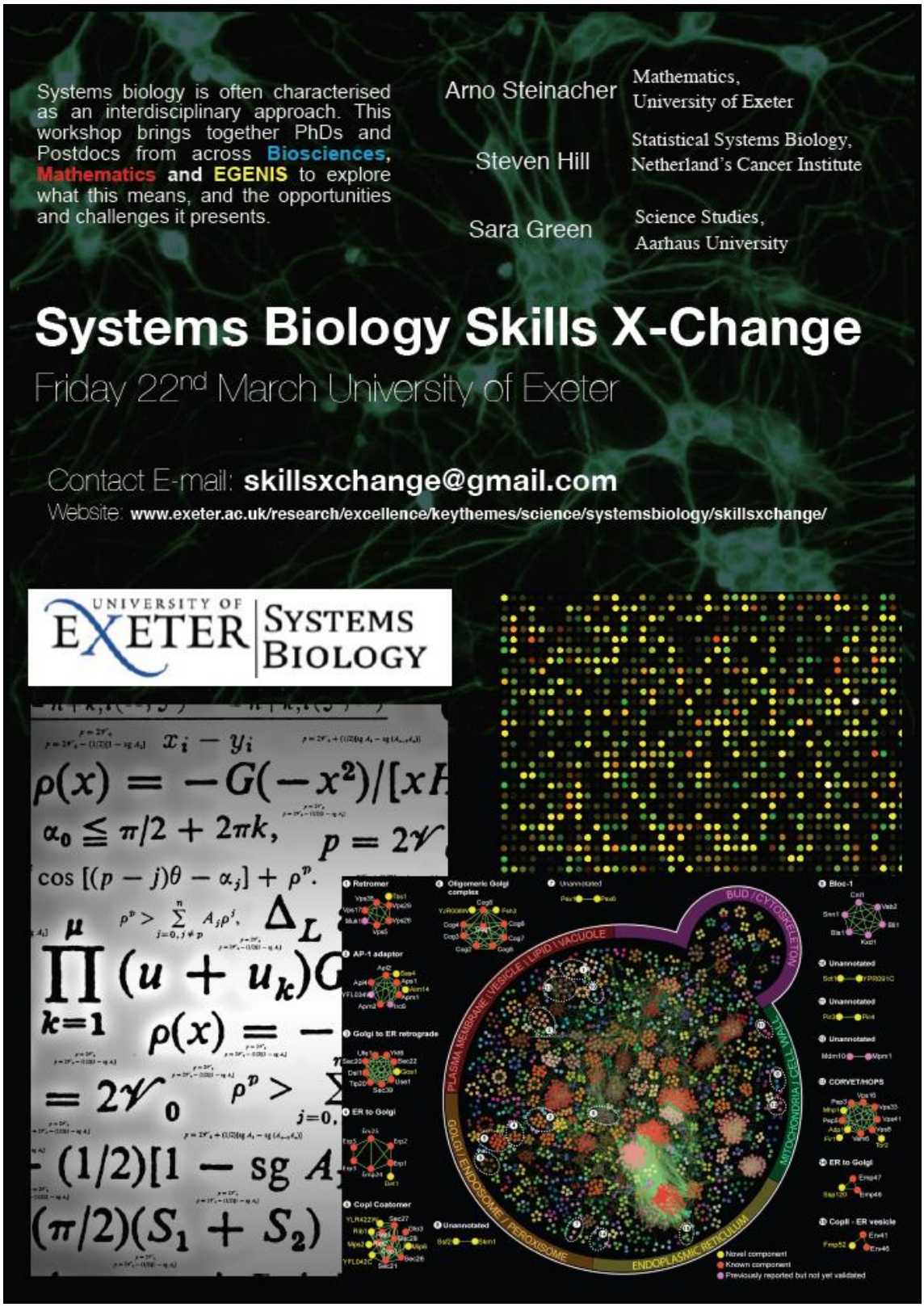
Historically there is much material on mathematical models of metabolism which I did not look at in my research. There are also many approaches to modelling used in current biochemical systems biology which this thesis does not examine, including detailed kinetic modelling (Le Novere et al., 2006) and network analysis (Alon, 2006). Some of this has been addressed in contemporary philosophy of science. See for instance Sara Green's body of work for an in depth analysis of philosophical issues related to network analysis and design principles (Green and Wolkenhauer, 2013; Green, 2013; Green et al., Forthcoming). As I highlighted in section 2.3, there is scope for more work on the history of mathematical models of the dynamic behaviour of individual reactions. It would be interesting to look at the development of Michaelis-Menten's (1913) initial mathematical model of the kinetic behaviour of individual reactions through to the coupling of these equations to construct mathematical models of the dynamic behaviour of systems of multiple reactions. The history of each of the case studies could have been examined over longer periods of time. Extending further into the past, several of the modellers mention being inspired by earlier researchers who developed more abstract theoretical models of biochemical systems. For instance, Joseph Higgins cites Alfred

Lotka, Ilya Prigogine, and Brian Goodwin as significant influences on his work (Higgins, 1965).

Moving further forward in time, whilst the thesis examines the uptake of constraint based optimisation in current data intensive science in chapter six and uses Nersessian and McLeod's (2013c) research to comment on the current use of biochemical systems biology in chapter five, it would be fruitful to examine the history of a wider diversity of metabolic models over the transition into the data intensive era. The breadth of research around each of the cases could also be widened to examine the social and political context in which research was taking place. Such work could help highlight the wider cultural context associated with researchers adopting commitments which led to hostile or open interactions with members of other research cultures. Given the philosophical contributions this thesis has made by analysing a handful of historical cases there is potential for further productive work in this area by expanding upon the range of models, as well as the temporal and cultural dimensions along which they are analysed.

Appendices

Appendix A: Poster for the Systems Biology Skills X-Change



Appendix B: Article on the Dance your PhD project in the Autumn 2012 SPSP newsletter

P-o-S vs. S-i-P Cont.

She discussed our questions with other members of her group, i.e. Susann Wagenknecht, Sara Green, Mads Goddiksen and Brian Hepburn, and together they came up with the following answer:

Philosophy of Contemporary Science in Practice is a project that aims at achieving a detailed understanding of characteristic changes in contemporary science and their implications for the future development of science. Our group conducts case studies that investigate contemporary science on an empirical basis, focusing on a range of topics including the increased mathematical embedding and use of engineering approaches in areas of contemporary biology, explanatory practices in nanoscience education, trust and testimonial practices in collaboration among scientists on a group level, authorship practices, exploratory experimentation, conceptual development in interdisciplinary research, and scientific misconduct and other malpractices in science. Hence, some of our work is engaged with the practices of how science is actually done, and some of our work is engaged with how best to investigate traditional philosophical problems in the sciences – most of the time it is engaged with both.

Our work stands on both sides of Dupré's distinction between philosophy-of-science in practice and philosophy of science-in-practice. Our position is that philosophy of science is improved by a proper understanding of scientific practice, at the same time recognizing that to understand scientific practice properly requires the application of the philosophy of science. The two evolve together. We would widen the application of philosophy-of-science in practice to include not only, as Kevin C.

Graduate Students Speak Out!

Life as a graduate student is full of new experiences. These often form the most memorable parts of the years we spend toiling away on MAs and PhDs – though surely there are some things we'd rather forget. I recall the first time I was able to share my work (on gene concepts) with a practicing geneticist. He made my neat-and-tidy philosophy crumble when he revealed, "I don't care what a gene is. I never use the word and the concept plays no role in my practice. The term 'gene' only crops up when we start explaining our findings to the PR department." After mumbling a few incoherent words about *a priori generalizations* and *the theory-ladenness of observation*, I left the room and quietly admitted defeat. My first interaction with a scientist lead to my first trip back to the drawing board. I am told these things happen often.

Often comical, sometimes disastrous, other-times uplifting, these firsts are important for grad students. My first encounter with a scientist and my first taste of failure prepared me for yet other firsts, like my first *successful* encounter with a scientist, and the first time I integrated practice into my studies of science.

Many graduate students will share the same firsts: first conferences, first publication rejections, first acceptances(!), and first collaborations. This month, I had the opportunity to talk to one SPSP grad student who is not content with this standard list. I caught up with Jo Donaghy, a student researcher with EGENIS at the University of Exeter, to talk about her very unique experience – surely an SPSP first.

If, like Jo, you've had to move to a small quiet town to do your PhD, you'll probably be

P-o-S vs. S-i-P Cont.

Elliot did in the last newsletter, how philosophical insights can be used to assist policy makers and citizens in addressing science related questions, but also how and to what extent philosophical insights into the nature of science might be used by science educators to enhance scientific proficiency. The comparative, empirical investigation of the role of explanations in textbooks from different disciplines is an example of how philosophical insights can be of benefit for educators and students in interdisciplinary programs.

We see science as a multi-faceted enterprise that can vary considerably from one area of inquiry to the next. We therefore also think that philosophy of science-in-practice needs to draw on many methodologies, not only from philosophy but also from history, sociology, psychology and the cognitive sciences. Although the demarcation lines between philosophy of science-in-practice and the many other modes of studying science may not dissolve, they can be made more permeable and less rigid in practice. In the same way, if philosophy-of-science in practice engages with scientific research through interaction with scientists about philosophical problems in the sciences, and also engages with the application of scientific research in society or the training of future scientists, it will necessarily have closer relations to neighboring disciplines such as science education.

Grad Students Speak Out! Cont.

looking for ways to make life a little more interesting. Jo found one: Submitting a proposal to the annual 'Dance your PhD' contest, a web-run contest asking PhD students to create interpretive dances of their research (and you thought writing an abstract was hard!).

Many fellow staff and students encouraged me from the start, bombarding me with links to their favourite entries from previous years. A few just gave me slightly concerned looks. Unfortunately none of them could be tempted to get in front of the camera. Sadly I didn't win. The department continued to show their support at a post research seminar showing of the dance film, and I gave a talk about the project during the ESRC festival of social science.

I learnt a lot about public engagement in research. I was working with a team of 20 local performers and filmmakers. I was taken by surprise by how incredibly interested they all were in my research. In preparing for the project I had been so busy with logistics and the final piece that I failed to realise that the process of making the piece was an opportunity for public engagement. The performers and filmmakers were donating their time and expertise; I could have offered them a much better insight into current philosophy of science in return.

I also learnt a lot about the importance of paying close attention to the history of science. I made the piece towards the beginning of my PhD when I was still grappling with what it meant to do academic research and what exactly it was I was researching. Whilst this didn't make for a particularly clear and engaging final piece it meant that I was more open to learning about my topic during the process. I was trying to make a piece about methodological

Grad Students Speak Out! Cont.

integration. A problem I encountered whilst creating the piece made me appreciate that I needed to pay more attention to the historical trajectories of the different methods when I returned to my research.

Jo's video is, to my knowledge, an SPSP first. Perhaps Jo's story will inspire another entrant, perhaps even the first SPSP *Dance your PhD* winner? … but I'm afraid it won't be me.

Jo's video can be found here: http://vimeo.com/30026966 and be sure to check out the winning 2011 videos here: http://tinyurl.com/3b3xh2b. If you're interested in the 2012 contest, go here: http://gonzolabs.org/dance/, Jo highly recommends it!

Jo Donaghy is a 2nd year PhD student at the ESRC Centre for Genomics and Society, University of Exeter. She is working on the AHRC-funded project Philosophical and Historical Perspectives on the Systems Biology of Metabolism. This project is supervised by Sabina Leonelli and John Dupré. Jo moved into HPS after taking a BSc in Human Sciences at the University of Sussex.

Jordan Bartol is a 2nd year PhD student in the Centre for History and Philosophy of Science at the University of Leeds. He does not dance.

Talk of the Town

Workshop Reports: Practice and Philosophy

We attended two workshops that examined the relations between philosophy and practice, with interesting outcomes.

"*Empirical philosophy of science – qualitative methods*" March 21-23, 2012, Sandbjerg Estate (Denmark), organized by the project *Philosophy of Contemporary Science in Practice,* led by Professor Hanne Andersen at the Centre for Science Studies, Aarhus University, Denmark.

This workshop examined how empirical methods are being used and could be used to answer and reformulate philosophy of science questions. The meeting was attended by almost 20 people, mostly philosophers but of various interdisciplinary backgrounds. It provided a forum for discussing a wide variety of empirically inspired philosophy and science studies work. 'Empirically immersed' philosophers choose to do empirical work themselves, whereas 'empirically informed' philosophers saw themselves in the role of 'second-order observers' and draw on empirical data that others have established – both roles embodied by Erik Angner of George Mason University, and his work on happiness. It also became clear that empirical data of interest to the philosopher arise from a range of different empirical methods, including historical analyses, ethnography, in-depth interviews and quantitative statistical analyses.

Among the speakers, Susann Wagenknecht of Aarhus University spoke about the possibility of a naturalist framework justifying an empirical philosophy of science approach. An

References

Abir-Am, P. G. (1992) the Politics of Macromolecules: Molecular Biologists, Biochemists, and Rhetoric. *Osiris,* 7**,** 164-191.

Allchin, D. (1996) Cellular and Theoretical Chimeras: Piecing Together how Cells Process Energy. *Studies in History and Philosophy of Science Part A,* 27**,** 31-41.

Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits,* Florida, Chapman and Hall / CRC.

Arabatzis, T. & Schickore, J. (2012) Ways of Integrating History and Philosophy of Science. *Perspectives on Science,* 20**,** 395-408.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. & Eppig, J. T. (2000) Gene Ontology: A Tool for the Unification of Biology. *Nature Genetics,* 25**,** 25-29.

Atkinson, D. E. (1990a) An Experimentalist's View of Control Analysis. IN Cornish-Bowden, A. & Cárdenas, M. L. (Eds.) *Control of Metabolic Processes.* New York, Plenum Press.

Atkinson, D. E. (1990b) Prologue: What Should a Theory offer? IN Cornish-Bowden, A. & Cárdenas, M. L. (Eds.) *Control of Metabolic Processes.* New York and London, Plenum Press.

Bailer-Jones, D. (2009) *Scientific Models in Philosophy of Science,* Pittsburgh, University of Pittsburgh Press.

Bapteste, E. & Dupré, J. (2013) Towards a Processual Microbial Ontology. *Biology & Philosophy,* 28**,** 379-404.

Barnes, B. & Dupré, J. (2008) *Genomes and What to Make of Them,* Chicago, University of Chicago Press.

Bechtel, W. (1984) Reconceptualizations and Interfield Connections: The Discovery of the Link Between Vitamins and Coenzymes. *Philosophy of Science,* 51**,** 265-292.

Bechtel, W. (1986) Biochemistry: A Cross-Disciplinary Endeavor that Discovered a Distinctive Domain. IN Bechtel, W. (Ed.) *Integrating Scientific Disciplines.* Netherlands, Martinus Nijhoff Publishers.

Bechtel, W. (1990) Scientific Evidence: Creating and Evaluating Experimental Instruments and Research Techniques. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association.* JSTOR.

Bechtel, W. (2000) From Imaging to Believing: Epistemic Issues in Generating Biological Data. IN Creath, R. & Maienschein, J. (Eds.) *Biology and Epistemology.* Cambridge, Cambridge University Press.

Bechtel, W. (2006) *Discovering Cell Mechanisms: The Creation of Modern Cell Biology,* Cambridge, Cambridge University Press.

Bechtel, W. (2011) Mechanism and Biological Explanation*. *Philosophy Of Science,* 78**,** 533-557.

Bechtel, W. & Abrahamsen, A. (2005) Explanation: A Mechanist Alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 36**,** 421-441.

Bechtel, W. & Abrahamsen, A. (2010) Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science. *Studies in History and Philosophy of Science Part A,* 41**,** 321-333.

Bechtel, W. & Richardson, R. C. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research,* London, The MIT Press.

Becker, S. A. & Palsson, B. Ø. (2008) Context-Specific Metabolic Networks are Consistent with Experiments. *Plos Computational Biology,* 4**,** e1000082.
Blackman, F. (1905) Optima and Limiting Factors. *Annals of Botany,* 19**,** 281-295.
Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & Mayhew, G. F. (1997) the Complete Genome Sequence of Escherichia coli K-12. *Science,* 277**,** 1453-1462.
Bogen, J. & Woodward, J. (1988) Saving the Phenomena. *the Philosophical Review,* 97**,** 303-352.
Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J. H. S. & Westerhoff, H. V. (2007) Towards Philosophical Foundations of Systems Biology: Introduction. IN Boogerd, F. C., Bruggeman, F. J., Jan-Hendrik, S., Hofmeyr, J. H. S. & Westerhoff, H. V. (Eds.) *Systems Biology: Philosophical Foundations.* Elsevier.
Boogerd, F. C., Bruggeman, F. J. & Richardson, R. C. (2013) Mechanistic Explanations and Models in Molecular Systems Biology. *Foundations of Science,* 18**,** 725-744.
Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø. & Jamshidi, N. (2010) Insight Into Human Alveolar Macrophage and M. tuberculosis Interactions via Metabolic Reconstructions. *Molecular Systems Biology,* 6**,** 1-14.
Borodina, I., Krabben, P. & Nielsen, J. (2005) Genome-scale Analysis of Streptomyces coelicolor A3 (2) Metabolism. *Genome Research,* 15**,** 820-829.
Bosca, L. & Corredor, C. (1984) Is Phosphofructokinase the Rate-Limiting Step of Glycolysis? *Trends in Biochemical Sciences,* 9**,** 372-373.
Bostanci, A. (2004) Sequencing human genomes. IN Gaudillière, J. & Rheinberger, H. J. (Eds.) *From Molecular Genetics to Genomics: the Mapping Cultures of Twentieth-Century Genetics.* London and New York, Routledge.
Boumans, M. (1999) Built-in justification. IN Morgan, M. S. & Morrison, M. (Eds.) *Models as Mediators.* Cambridge, Cambridge University press.
Boumans, M. (2005) *How Economists Model the World into Numbers,* Oxon, Routledge
Brigandt, I. (2013a) Integration in Biology: Philosophical Perspectives on the Dynamics of Interdisciplinarity. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 44**,** 461-465.
Brigandt, I. (2013b) Systems Biology and the Integration of Mechanistic Explanation and Mathematical Explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 4**,** 477-492.
Brigandt, I. & Love, A. (2012) Reductionism in Biology. IN Zalta, E., N. (Ed.) *Stanford Encyclopedia of Philosophy*
Briggs, G. E. & Haldane, J. B. S. (1925) A Note on the Kinetics of Enzyme Action. *Biochemical Journal,* 19**,** 338.
Brown, W. E. L. & Hill, A. V. (1923) The Oxygen-Dissociation Curve of Blood, and its Thermodynamical Basis. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character,* 94**,** 297-334.
Bruggeman, F. J. & Westerhoff, H. V. (2007) The Nature of Systems Biology. *Trends in Microbiology,* 15**,** 45-50.
Burian, R. (1977) More than a Marriage of Convenience: On the Inextricability of History and Philosophy of Science. *Philosophy of Science,* 44**,** 1-42.

Burian, R. (2001) the Dilemma of Case Studies Resolved: The Virtues of Using Case Studies in the History and Philosophy of Science. *Perspectives on Science,* 9**,** 383-404.
Burian, R. (2004) Molecular Epigenesis, Molecular Pleiotropy, and Molecular Gene Definitions. *History and Philosophy of the Life Sciences,* 26**,** 59-80.
Burian, R. (2013) Hans-Jorg Rheinberger on Biological Time Scales. *History and Philosophy of the Life Sciences,* 35**,** 19-25.
Burns, J. A., Cornish-Bowden, A., Groen, A. K., Heinrich, R., Kacser, H., Porteous, J. W., Rapoport, S. M., Rapoport, T. A., Stucki, J. W. & Tager, J. M. (1985) Control Analysis of Metabolic Systems. *Trends in Biochemical Sciences,* 10**,** 16.
Burns, J. A., Fell, D., Heinrich, R., Mazat, J. P., Small, R. & Westerhoff, H. (1995) Henrik Kacser, 1918-1995. *Trends in Biochemical Sciences,* 20**,** 297-298.
Butcher, E. C., Berg, E. L. & Kunkel, E. J. (2004) Systems Biology in Drug Discovery. *Nature Biotechnology,* 22**,** 1253-1259.
Calvert, J. (2010) Systems Biology, Interdisciplinarity and Disciplinary Identity. IN Parker, J. N., Vermeulen, N. & Penders, B. (Eds.) *Collaboration in the New Life Sciences.* Farnham, Ashgate.
Calvert, J. & Fujimura, J. H. (2011) Calculating Life? Duelling Discourses in Interdisciplinary Systems Biology. *Studies in History and Philosophy of Biological and Biomedical sciences,* 42**,** 155-163.
Cartwright, N. (1983) *How the Laws of Physics Lie,* Oxford, Clarendon Press.
Cartwright, N. (1994) *Nature's Capacities and their Measurement,* Oxford, Oxford University Press.
Cartwright, N. (1999a) *the Dappled World: A Study of the Boundaries of Science,* Cambridge, Cambridge University Press.
Cartwright, N. (1999b) Models and the Limits of Theory: Quantum Hamiltonians and the BCS Models of Superconductivity. IN Morgan, M. S. & Morrison, M. (Eds.) *Models as Mediators.* Cambridge, Cambridge University Press.
Cascante, M., Boros, L. G., Comin-Anduix, B., De Atauri, P., Centelles, J. J. & Lee, P. W.-N. (2002) Metabolic Control Analysis in Drug Discovery and Disease. *Nature Biotechnology,* 20**,** 243-249.
Chang, H. (2011) Beyond Case-Studies: History as Philosophy. IN Mauskopf, S. & Schmaltz, T. (Eds.) *Integrating History and Philosophy of Science: Problems and Prospects.* London, Springer.
Cornish-Bowden, A. (1989) Metabolic Control Theory and Biochemical Systems Theory: Different Objectives, Different Assumptions, Different Results. *Journal of Theoretical Biology,* 136**,** 365-377.
Cornish-Bowden, A. (1995) *Fundamentals of Enzyme Kinetics*, Portland Press London.
Cornish-Bowden, A. & Cárdenas, M. (2005) Systems Biology May Work When we Learn to Understand the Parts in Terms of the Whole. *Biochemical Society Transactions,* 33**,** 516-519.
Cornish-Bowden, A. & Cárdenas, M. L. (Eds.) (1990) *Control of Metabolic Processes,* New York and London, Plenum Press.
Covert, M. W., Schilling, C. H. & Palsson, B. (2001) Regulation of Gene Expression in Flux Balance Models of Metabolism. *Journal of Theoretical Biology,* 213**,** 73-88.
Craver, C. F. (2006) When Mechanistic Models Explain. *Synthese,* 153**,** 355-376.

Creager, A. N. H. (2013a) *Life Atomic: Radioisotopes in Science and Medicine,* Chicago, University of Chicago Press.
Creager, A. N. H. (2013b) Timescapes of Radioactive Tracers in Biochemistry and Ecology. *History and Philosophy of the Life Sciences,* 35**,** 83-90.
Creager, A. N. H. & Gaudillière, J. (1996) Meanings in Search of Experiments and Vice-Versa: the Invention of Allosteric Regulation in Paris and Berkeley, 1959-1968. *Historical Studies in the Physical and Biological Sciences,* 27**,** 1-89.
Darden, L. & Maull, N. (1977) Interfield Theories. *Philosophy of Science,* 44**,** 43-64.
Davies, G., Frow, E. & Leonelli, S. (2013) Bigger, Faster, Better: Rhetorics and Practices of Large-Scale Research in Contemporary Bioscience. *Biosocieties,* 8**,** 386-396.
De Chadarevian, S. (1996) Sequences, Conformation, Information: Biochemists and Molecular Biologists in the 1950s. *Journal of the History of Biology,* 29**,** 361-386.
De Chadarevian, S. (2002) *Designs for Life: Molecular Biology after World War II,* Cambridge, Cambridge University Press.
Dejongh, M., Formsma, K., Boillot, P., Gould, J., Rycenga, M. & Best, A. (2007) Toward the Automated Generation of Genome-Scale Metabolic Networks in the SEED. *BMC Bioinformatics,* 8**,** 139.
Donaghy, J. (2013) Autonomous Mathematical Models: Constructing Theories of Metabolic Control. *History and Philosophy of the Life Sciences,* 35**,** 533-552.
Donaghy, J. (Forthcoming) Temporal Decomposition: A Strategy for Building Mathematical Models of Complex Metabolic Systems. *Studies in History and Philosophy of Biological and Biomedical Sciences*.
Duarte, N. C., Herrgård, M. J. & Palsson, B. Ø. (2004) Reconstruction and Validation of Saccharomyces Cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Research,* 14**,** 1298-1309.
Dupré, J. (1995) *The Disorder of Things: Metaphysical Foundations of The Disunity of Science,* London, Harvard University Press.
Dupré, J. (2009) It Is Not Possible to Reduce Biological Explanations to Explanations in Chemistry and/or Physics. IN Ayala, F. J. & Arp, R. (Eds.) *Contemporary Debates in Philosophy of Biology.* Malden, Mass, Wiley-Blackwell.
Dupré, J. (2012) *Processes of Life: Essays in the Philosophy of Biology,* Oxford, Oxford University Press.
Dupré, J. (2013) I-Living Causes. *Aristotelian Society Supplementary Volume,* 87**,** 19-37.
Edwards, J. S. & Palsson, B. Ø. (1998) How Will Bioinformatics Influence Metabolic Engineering? *Biotechnology and Bioengineering,* 58**,** 162-169.
Edwards, J. S. & Palsson, B. Ø. (1999) Systems Properties of the Haemophilus influenzaeRd Metabolic Genotype. *Journal of Biological Chemistry,* 274**,** 17410.
Edwards, J. S. & Palsson, B. Ø. (2000a) The Escherichia coli MG1655 in silico Metabolic Genotype: Its Definition, Characteristics, and Capabilities. *Proceedings of the National Academy of Sciences,* 97**,** 5528.
Edwards, J. S. & Palsson, B. Ø. (2000b) Robustness Analysis of the Escherichia coli Metabolic Network. *Biotechnology Progress,* 16**,** 927-939.
Eigen, M. (1968) New Looks and Outlooks on Physical Enzymology. *Quarterly Review of Biophysics* 1**,** 3-33.
Falk, R. (2001) The Rise and Fall of Dominance. *Biology and Philosophy,* 16**,** 285-323.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. & Palsson, B. Ø. (2007) A Genome-Scale Metabolic Reconstruction for Escherichia coli K-12 MG1655 that Accounts for 1260 ORFs and Thermodynamic Information. *Molecular Systems Biology,* 3**,** 1-18.

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B. Ø. (2008) Reconstruction of Biochemical Networks in Microorganisms. *Nature Reviews Microbiology,* 7**,** 129-143.

Fell, D. A. (1984) Phosphofructokinase and Glycolytic Flux. *Trends in Biochemical Sciences,* 9**,** 515-516.

Fell, D. A. (1992) Metabolic Control Analysis: A Survey of its Theoretical and Experimental Development. *Biochemical Journal,* 286**,** 313-330.

Fell, D. A. (1997) *Understanding the Control of Metabolism,* London, Portland press

Fell, D. A. (2007) How can we Understand Metabolism. IN Boogerd, F. C., Bruggeman, F., Hofmeyr, J. H. S. & Westerhoff, H. (Eds.) *Systems Biology: Philosophical Foundations.* Netherlands, Elsevier.

Fell, D. A. & Sauro, H. M. (1985) Metabolic Control and its Analysis: Additional Relationships between Elasticities and Control Coefficients. . *European Journal of Biochemistry,* 148**,** 555-561.

Fell, D. A. & Small, J. R. (1986) Fat Synthesis in Adipose Tissue. An Examination of Stoichiometric Constraints. *Biochemical Journal,* 238**,** 781-786.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A. & Merrick, J. M. (1995) Whole-Genome Random Sequencing and Assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.

Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. (2003) Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome Research,* 13**,** 244-253.

Foucault, M. (1973) *The Order of Things* New York, Vintage.

Galison, P. (1987) *How Experiments End,* Chicago, London, University of Chicago Press.

Galison, P. (1995) Context and Constraints. IN Buchwald, J. Z. (Ed.) *Scientific Practice: Theories and Stories of Doing Physics.* Chicago, University of Chicago Press

Galison, P. (1997) *Image and logic: A Material Culture of Microphysics,* Chicago, University of Chicago Press.

Galison, P. (1999) Trading Zone: Coordinating Action and Belief. IN Biagioli, M. (Ed.) *the Science Studies Reader.* London, Routledge.

Galison, P. (2008) Ten Problems in History and Philosophy of Science. *Isis,* 99**,** 111-124.

Garfinkel, D. (1969) Construction of Biochemical Computer Models. *FEBS Letters,* 2**,** S9-S13.

Garfinkel, D., Garfinkel, L., Pring, M., Green, S. B. & Chance, B. (1970) Computer Applications to Biochemical Kinetics. *Annual Review of Biochemistry,* 39**,** 473-498.

Gianchandani, E. P., Chavali, A. K. & Papin, J. A. (2010) The Application of Flux Balance Analysis in Systems Biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine,* 2**,** 372-382.

Giere, R. N. (1973) History and Philosophy of Science: Intimate Relationship or Marriage of Convenience? . *The British Journal for the Philosophy of Science,* 24**,** 282-297.
Gilbert, S. & Sarkar, S. (2000) Embracing Complexity: Organicism for the 21st century. *Developmental Dynamics,* 219**,** 1-9.
Gillies, R. (1983) Understanding Pathway Control (Without the aid of computers). *Trends in Biochemical Sciences,* 8**,** 2-3.
Glennan, S. (2002) Rethinking Mechanistic Explanation. *Philosophy of Science,* 69**,** S342-S353.
Goto, S., Okuno, Y., Hattori, M., Nishioka, T. & Kanehisa, M. (2002) LIGAND: Database of Chemical Compounds and Reactions in Biological Pathways. *Nucleic Acids Research,* 30**,** 402-404.
Gramelsberger, G. (2011) What do Numerical (Climate) Models Really Represent? *Studies in History and Philosophy of Science Part A,* 42**,** 296-302.
Green, S. (2013) When One Model is Not Enough: Combining Epistemic Tools in Systems Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 44**,** 170-180.
Green, S., Levy, A. & Bechtel, W. (Forthcoming) Design sans Adaptation. *European Journal for Philosophy of Science*.
Green, S. & 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. (2001) Reproduction and Reduction of Genetics IN Beurton, P. J., Falk, R. & Rheinberger, H. J. (Eds.) *the Concept of the Gene in Development and Evolution* Cambridge, Cambridge University Press.
Griesemer, J. (2002) Space<=> Time: Temporality and Attention in Iconographies of the Living. IN Schmidgen, H. (Ed.) *Experimental Arcades: The Materiality of Time Relations in Life Sciences, Art, and Technology (1830–1930).* Bauhaus University Weimar
Griesemer, J. (2007) Tracking Organic Processes: Representations and Research Styles in Classical Embryology and Genetics. IN Laubichler, M. D. & Maienschein, J. (Eds.) *From Embryology to Evo-Devo: A History of Developmental Evolution.* London, MIT Press
Griffiths, P. & Stotz, K. (2013) *Genetics and Philosophy: An Introduction,* Cambridge, Cambridge University Press.
Groen, A. K., Wanders, R., Westerhoff, H., Van Der Meer, R. & Tager, J. (1982) Quantification of the Contribution of Various Steps to the Control of Mitochondrial Respiration. *Journal of Biological Chemistry,* 257**,** 2754-2757.
Groen, A. K. & Westerhoff, H. (1990) Modern Control Theories: A Consumers' Test. IN Cornish-Bowden, A. & Cárdenas, M. L. (Eds.) *Control of Metabolic Processes.* New York and London, Plenum Press.
Hacking, I. (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science,* New York, Cambridge University Press.
Hacking, I. (1992a) The Self-Vindication of the Laboratory Sciences. IN Pickering, A. (Ed.) *Science as Practice and Culture.* Chicago, University of Chicago press.
Hacking, I. (1992b) 'Style'for Historians and Philosophers. *Studies in History and Philosophy of Science Part A,* 23**,** 1-20.
Heinemann, M. & Sauer, U. (2011) From Good Old Biochemical Analyses to High-Throughput Omics Measurements and Back. *Current Opinion in Biotechnology,* 22**,** 1-2.

Heinrich, R., Rapoport, S. & Rapoport, T. (1977) Metabolic Regulation and Mathematical Models. *Prog. Biophys. Mol. Biol,* 32**,** 1-82.
Heinrich, R. & Rapoport, T. A. (1974a) A Linear Steady State Treatment of Enzymatic Chains: Critique of the Crossover Theorem and a General proceedure to Identify Interaction Sites with an Effector. *European Journal of Biochemistry,* 42**,** 97-105.
Heinrich, R. & Rapoport, T. A. (1974b) A Linnear Steady State Treatment of Enzymatic Chains: General Properties Control and Effector Strength. *European Journal of Biochemistry,* 42**,** 89-95.
Heinrich, R. & Schuster, S. (1996) *The Regulation of Cellular Systems,* New York, Chapman and Hall.
Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R. & Heinemann, M. (2008) A Consensus Yeast Metabolic Network Reconstruction Obtained from a Community Approach to Systems Biology. *Nature Biotechnology,* 26**,** 1155-1160.
Higgins, J. (1961) Use of Computers. IN Friess, S. L., Lewis, E. S. & Weissberg, A. (Eds.) *Investigations of Rates and Mechanisms of Reactions.* London, Interscience Publishers.
Higgins, J. (1963) Analysis of Sequential Reactions. *Annals of the New York Academy of Sciences,* 108**,** 305-321.
Higgins, J. (1964) A Chemical Mechanism for Oscillation of Glycolytic Intermediates in Yeast Cells. *Proceedings of the National Academy of Sciences of the United States of America,* 51**,** 989-994.
Higgins, J. (1965) Dynamics and Control in Cellular Reactions. IN Chance, B., Estabrook, R. & Williamson, J. (Eds.) *Control of Energy Metabolism.* New York, Academic Press.
Higgins, J. (1990) History and Original Thoughts on the Control Theoretic Approach. IN Cornish-Bowden, A. & Maria, L. C. (Eds.) *Control of Metabolic Processes.* New york and London, Plenum Press.
Hilgartner, S. (1995) Biomolecular Databases New Communication Regimes for Biology? *Science Communication,* 17**,** 240-263.
Holmes, F. L. (1986) Intermediary Metabolism in the Early Twentieth Century. IN Bechtel, W. (Ed.) *Integrating Scientific Disciplines.* Dordrecht, Martinus Nijhoff Publishers.
Holmes, F. L. (1992) *Between Biology and Medicine: The Formation of Intermediary Metabolism*, University of California, Office for History of Science & Technology.
Holmes, F. L. (2004) *Investigative Pathways: Patterns and Stages in the Careers of Experimental Scientists,* London, Yale University Press.
Hopcroft, J. E. & Ullman, J. D. (1979) *Introduction to Automata Theory, Languages, and Computation,* Reading, Massachusetts, Addison-Wesley
Hüttemann, A. (2005) Explanation, Emergence, and Quantum Entanglement*. *Philosophy of Science,* 72**,** 114-127.
Hüttemann, A. & Love, A. C. (2011) Aspects of Reductive Explanation in Biological Science: Intrinsicality, Fundamentality, and Temporality. *The British Journal for the Philosophy of Science,* 62**,** 1-31.
Irvine, D. H. (1990) Objectives, Assumptions and Results of Metabolic Control Theory and Biochemical Systems Theory. *Journal of Theoretical Biology,* 143**,** 139.

Jamshidi, N. & Palsson, B. Ø. (2010) Mass Action Stoichiometric Simulation Models: Incorporating Kinetics and Regulation into Stoichiometric Models. *Biophysical Journal,* 98**,** 175-185.
Joshi, A. & Palsson, B. Ø. (1989) Metabolic Dynamics in the Human Red Blood Cell:: Part I--A Comprehensive Kinetic Model. *Journal of Theoretical Biology,* 141**,** 515-528.
Kacser, H. (1957) Some Physico-Chemical Aspects of Biological Organisation (Appendix). IN Waddington, C. (Ed.) *the Strategy of the Genes.* London, George Allen & Unwin.
Kacser, H. (1960) Kinetic Models of Development and Heredity. *Symposia of the Society for Experimental Biology,* 14**,** 13-27.
Kacser, H. (1983) Biochemists Alcohol Problem: A Case of Addiction to the Wrong Concepts? *Trends in Biochemical Sciences,* 8**,** 310-311.
Kacser, H. (1991) A Superor Theory? *Journal of Theoretical Biology,* 149**,** 141-144.
Kacser, H. & Burns, J. A. (1967) Causality, Complexity and Computers. IN Locker, A. (Ed.) *Quantitative Biology of Metabolism.* Berlin, Speinger Verlag.
Kacser, H. & Burns, J. A. (1973) The Control of Flux. *Symposia of the Society for Experimental Biology,* 27**,** 65-104.
Kacser, H. & Burns, J. A. (1981) The Molecular Basis of Dominance. *Genetics,* 97**,** 639-666.
Kacser, H. & Porteous, D. (1987a) Replies from Kacser and Porteous: Reply to Savageau. *Trends in Biochemical Sciences,* 12**,** 222-223.
Kacser, H. & Porteous, J. (1987b) Control of Metabolism: What do we have to measure? *Trends in Biochemical Sciences,* 12**,** 5-14.
Kaplan, D. M. & Craver, C. F. (2011) The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective*. *Philosophy Of Science,* 78**,** 601-627.
Karp, P. D., Paley, S. & Romero, P. (2002) The Pathway Tools Software. *Bioinformatics,* 18**,** S225-S232.
Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M. & Pellegrini-Toole, A. (2000) The Ecocyc and Metacyc Databases. *Nucleic Acids Research,* 28**,** 56.
Kauffman, S. A. (1970) Articulation of Parts Explanation in Biology and the Rational Search for Them. IN Grene, M. & Everette, M. (Eds.) *Topics in the Philosophy of Biology.* Dordrecht, D. Reidel Publishing Company.
Kay, L. E. (2000) *Who Wrote the Book of Life?: A History of the Genetic Code*, Stanford University Press.
Keller, E. F. (2010) It is Possible to Reduce Biological Explanations to Explanations in Chemistry and/or Physics. IN Ayala, F. J. & Arp, R. (Eds.) *Contemporary Debates in Philosophy of Biology.* Malden, Mass, Wiley-Blackwell.
Kincaid, H., Dupré, J. & Wylie, A. (2007) Value-Free Science? Ideals and Illusions. New York, Oxford University Press.
Kitano, H. (2001) Systems Biology: Toward System-Level Understanding of Biological Systems. IN Kitano, H. (Ed.) *Foundations of Systems Biology.* Cambridge, Mass and London MIT press.
Kitano, H., Ghosh, S. & Matsuoka, Y. (2011) Social Engineering for Virtual 'Big Science' In Systems Biology. *Nature Chemical Biology,* 7**,** 323-326.
Klitgord, N. & Segrè, D. (2010) Environments that Induce Synthetic Microbial Ecosystems. *Plos Computational Biology,* 6**,** e1001002.
Knorr Cetina, K. (1999) *Epistemic Cultures: How the Sciences Make Knowledge,* Cambridge, Mass and London, Harvard University Press.

Knuuttila, T. (2011) Modelling and Representing: An Artefactual Approach to Model-Based Representation. *Studies in History and Philosophy of Science Part A,* 42**,** 262-271.
Knuuttila, T. & Voutilainen, A. (2003) A Parser as an Epistemic Artifact: A Material View on Models. *Philosophy of Science,* 70**,** 1484-1495.
Kohler, R. E. (1973) The Enzyme Theory and the Origin of Biochemistry. *Isis,* 64**,** 181-196.
Kohler, R. E. (1975) The History of Biochemistry: A Survey. *Journal of the History of Biology,* 8**,** 275-318.
Kolker, E., Purvine, S., Picone, A., Cherny, T., Akerley, B. J., Munson Jr, R. S., Palsson, B. Ø., Daines, D. A. & Smith, A. L. (2002) H. influenzae Consortium: Integrative Study of H. influenzae-Human Interactions. *Omics: A Journal of Integrative Biology,* 6**,** 341-348.
Krebs, H. A., Kornberg, H. L. & Burton, K. (1957) *The Energy Transformations in Living Matter,* Berlin Heidelberg, Springer.
Krohs, U. (2008) How Digital Computer Simulations Explain Real-World Processes. *International Studies in the Philosophy of Science,* 22**,** 277-292.
Krohs, U. (2010) Epistemic Consequences of Two Different Strategies for Decomposing Biological Networks. IN Suárez, M., Dorato, M. & Rédei, M. (Eds.) *EPSA Philosophical Issues in the Sciences.* Springer.
Krohs, U. (2011) Convenience Experimentation. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 43**,** 52-57.
Krohs, U. & Callebaut, W. (2007) Data Without Models Merging with Models Without Data. IN Boogerd, F. C., Bruggeman, F., Hofmeyr, J. H. S. & Westerhoff, H. (Eds.) *Systems Biology: Philosophical Foundations.* Elsevier.
Kuepfer, L., Sauer, U. & Blank, L. M. (2005) Metabolic Functions of Duplicate Genes in Saccharomyces cerevisiae. *Genome Research,* 15**,** 1421-1430.
Lakatos, I. (1978) *Philosophical Papers: The Methodology of Scientific Research Programmes,* Cambridge, MA., Cambridge University Press. .
Landecker, H. (2002) New Times for Biology: Nerve Cultures and the Advent of Cellular Life in Vitro. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 33**,** 667-694.
Landecker, H. (2009) Living Differently in Time: Plasticity, Temporality and Cellular Biotechnologies. *Technologized Images: Technologized Bodies.* Oxford, Berghahn Books.
Landecker, H. (2013) Postindustrial Metabolism: Fat Knowledge. *Public Culture,* 25**,** 495-522.
Latour, B. & Woolgar, S. (1986) *Laboratory Life: The Construction of Scientific Facts,* Princeton, Chicester, Princeton University Press.
Laubichler, M. D. & Wagner, G. P. (2001) How Molecular is Molecular Developmental Biology? A Reply to Alex Rosenberg's Reductionism Redux: Computing the Embryo. *Biology and Philosophy,* 16**,** 53-68.
Laudan, L. (1989) Thoughts on HPS: 20 Years Later. *Studies in History and Philosophy of Science Part A,* 20**,** 9-13.
Le Novere, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M. & Shapiro, B. (2006) BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research,* 34**,** D689-D691.
Leonelli, S. (2009a) On the Locality of Data and Claims about Phenomena. *Philosophy of Science,* 76**,** 737-749.

Leonelli, S. (2009b) Understanding in Biology: The impure nature of biological knowledge. IN De Regt, H., Leonelli, S. & Eigner, K. (Eds.) *Scientific Understanding: Philosophical Perspectives.* Pittsburgh, Pittsburgh University Press.

Leonelli, S. (2010a) Documenting the Emergence of Bio-Ontologies: or, Why Researching Bioinformatics Requires HPSSB. *History and Philosophy of the Life Sciences,* 32**,** 105-125.

Leonelli, S. (2010b) Packaging Small Facts for Re-use: Databases in Model Organism Biology. IN Howlett, P. & Morgan, M. (Eds.) *How Well Do Facts Travel? The Dissemination of Reliable Knowledge.* Cambridge, Cambridge University Press.

Leonelli, S. (2012) Introduction: Making Sense of Data-Driven Research in the Biological and Biomedical Sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences***,** 1-3.

Leonelli, S. (2013) Integrating Data to Aquire New Knowledge: Three Modes of Integration in Plant Science. *Studies in the History and the Philosophy of the Biological and Biomedical Sciences,* 4**,** 503-514.

Leonelli, S. (Forthcoming) Data Interpretation in the Digital Age. *Perspectives on Science*.

Leonelli, S. & Ankeny, R. A. (2011) Re-thinking Organisms: The Impact of Databases on Model Organism Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 43**,** 29-36.

Levins, R. (1966) The Strategy of Model Building in Population Biology. *American Scientist,* 54**,** 421-431.

Liao, J. C. & Delgado, J. (1993) Advances in Metabolic Control Pnalysis. *Biotechnology Progress,* 9**,** 221-233.

Liao, J. C. & Lightfoot, E. N. (1987) Extending the Quasi-Steady State Concept to Analysis of Metabolic Networks. *Journal of Theoretical Biology,* 126**,** 253-273.

Longino, H. E. (1979) Evidence and Hypothesis: An Analysis of Evidential Relations. *Philosophy of Science,* 46**,** 35-56.

Longino, H. E. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific inquiry,* Chichester, Princeton University Press.

Longino, H. E. (2002) *The Fate of Knowledge,* Oxford, Princeton University Press.

Love, A. C. & Hüttemann, A. (2011) Comparing Part-Whole Reductive Explanations in Biology and Physics. IN Dieks, D., Gonzalez, W. J., Hartmann, S., Uebel, T. & Weber, M. (Eds.) *Explanation, Prediction, and Confirmation.* Berlin, Springer.

Machamer, P. (2011) Phenomena, Data and Theories: A Special Issue of Synthese. *Synthese,* 182**,** 1-5.

Machamer, P., Darden, L. & Craver, C. F. (2000) Thinking about Mechanisms. *Philosophy of Science,* 67**,** 1-25.

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

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

Macleod, M. & Nersessian, N. J. (2013c) The Creative Industry of Integrative Systems Biology. *Mind & Society,* 12**,** 35-48.

Majewski, R. A. & Domach, M. M. (1990) Simple Constrained-Optimization View of Acetate Overflow in E. coli. *Biotechnology and Bioengineering,* 35**,** 732-738.
Mccloskey, D., Gangoiti, J. A., King, Z. A., Naviaux, R. K., Barshop, B. A., Palsson, B. Ø. & Feist, A. M. (2013) A Model-Driven Quantitative Metabolomics Analysis of Aerobic and Anaerobic Metabolism in E. coli K-12 MG1655 that is Biochemically and Thermodynamically Consistent. *Biotechnology and Bioengineering,* 111**,** 803-815.
Mcmullin, E. (1976) History and Philosophy of Science: A Marriage of Convenience? IN Cohen, R. S., Hooker, C. A., Michalos, A. C. & Van Evra, J. W. (Eds.) *Psa 1974: Proceedings of the 1974 Biennial Meeting Philosophy of Science Association.* Dordrecht, Reidel Publishing Company.
Méthot, P. (2013) On the Genealogy of Concepts and Experimental Practices: Rethinking Georges Canguilhem's Historical Epistemology. *Studies in History and Philosophy of Science Part A,* 44**,** 112-123.
Michaelis, L. & Menten, M. L. (1913) Die Kinetik Der Invertinwirkung. *Biochemische Zeitschrift,* 49**,** 333-369.
Mitchell, S. D. (2003) *Biological Complexity and Integrative Pluralism,* Cambridge, Cambridge University Press.
Mo, M. L. & Palsson, B. Ø. (2009) Understanding Human Metabolic Physiology: A Genome-to-Systems Approach. *Trends in Biotechnology,* 27**,** 37-44.
Monod, J., Changeux, J. P. & Jacob, F. (1963) Allosteric Proteins and Cellular Control Systems. *Journal of Molecular Biology,* 6**,** 306-329.
Morange, M. (1998) *History of Molecular Biology,* London, Harvard University Press.
Morange, M. (2012) What History Tells Us XXVII. A New Life for Allostery. *Journal of Biosciences,* 37**,** 13-17.
Moreno-Sánchez, R., Saavedra, E., Rodríguez-Enríquez, S. & Olín-Sandoval, V. (2008) Metabolic Control Analysis: A Tool for Eesigning Strategies to Manipulate Metabolic Pathways. *Journal of Biomedicine and Biotechnology,* 2008**,** 1-30.
Morgan, M. (2012) Case Studies: One Observation or Many? Justification or Discovery? *Philosophy of Science,* 79**,** 667-677.
Morgan, M. S. (1988) Finding a Satisfactory Empirical Model. IN De Marchi, N. (Ed.) *The Popperian Legacy in Economics: Papers Presented At A Symposium in Amsterdam, December 1985.* Cambridge, Cambridge University Press.
Morgan, M. S. (2003) Experiments Without Material Intervention. IN Radder, H. (Ed.) *The Philosophy of Scientific Experimentation.* Pittsburgh, University of Pittsburgh Press.
Morgan, M. S. (2005) Experiments versus Models: New Phenomena, Inference and Surprise. *Journal of Economic Methodology,* 12**,** 317-329.
Morgan, M. S. & Knuuttila, T. (2012) Models and Modelling in Economics. IN Mäki, U. (Ed.) *Handbook of the Philosophy of Economics.* Oxford, Elsevier.
Morgan, N. (1990) From Physiology to Biochemistry. IN Olby, R. C., Cantor, G. N., Christie, J. R. R. & Hodge, M. J. S. (Eds.) *Companion to the History of Modern Science.* London, Routledge.
Morrison, M. (1999) Models as Autonomous Agents. IN Morrison, M. & Morgan, M. S. (Eds.) *Models as Mediators.* Cambridge University Press.
Morrison, M. & Morgan, M. S. (1999a) Introduction. IN Morgan, M. S. & Morrison, M. (Eds.) *Models as Mediators.* Cambridge, Cambridge University Press.

Morrison, M. & Morgan, M. S. (1999b) Models as Mediating Instruments. IN Morrison, M. & Morgan, M. S. (Eds.) *Models as Mediators.* Cambridge, Cambridge University press.
Moss, L. (2004) *What Genes Can't Do*, MIT press.
Nersessian, N. J. (1999) Model-Based Reasoning in Conceptual Change. IN Magnani, L., Nersessian, N. J. & Thagard, P. (Eds.) *Model-Based Reasoning in Scientific Discovery.* New York, Kluwer Academic.
Nersessian, N. J. (2005) Interpreting Scientific and Engineering Practices: Integrating the Cognitive, Social, and Cultural Dimensions. IN Gorman, M., Tweney, R., Gooding, D. & Kincannon, A. (Eds.) *Scientific and Technological Thinking.* London, Lawrence Erlbaum Associates.
Nersessian, N. J. & Newstetter, W. C. (Forthcoming) Interdisciplinarity in Engineering Research and Learning. IN Johri, A. & Olds, B. M. (Eds.) *Cambridge Handbook of Engineering Education Research.* New York, Cambrideg University Press.
Newsholme, E. A. & Start, C. (1973) *Regulation in Metabolism,* London, John Wiley & Sons
Nicholson, D. J. (2014) The Return of the Organism as a Fundamental Explanatory Concept in Biology. *Philosophy Compass,* 9**,** 347-359.
Nickelsen, K. (2007) Otto Warburg's First Approach to Photosynthesis. *Photosynthesis research,* 92**,** 109-120.
Nickles, T. (1990) Discovery Logics. *Philosophica,* 45**,** 7-32.
O'malley, M. A. & Dupré, J. (2005) Fundamental Issues in Systems Biology. *Bioessays,* 27**,** 1270-1276.
O'malley, M. A. & Soyer, O. S. (2012) The Roles of Integration in Molecular Systems Biology. *Studies in the History and Philosophy of Biological and Biomedical Sciences,* 43**,** 58-68.
Oberhardt, M. A., Palsson, B. Ø. & Papin, J. A. (2009) Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology,* 5**,** 1-15.
Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. & Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research,* 27**,** 29-34.
Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M. & Palsson, B. Ø. (2011) A Comprehensive Genome-Scale Reconstruction of Escherichia coli Metabolism—2011. *Molecular Systems Biology,* 7**,** 1-9.
Palsson, B. Ø. (1997) What Lies Beyond Bioinformatics? *Nature Biotechnology,* 15**,** 3-4.
Palsson, B. Ø. (2006) *Systems Biology: Properties of Reconstructed Networks,* Cambridge, Cambridge University Press.
Palsson, B. Ø. & Zengler, K. (2010) The Challenges of Integrating Multi-Omic Data Sets. *Nature Chemical Biology,* 6**,** 787-789.
Papoutsakis, E. T. (1984) Equations and Calculations for Fermentations of Butyric Acid Bacteria. *Biotechnology and Bioengineering,* 26**,** 174-187.
Park, D. J. M. (1974) The Hierarchical Structure of Metabolic Networks and the Construction of Efficient Metabolic Simulators. *Journal of Theoretical Biology,* 46**,** 31-74.
Parker, J. N., Vermeulen, N. & Penders, B. (2010) *Collaboration in the New Life Cciences,* Farnham, Ashgate Publishing.
Parker, W. S. (2009) Does Matter really Matter? Computer Simulations, Experiments, and Materiality. *Synthese,* 169**,** 483-496.

Penders, B., Horstman, K. & Vos, R. (2008) Walking the Line between Lab and Computation: The “Moist” Zone. *Bioscience,* 58**,** 747-755.
Pitt, J. C. (2001) The Dilemma of Case Studies: Toward a Heraclitian Philosophy of Science. *Perspectives on Science,* 9**,** 373-382.
Plutynski, A. (2008) Explaining How and Explaining Why: Developmental and Evolutionary Explanations of Dominance. *Biology & Philosophy,* 23**,** 363-381.
Porteous, J. (1985) Control Confused. *Trends in Biochemical Sciences,* 10**,** 14-15.
Porteous, J. W. (1983) Sound Practice Follows from Sound Theory—The Control Analysis of Kacser and Burns Evaluated. *Trends in Biochemical Sciences,* 8**,** 200-202.
Powell, A. & Dupré, J. (2009) From Molecules to Systems: The Importance of Looking Both Ways. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 40**,** 54-64.
Powell, A., O'malley, M. A., Müller-Wille, S., Calvert, J. & Dupré, J. (2007) Disciplinary Baptisms: A Comparison of the Naming Stories of Genetics, Molecular Biology, Genomics, and Systems Biology. *History and Philosophy of the Life Sciences,* 29**,** 5-32.
Puchałka, J., Oberhardt, M. A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K. N., Papin, J. A. & Dos Santos, V. a. M. (2008) Genome-Scale Reconstruction and Analysis of the Pseudomonas putida KT2440 Metabolic Network Facilitates Applications in Biotechnology. *PLOS Computational Biology,* 4**,** e1000210.
Racker, E. (1985) Concepts Confused. *Trends in Biochemical Sciences,* 10**,** 270.
Rajan, K., Sunder. & Leonelli, S. (2013) Introduction: Biomedical Trans-actions, Postgenomics, and Knowledge/Value. *Public Culture,* 25**,** 463-475.
Raman, K. & Chandra, N. (2009) Flux Balance Analysis of Biological Systems: Applications and Challenges. *Briefings in Bioinformatics,* 10**,** 435-449.
Rankin Small, J. & Fell, D. A. (1989) The Matrix Method of Metabolic Control Analysis: Its Validity for Complex Pathway Structures. *Journal of Theoretical Biology,* 136**,** 181-197.
Rapoport, T. A., Heinrich, R. & Rapoport, S. M. (1976) The Regulatory Principles of Glycolysis in Erythrocytes In Vivo and In Vitro. A Minimal Comprehensive Model Describing Steady States, Quasi-Steady States and Time-Dependent Processes. *Biochemical Journal,* 154**,** 449-469.
Reed, J. L. & Palsson, B. Ø. (2003) Thirteen Years of Building Constraint-Based In Silico Models of Escherichia coli. *Journal of Bacteriology,* 185**,** 2692-2699.
Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. Ø. (2003) An Expanded Genome-Scale Model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology,* 4**,** 1-12.
Resnik, D. (1996) Adaptationism: Hypothesis or Heuristic? *Biology and Philosophy,* 12**,** 39-50.
Rheinberger, H. J. (1997a) Experimental Complexity in Biology: Some Epistemological and Historical Remarks. *Philosophy of Science,* 64**,** S245-S254.
Rheinberger, H. J. (1997b) *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube,* Stanford, Stanford University Press.
Rheinberger, H. J. (2002) A Note on Time and Biology. IN Auxier, R. E. & Hahn, L. E. (Eds.) *The Philosophy of Marjorie Grene.* Illinois, Open Court.
Rheinberger, H. J. (2010) *An Epistemology of the Concrete: Twentieth-Century Histories of Life,* London, Duke University Press.

Sarkar, S. (1992) Models of Reduction and Categories of Reductionism. *Synthese,* 91**,** 167-194.
Sarkar, S. (1998) *Genetics and Reductionism,* Cambridge, Cambridge University Press.
Sarkar, S. (Ed.) (2001) *The Philosophy and History of Molecular Biology: New Perspectives,* London, Kluwer Academic Publishers.
Sarkar, S. (2003) Reduction: A Philosophical Analysis. *Encyclopedia of Life Sciences.* Macmillan Publishers Ltd.
Savageau, M. (1990) Biochemical Systems Theory: Alternative Views of Metabolic Control. IN Cornish-Bowden, A. & M.L., C. (Eds.) *Control of Metabolic Processes.* London, Plenum Press.
Savageau, M. A. (1969a) Biochemical Systems Analysis: I. Some Mathematical Properties of the Rate Law for the Component Enzymatic Reactions. *Journal of Theoretical Biology,* 25**,** 365-369.
Savageau, M. A. (1969b) Biochemical Systems Analysis: II. The Steady-State Solutions for an n-pool System Using a Power-Law Approximation. *Journal of Theoretical Biology,* 25**,** 370-379.
Savageau, M. A. (1971) Concepts Relating the Behavior of Biochemical Systems to their Underlying Molecular Properties. *Archives of Biochemistry and Biophysics,* 145**,** 612-621.
Savageau, M. A. (1972) The Behaviour of Intact Biochemical Control Systems *Current Topics in Cellular Regulation* 6**,** 63-130.
Savageau, M. A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology,* Reading Massachusetts, Addison-Wesley Publishing Company.
Savageau, M. A. (1987) Control of Metabolism: Where is the Theory? *Trends in Biochemical Sciences,* 12**,** 219-220.
Savageau, M. A. (1991) Biochemical Systems Theory: Operational Differences Among Variant Representations and their Significance. *Journal of Theoretical Biology,* 151**,** 509-530.
Savageau, M. A. (1992) Dominance According to Metabolic Control Analysis: Major Achievement or House of Cards? *Journal of Theoretical Biology,* 154**,** 131-136.
Savageau, M. A. & Sorribas, A. (1989) Constraints Among Molecular and Systemic Properties: Implications for Physiological Genetics. *Journal of Theoretical Biology,* 141**,** 93-115.
Savageau, M. A., Voit, E. O. & Irvine, D. H. (1987a) Biochemical Systems Theory and Metabolic Control Theory: 1. Fundamental Similarities and Differences. *Mathematical Biosciences,* 86**,** 127-145.
Savageau, M. A., Voit, E. O. & Irvine, D. H. (1987b) Biochemical Systems Theory and Metabolic Control Theory: 2. The Role of Summation and Connectivity Relationships. *Mathematical Biosciences,* 86**,** 147-169.
Savinell, J. M. & Palsson, B. Ø. (1992) Network Analysis of Intermediary Metabolism using Linear Optimization. I. Development of Mathematical Formalism. *Journal of Theoretical Biology,* 154**,** 421-454.
Schaffner, K. F. (1969) The Watson-Crick Model and Reductionism. *The British Journal for the Philosophy of Science,* 20**,** 325-348.
Schellenberger, J., Park, J., Conrad, T. & Palsson, B. Ø. (2010) BiGG: A Biochemical Genetic and Genomic Knowledgebase of Large Scale Metabolic Reconstructions. *BMC Bioinformatics,* 11**,** 1-10.

Schickore, J. (2011) More Thoughts on HPS: Another 20 Years Later. *Perspectives on Science,* 19**,** 453-481.
Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S. & Palsson, B. Ø. (2002) Genome-scale Metabolic Model of Helicobacter pylori 26695. *Journal of Bacteriology,* 184**,** 4582-4593.
Schilling, C. H., Edwards, J. S. & Palsson, B. Ø. (1999a) Toward Metabolic Phenomics: Analysis of Genomic Data using Flux Balances. *Biotechnology Progress,* 15**,** 288-295.
Schilling, C. H. & Palsson, B. Ø. (1998) The Underlying Pathway Structure of Biochemical Reaction Networks. *Proceedings of the National Academy of Sciences,* 95**,** 4193-4198.
Schilling, C. H., Schuster, S., Palsson, B. Ø. & Heinrich, R. (1999b) Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-Genomic Era. *Biotechnology Progress,* 15**,** 296-303.
Schomburg, I., Chang, A. & Schomburg, D. (2002) BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Research,* 30**,** 47-49.
Schuetz, R., Kuepfer, L. & Sauer, U. (2007) Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in Escherichia coli. *Molecular Systems Biology,* 3**,** 1-15.
Segre, D., Vitkup, D. & Church, G. M. (2002) Analysis of Optimality in Natural and Perturbed Metabolic Networks. *Proceedings of the National Academy of Sciences,* 99**,** 15112-15117.
Serres, M. H., Gopal, S., Nahum, L. A., Liang, P., Gaasterland, T. & Riley, M. (2001) A Functional Update of the Escherichia coli K-12 Genome. *Genome Biology,* 2**,** 1-7.
Shapere, D. (1984) What Can the Theory of Knowledge Learn from the History of Knowledge? *the Monist,* 60**,** 488-508.
Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø. & Ruppin, E. (2008) Network-Based Prediction of Human Tissue-Specific Metabolism. *Nature Biotechnology,* 26**,** 1003-1010.
Shlomi, T., Herrgard, M., Portnoy, V., Naim, E., Palsson, B. Ø., Sharan, R. & Ruppin, E. (2007) Systematic Condition-Dependent Annotation of Metabolic Genes. *Genome Research,* 17**,** 1626-1633.
Simon, H. A. (1962) The Architecture of Complexity. *Proceedings of the American Philosophical Society,* 106**,** 467-482.
Simon, H. A. (1977) The Organization of Complex Systems. IN Simon, H. A. (Ed.) *Models of Discovery.* Dordrecht, D. Reidel Publishing Company.
Simon, H. A. & Ando, A. (1961) Aggregation of Variables in Dynamic Systems. *Econometrica,* 29**,** 111-138.
Sismondo, S. (2000) Island Biogeography and the Multiple Domains of Models. *Biology and Philosophy,* 15**,** 239-258.
Sols, A. & Marco, R. (1970) Concentrations of Metabolites and Binding Sites. Implications in Metabolic Regulation. *Current Topics in Cellular Regulation,* 2**,** 227-273.
Sorribas, A. & Savageau, M. A. (1989a) A Comparison of Variant Theories of Intact Biochemical Systems. I. Enzyme-Enzyme Interactions and Biochemical Systems Theory. *Mathematical Biosciences,* 94**,** 161-193.
Sorribas, A. & Savageau, M. A. (1989b) A Comparison of Variant Theories of Intact Biochemical Systems. II. Flux-oriented and Metabolic Control Theories. *Mathematical Biosciences,* 94**,** 195-238.

Sorribas, A. & Savageau, M. A. (1989c) Strategies for Representing Metabolic Pathways within Biochemical Systems Theory: Reversible Pathways. *Mathematical Biosciences,* 94**,** 239-269.
Srere, P. A. (1967) Enzyme Concentrations in Tissues. *Science,* 158**,** 936-937.
Steinle, F. & Burian, R. (2002) Introduction: History of Science and Philosophy of Science. *Perspectives on Science,* 10**,** 391-397.
Strand, R. (1999) Towards a Useful Philosophy of Biochemistry: Sketches and Examples. *Foundations of Chemistry,* 1**,** 269-292.
Strasser, B. J. (2008) GenBank--Natural History in the 21st Century? *Science,* 322**,** 537-538.
Suárez, M. (Ed.) (2008) *Fictions in Science: Philosophical Essays on Modeling and Idealisation,* Abingdon, Oxon, Routledge.
Thiele, I., Hyduke, D. R., Steeb, B., Fankam, G., Allen, D. K., Bazzani, S., Charusanti, P., Chen, F.-C., Fleming, R. M. & Hsiung, C. A. (2011) A Community Effort Towards a Knowledge-Base and Mathematical Model of the Human Pathogen Salmonella Typhimurium LT2. *BMC systems biology,* 5**,** 1-9.
Thiele, I. & Palsson, B. Ø. (2010a) A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nature Protocols,* 5**,** 93-121.
Thiele, I. & Palsson, B. Ø. (2010b) Reconstruction Annotation Jamborees: A Community Approach to Systems Biology. *Molecular Systems Biology,* 6**,** 1-4.
Thiele, I., Swainston, N., Fleming, R. M., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O. & Stobbe, M. D. (2013) A Community-Driven Global Reconstruction of Human Metabolism. *Nature Biotechnology,* 31**,** 419-425.
Umbarger, H. E. (1956) Evidence for a Negative-Feedback Mechanism in the Biosynthesis of Isoleucine. *Science,* 123**,** 848.
Van Dam, K. (1986) Biochemistry is a Quantitative Science. *Trends in Biochemical Sciences,* 11**,** 13-14.
Varma, A. & Palsson, B. Ø. (1993) Metabolic Capabilities of Escherichia coli: I. Synthesis of Biosynthetic Precursors and Cofactors. *Journal of Theoretical Biology,* 165**,** 477-502.
Varma, A. & Palsson, B. Ø. (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nature Biotechnology,* 12**,** 994-998.
Vermeulen, N., Parker, J. N. & Penders, B. (2013) Understanding Life Together: A Brief History of Collaboration in Biology. *Endeavour***,** 162-171.
Voit, E. O. & Savageau, M. A. (1987) Accuracy of Alternative Representations for Integrated Biochemical Systems. *Biochemistry,* 26**,** 6869-6880.
Waddington, C. H. & Kacser, H. (1957) *The Strategy of the Genes,* London, Allen and Unwin.
Wagner, A. (2005a) *Robustness and Evolvability in Living Systems,* Oxford, Princeton.
Wagner, A. (2005b) Robustness, Evolvability, and Neutrality. *FEBS Letters,* 579**,** 1772-1778.
Wagner, A. (2011) *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems,* New York, Oxford University Press
Watson, J. D. & Crick, F. H. (1953) Molecular Structure of Nucleic Acids. *Nature,* 171**,** 737-738.
Watson, M. R. (1984) Use of the Apple II in Teaching Metabolism. IN Wood, E., J. (Ed.) *Microcomputers in Biochemical Education.* London, Taylor and Francis.

Watson, M. R. (1986) A Discrete Model of Bacterial Metabolism. *Computer Applications in the Biosciences: Cabios,* 2**,** 23-27.
Weber, M. (2002) Theory Testing in Experimental Biology: The Chemiosmotic Mechanism of ATP Synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences,* 33**,** 29-52.
Weber, M. (2004) *Philosophy of Experimental Biology,* New York, Cambridge University Press.
Weisberg, M. (2006) Forty years of 'The strategy': Levins on Model Building and Idealization. *Biology and Philosophy,* 21**,** 623-645.
Weisberg, M. (2007) Three Kinds of Idealization. *the Journal of Philosophy,* 104**,** 639-659.
Westerhoff, H. V. & Palsson, B. Ø. (2004) The Evolution of Molecular Biology into Systems Biology. *Nature Biotechnology,* 22**,** 1249-1252.
Wimsatt, W. C. (1972) Complexity and Organization. IN Grene, M. & Mendelsohn, E. (Eds.) *Topics in the Philosophy of Biology.* Dordrecht, D. Reidel Publishing Group.
Wimsatt, W. C. (1976) Reductionism, Levels of Organization, and the Mind-Body Problem. IN Globus, G. G., Maxwell, G. & Savodnik, I. (Eds.) *Consciousness and the Brain: A Sceinetific and Philosophical Enquiry.* New York, Plenum.
Wimsatt, W. C. (1986) Forms of Aggregativity. IN Donagan, B., Perovich, A. & Wedin, M. (Eds.) *Human Nature and Natural Knowledge.* Dordrecht, D.Reidel Publishing Company.
Wimsatt, W. C. (2007a) On Building Reliable Pictures with Unreliable Data: An Evolutionary and Developmental Coda for the New Systems Biology. *Systems Biology: Philosophical Foundations.*
Wimsatt, W. C. (2007b) *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality,* London, Harvard University press.
Winsberg, E. (2001) Simulations, Models, and Theories: Complex Physical Systems and their Representations. *Philosophy of Science,* 68**,** S442-S454.
Winsberg, E. (2003) Simulated Experiments: Methodology for a Virtual World. *Philosophy of Science,* 70**,** 105-125.
Winsberg, E. (2006) Models of Success versus the Success of Models: Reliability without Truth. *Synthese,* 152**,** 1-19.
Winther, R. (2006a) Fisherian and Wrightian Perspectives in Evolutionary Genetics and Model-Mediated Imposition of Theoretical Assumptions. *Journal of Theoretical Biology,* 240**,** 218-232.
Winther, R. G. (2006b) Parts and Theories in Compositional Biology. *Biology and Philosophy,* 21**,** 471-499.
Winther, R. G. (2011) Part-Whole Science. *Synthese,* 178**,** 397-427.
Wright, B. E. & Kelly, P. J. (1981) Kinetic Models of Metabolism in Intact Cells, Tissues, and Organisms. *Current Topics in Cellular Regulation,* 19**,** 103-158.
Wylie, A. (1994) Discourse, Practice, Context: From HPS to Interdisciplinary Science Studies. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association.* University of Chicago Press.
Yates, R. A. & Pardee, A. B. (1956) Control of Pyrimidine Biosynthesis in Escherichia coli by a Feed-Back Mechanism. *Journal of Biological Chemistry,* 221**,** 757-770.