

The roles of integration in molecular systems biology

Authors

Maureen A. O'Malley

Department of Philosophy, Quadrangle A14, University of Sydney, NSW 2066

Phone: +61 (0)2 90366591

Email: maureen.omalley@sydney.edu.au (corresponding author)

Orkun S. Soyer

School of Engineering, Computing and Mathematics, Harrison Building, North
Park Road, University of Exeter, EX4 4QF

Email: O.S.Soyer@exeter.ac.uk

Outline

Abstract

1. Introduction
 - 1.1. From DD and HD to integration
2. Integration: What is it?
 - 2.1. Methodological integration
 - 2.2. Data integration
 - 2.3. Explanatory integration
3. Exemplifying integration
 - 3.1. Vesicle transport in mathematical cell biology
 - 3.2. Noise biology
 - 3.3. Evolutionary systems biology
4. Integration in context
5. Conclusion
6. Acknowledgements
7. References

Abstract

A common way to think about scientific practice involves classifying it as hypothesis- or data-driven. We argue that although such distinctions might illuminate scientific practice very generally, they are not sufficient to understand the day-to-day dynamics of scientific activity and the development of programmes of research. One aspect of everyday scientific practice that is beginning to gain more attention is integration. This paper outlines what is meant by this term and how it has been discussed from scientific and philosophical points of view. We focus on methodological, data and explanatory integration, and show how they are connected. Then, using some examples from molecular systems biology, we will show how integration works in a range of inquiries to generate surprising insights and even new fields of research. From these examples we try to gain a broader perspective on integration in relation to the contexts of inquiry in which it is implemented. In today's environment of data-intensive large-scale science, integration has become both a practical and normative requirement with corresponding implications for meta-methodological accounts of scientific practice. We conclude with a discussion of why an understanding of integration and its dynamics is useful for philosophy of science and scientific practice in general.

Keywords

Integration, iteration, systems biology, data-driven research, hypothesis-driven research

1. Introduction

Data-driven (DD) science is often contrasted to hypothesis-driven (HD) science. Using this classification to distinguish between different scientific activities is part of a broader attempt to understand the dynamics of contemporary life sciences as they deal with a flood of molecular data. We start from the position that DD and HD categories are insufficient to capture in any fine-grained way the processes occurring in today's life sciences, and that instead, a focus on *integration* will be useful. In what follows, we will discuss what we mean by integration and how it works in a range of examples from recent molecular biology, especially research that is concerned with the molecular dynamics of biological systems. Our examples will show how the integration of methods, bodies of data and explanations can produce novel insights into biological phenomena, stimulate new fields of research, and generally reconfigure expectations of scientific practice. Finally, we will discuss integration more broadly, as part of a *translational* dynamic that shapes science today, from policy all the way down to what happens in the lab and at the computer screen. We conclude that understanding integration is useful for philosophy of science and scientific practice in general because it shows how novel modes of inquiry can

develop within a complex landscape of exploratory questioning, tool development, and transfers between very different systems of knowledge and practice.

1.1. From DD and HD to integration

The theme of this special issue, DD science, forms an important part of the historical context in which the contemporary scientific emphasis on integration has developed. Commentators who use the term DD are referring to research that is characterized by the generation, collection and potential interpretation of large bodies of biological data, from which subsequent analyses can detect new relationships, processes and phenomena. There is usually an assumption that no particular hypothesis is being tested in a standard DD inquiry. However, DD research is often subsumed by a larger aim, which is to generate new hypotheses in the process of being guided (often implicitly) by older ones. Conversely, HD approaches are discussed explicitly as strategies in which the first step in a cycle of research is to formulate a hypothesis on the basis of existing knowledge. Experiments are designed and carried out in order to assess particular hypotheses. At least in principle, this type of research is clearly defined with an obvious starting point and a specified set of procedures, the outcomes of which lead to the acceptance, modification or rejection of the original hypothesis.

Although many practitioners and observers of science believe both DD and HD modes of practice work in concert, there is also a general admission that in an era of high-throughput data collection (such as genome sequencing), emphases on the data-gathering side of scientific activity may have become stronger, with both positive and negative consequences (e.g., Brent, 2000; Allen, 2001; Strasser, 2008; Kell and Oliver, 2004; Kitano, 2002; Aebersold et al., 2000). For those with concerns about this turn, the advent of systems biology is perceived as a sensible return to hypothesis testing and an admission that new scientific insights cannot be brought about 'merely' by gathering data and hoping for inductive insight (White, 2010; Weston and Hood, 2004). Our aim in this paper is to investigate such claims not from the assumption that one or the other approach might be better, but that there is something important to both strategies for the production of new insight and practices. However, the level at which DD and HD approaches are usually discussed is too general to see exactly how they work, and indeed, how they could be more than idealizations that have the primary function of describing historical shifts in life-science methodologies.

In order to examine more closely what is actually occurring in contemporary molecular biology, we focus on an increasingly common emphasis in such sciences, which is that of the process of integration. This term is used to cover a multi-faceted dynamic, in which methods, bodies of data and explanations are synthesized in order to understand and intervene more effectively in biological systems. Integration can also include the combination of DD approaches with HD approaches, and emerging approaches to molecular biology often require this. Systems biology, the successor science to standard DD genomics, is usually

described as the field that has arisen out of a perceived need for more integration, particularly of data and methods but also of disciplines. Although systems biologists and philosophers have discussed the necessity of integrative strategies, we will suggest that existing accounts have yet to examine fully the dynamics of integration as they occur in today's molecular life sciences.

One stimulus for us to think more about integration came from a meeting held in 2010 at the University of Exeter, called 'Frontiers of Multidisciplinary Research: Mathematics, Engineering and Biology'.¹ This meeting brought together scientists from diverse backgrounds, including mathematics, physics, engineering and biology. Despite the different disciplines, the group shared a common interest in questions about dynamic biological processes and how to comprehend them. Presentations and discussions at the meeting repeatedly stressed the value of combining 'wet' experimental practices with 'dry' theoretical approaches (mathematical modelling of various sorts). Discussions following presentations repeatedly addressed the problems of achieving integration and how to solve them (see Section 5). We draw our examples in Section Three from research presented at the meeting, in part because of its focus on integration rather than merely the label 'systems biology'. While that label can obscure more than it reveals, a focus on integration can open up more clearly how systems biology operates.

Using the exemplars of a particular study in mathematical cell biology, and the more general projects of noise biology and evolutionary systems biology, we will explore the combinations of methods, data and explanations used in these programmes of research in order to show how these integrative efforts have produced novel insight into biological phenomena. In some of this exemplary work, more than new insight has been generated: wholly unexpected phenomena have been discovered and entire fields launched because of these discoveries. Integration, as a normative methodology, gains considerable momentum from such successes and this leads to the restructuring of accounts of how science works. Our analyses of these examples will show that amongst the factors conducive to integration are background conditions of exploratory questioning, technological development, and the ability to transfer existing systems of inquiry into new domains of phenomena and research. We suggest that these factors are of more importance to understand integration and scientific change than an emphasis on DD or HD approaches, even if these terms continue to have broad classificatory virtues.

2. Integration: What is it?

Many sorts of activities and entities can be integrated in scientific practice. Even in very traditional low-throughput ways of practising science, integration of a kind

¹ www.exeter.ac.uk/research/excellence/keythemes/systemsbiology/frontiers/

occurs when one line of evidence is used to back up a hypothesis already supported by another line of evidence. But in this guise, integration is only sometimes seen as something that needs to be addressed explicitly, when cross-testing, consilience or triangulation prove troublesome or in need of justification (e.g., Padian, 2001; Star, 1986; Wimsatt, 2007a [1981]). This is what has changed in contemporary molecular biology, which is blessed (or cursed) with a superabundance of data and no easy way in which to make sense of it. Although some progress can be made by focusing on specific subsets of large datasets, a general motivation has developed in the post-genomic era of molecular biology to create new ways of integrating different data types, methods and models, as well as different disciplinary approaches and expertise (Ideker et al., 2001; Gunawardena, 2010; Lerman and Palsson, 2010; Chen et al., 2010). Systems biology bases its justification on the recognition that in order to understand function, it is essential to develop multilevel interpretations and manipulations of datasets by addressing different levels of questions, set within different explanatory frameworks, with a range of methods (Auffray et al., 2003). Integration is the term commonly used to encompass this desideratum and its achievement.

In our analysis, integration encompasses the combination of methods and methodologies (sometimes including general methodological approaches such as HD and DD), the process of making data sets comparable and re-analysable, and the variety of ways in which explanations are brought together in a particular inquiry. Although some views of integration in systems biology take a circumscribed methodological perspective, in that they idealize integration within an experiment-modeling cycle of hypothesis-generation and testing (see Section 2.1), our view draws on the notion of integrating whatever methods, data and explanations might be usefully combined to develop a novel and productive line of inquiry about dynamic biological processes for which large datasets exist. We also draw on philosophical discussions about integration, sometimes to distinguish what we are saying from older philosophies of theoretical 'unification', and other times to expand on more practical philosophical accounts of integration.

2.1. Methodological integration

It is common in systems biology to think of integration methodologically. This sort of integration involves directing a range of methods at a particular biological system or research problem in order to gain a multidimensional understanding of how the system works. Understanding means not just explanation, but also (very importantly for systems biology) prediction and control. We are not referring here to older philosophical notions of 'unity of method', in which claims were made that *the* scientific method had a single logic (see Cat, 2007), but to the combination of specific methods and sometimes more general methodologies. This integration need not happen sequentially. Different methods can be employed simultaneously, as they might be in a collaboration of several researchers, but they remain focused on the same biological system. While the

range of methods is not prescribed in molecular systems biology, it very commonly involves the combination of experimental, observational, mathematical and bioinformatic methods along with visualization tools (Bruggeman and Westerhoff, 2007). Mathematical modelling is considered quite central to systems biology. In fact, if mathematical models are not brought into the inquiry at some point, most practitioners would think the research is not systems biology (Palsson, 2006; Westerhoff and Kell, 2007). In our discussion, we relax this definition so as to include research examining large-scale molecular dynamics in biological systems, even if mathematical models have not yet been developed to represent those dynamics.

Another concept that has been used to cover how methods are combined over time in systems analysis is iteration. Traditionally, iteration means to repeat a process over and over again, usually with the output of the previous round forming the basis of the next round of the operation. Mathematical iteration works on this basis, and the iteration of simple recursive equations can lead to complex patterns such as those in images of fractals, such as Mandelbrot and Julia sets. Iteration is also thought of more prosaically, in technology and software design in particular, where successive rounds of planning, testing and revision are conceived of as iteration in contrast to more linear and simplified 'waterfall' representations of the process (Boehm, 1988; Nielsen, 1993). In the manifesto statements of systems biology that proliferated in the early years of the 2000s, iteration was often discussed as the cyclic application of methods one after the other in attempts to generate functional understanding from large-scale datasets (e.g., Kitano, 2002; Aderem, 2005; Bruggeman and Westerhoff, 2006). Even prior to systems biology, however, multiple iterations of hypothesis formulation, testing and revision were seen as desirable and necessary to scientific progress (Understanding Science, 2010; Good, 1999).²

We think it is useful to distinguish between two different senses of iteration in order to understand integration better. One sense entails the iteration of very specific methods in an anticipated cycle of inquiry, especially the serial repetition of experiment and mathematical modelling in order to refine models by individualized hypothesis testing (e.g., Aderem, 2005). This perspective on iteration matches most closely the HD account of scientific practice. The other sense of iteration fits more closely what we mean by the methodological dimension of integration. It covers a much broader and less predictable combination of approaches. It can involve not only the iteration of particular methods, but also the interplay between hypothesis-driven and data-driven approaches (Kell and Oliver, 2004; O'Malley et al., 2009). Integration in this broader methodological sense is not generally guided by a specific hypothesis, but more by a general determination to probe into an area of biological inquiry or

² We are leaving out from this section some philosophical accounts of iteration because we think they are more usefully discussed in the section on explanatory integration (Section 3.3).

a biological phenomenon. It is not data-driven per se, although it depends on the existence of large datasets and may well give rise to new bodies of data. Methodological integration in this broad sense can often be method- or technique-driven, in that the mere availability of a method or a technology inspires its application to a new domain of phenomena or a novel problem to which other methods have already been applied.

A useful way to consider methodological integration is by looking at the relationship between systems and synthetic biology. Very often, the two are connected — at the very least because synthetic biology relies on systems biology for the design underpinning any synthetic construction (Smolke and Silver, 2011). Methodological integration is crucial to the systems-synthetic biology interface because of the way each set of practices combines engineering and biological methods. This integration is sometimes described as 'combinatorial' methodology because it combines mathematical representations and experimental data in a synthetic biological construction (Knuutila and Loettgers, 2010).³ This materialized model can be said to embody the combinations of methods used in its creation. In addition, synthetic biology is built on the interchangeability of material components in a way that has affinities with the idea of data integration (see Section 3.2).

A further set of practices that should be considered under methodological integration are those referred to as machine-based approaches, which consist of roboticized experimentation, automated literature mining for hypotheses, and automated analysis of models. Although most accounts of machine science suggest they will be integrated with human cognitive activities (King et al., 2011; Evans and Rzhetsky, 2010), proponents also look forward to less of the 'messy' discovery-oriented biology carried out by human researchers. As some of the participants suggested at the Frontiers workshop (see Section 2), machine science advocates conceive of a thorough design process, based on complete data, that results in the true engineering of biological systems and the effective integration of methods from engineering and biology.

In summary, whatever the techniques and technologies used, methodological integration in the life sciences relies on the combination of methods and methodologies — not necessarily in any predetermined order — to produce knowledge not obtained (and probably not obtainable) by single-method or even single-discipline approaches. The linear reapplication of methods, which we would call iteration, is the simplest way of achieving methodological integration in systems biology, but integration often occurs in much more methodologically complex forms (see Section 3.2). Some philosophers of science have used case-

³ It is combinatorial in a more exact methodological sense because it integrates semi-randomized DNA rearrangement (called 'combinatorial synthesis') with the rational design of the construct (Guet et al., 2002; Blake and Issacs, 2004; Michalodimitrakis and Isalan, 2009; Blake and Issacs, 2004).

studies to argue for the importance of methodological integration, going so far as to suggest that only by understanding the dynamics of underlying methods can theoretical integration be comprehended (Bechtel, 1993; Grantham, 2004). While it is certainly the case that combinations of methods were not uncommon in earlier molecular biology and much farther afield (Grantham, 2004; Bechtel, 1993), what is different now is both the self-consciousness and the scale on which methodological integration has been implemented. A further distinction can be made on the basis of the data with which these combinations of methods interact.

2.2. Data integration

The establishment of molecular systems biology as the ongoing development of genomics was built on the creation of shared databases of sequence and other information, as well as the recognition of the difficulties of squeezing suitable amounts of biological meaning out of such copious datasets. In molecular systems biology, data integration can sometimes be discussed as the process of merging disparate sources of data through appropriate methods, as if data integration were in fact a form of methodological integration (e.g., Schmid and Blank, 2010). Much more commonly, however, data integration is understood as the legacy of the era in which a flood of sequence and other data was produced by high-throughput methods, in close association with bioinformatic analysis. Often conceived as a problem or at least a major challenge, data integration is the activity of making comparable different data types from a huge variety of potentially inconsistent sources.

Data integration thus refers to the process of theorizing and modelling databases, quantifying data accurately, developing standardization procedures, cleaning data, and providing efficient and user-friendly interfaces to enable data not only to be reused, but reanalysed and combined in novel ways (Lanzerini, 2002; Ge et al., 2003; Leonelli, 2008). It has the aim of amalgamating diverse datasets of sub-systems so that system-level analyses can be performed and 'ordinary' biology (which produced the sub-system data) transformed into systems biology. Autonomous datasets have to be reconciled, regardless of their sources and original mode of analysis, so that users are able to query and re-examine the data in ways that were not even thought of when the data were generated. If data are integrated successfully, users can take advantage of multiple sets of data without needing to worry about how the data were originally produced and how they were merged (Calvi et al., 2002; Calvanese and Giacomo, 2005; Cary et al., 2005). This is also the rationale that underlies the repositories of standardized biological parts produced for sharing by synthetic biologists (Endy, 2005).

In other words, by abstracting data from their original sources, the integration procedure forms a new body of information that can be treated as a unified whole (Ideker et al., 2007). Obviously, there are many practical and theoretical routes to such integration, and not all of them are trouble-free, but much progress has

been made in molecular systems biology over the last decade (Hwang et al., 2005; Ge et al., 2003; Ng et al., 2006). Numerous future challenges remain in regard to synthesizing different types of data, diverse standardization regimes, different scales of datasets, biased data accumulations (e.g., from favouring particular interactions or organisms) and noisy data, as well as achieving efficiency of use (Triplet and Butler, 2011; Sullivan et al., 2010; Kitano, 2002; Hwang et al., 2005). Many molecular systems biologists believe that data integration is the crucial computational task for systems biology, and that it may even be the fundamental basis of systems biology (e.g., Philippi and Köhler, 2006). However, it goes without saying that database integration relies on conceptual and methodological resources, and just as importantly, the visualization tools needed to accompany integrated datasets so biologists can comprehend intuitively the masses of data they are analysing (Fox and Hendler, 2011; Aderem, 2005). Just being able to access computationally integrated data would be meaningless, and to us this indicates that integration has also to be discussed at the explanatory level.

2.3. Explanatory integration

The two preceding processes of integration are intimately connected to explanatory integration, which refers to both the synthesis of different explanations as well as the import of explanatory (and predictive) models from other research into a specified domain of inquiry. These models are not usually full-blown theories, but are mechanistic or statistical representations of particular phenomena (Arkin and Schaffer, 2011). They consist of conceptual resources and mathematical formalisms, with uncertain limits to their application. Often, methodological integration implies or leads to explanatory integration, when combinations of different methods invoke the fusion of explanatory frameworks, commonly in the accommodation of multiple levels of phenomena (Gudelj et al., 2010). Similarly, explanatory frameworks are used to probe as well as structure integrated databases (Vidal et al., 2011).

Because our examples in Section Three elaborate on the transfer and integration of explanations from one research domain to another, we will outline in this section only explanatory synthesis — perhaps the most obvious form of explanatory integration. In molecular systems biology, explanatory synthesis usually refers to the process by which a variety of techniques are used to unify models and their associated explanatory resources (Patel and Nagi, 2010). Sometimes this process of integration is described as ‘piecing together’ the explanatory puzzle of a complex system, such as disease (Kirkwood et al., 2003). In order to make individual mathematical models available for explanatory integration, shared-use databases have been created (e.g., www.ebi.ac.uk/biomodels; www.basis.ncl.ac.uk). These resources allow the collective synthesis and refinement of models, as well as the identification of missing explanatory resources. The concept of explanatory integration does not, however, involve the goal of a complete, unified explanation of all biology. Explanatory integration, particularly in the case of model integration, produces a

'complex ecology of models, embedded in a framework that enables debate and collaboration' (Finkelstein et al., 2004, p. 30).

Explanatory integration may have immediate resonance for philosophers of science, because the field's discussions of integration have almost always been in the form of explanatory or theoretical unification (e.g., Oppenheim and Putnam, 1958). By focusing on high-level explanatory theories, able to integrate increasingly diverse phenomena, philosophers see the potential for science to be unified, not always reductively (e.g., Kitcher, 1981; Mäki, 2001). One motivation is to seek explanations that explain as many facts as possible with as few explanatory principles and 'brute facts' as is manageable. Another well known philosophical account of integration involves 'interfield theories', in which theoretical interactions between fields (distinguished theoretically) produce new fields of research (e.g., Darden and Maull, 1977; Maull 1977). While this account resonates with some of our examples in Section Three, it is still too theory-focused to account for what we see going on in molecular systems biology.

A more recent philosophical perspective on integration emphasizes 'integrative pluralism' as the means by which complex, multifactorial biological processes can be understood (Mitchell, 2004). Single-mechanism explanations are inadequate for such complex realities, runs this argument. This view of integration is also an exclusively explanatory one. While Mitchell specifies three types of integration (mechanical rule integration, local theoretical integration, explanatory concrete integration), her account is still fixed on how scientists use theories in relation to biological systems. We argue that focusing exclusively on explanatory integration will not capture integration as it occurs in molecular systems biology and other sciences of 'complexity'. Our examples will show the necessity of considering integration more inclusively than is captured in purely theoretical explanatory integration (as if methods and data follow sheepishly behind theory). Another tendency amongst philosophers is to discuss explanatory integration in terms of disciplinary integration and this is something we will discuss briefly in Section Four.

More pragmatic philosophical accounts (of which we hope ours is one) are concerned with a range of connections between fields and research areas, rather than any overarching unification or theoretically driven integration of fields (e.g., Grantham, 2004; also Brigandt, 2010). From our observations, grander unifying senses of integration are not particularly good descriptions of day-to-day instances of the practical application of explanatory integration, in part because the models, explanations and theoretical resources are contextualized, and connections are made for practical rather than theoretical reasons. In molecular systems biology's explanatory integration, the plurality of data, methods and models on which it draws constitutes a general process of integration, for which the outcome is not primarily theoretical. Philosopher Todd Grantham (2004) sets out a range of 'unifying' (integrating) activities in biology by focusing on practical unity, in which heuristics (or models or explanations) are transferred from one

field to another. He notes that this can also occur with methods, and we will elaborate on the importance of transfer to integrative practice in Sections Three and Four.

We remarked in Section 2.1 that there are philosophical accounts of iteration but that we did not consider them as strictly or even mainly about methodological iteration. Instead, we think these treatments of iteration by philosophers of science fit fairly well within what we understand by explanatory integration. For example, Hasok Chang has advanced the notion of 'epistemic iteration', by which he means the way in which more precise understandings are produced at each stage of inquiry in a process of 'corrective evolution' (Chang, 2004, p. 253, 226, 46). Other philosophers have discussed iteration more implicitly, as a response to anomalies (Darden, 1991; Elliott, 2004), or as a satisficing stepwise progression of understanding based on imperfect knowledge and models (Nickles, 1997; Wimsatt, 2007b [1987]). We suggest these concepts of iteration accord with our discussion of explanatory integration because they involve the reconfiguration of explanations that functioned at an earlier stage of inquiry and now have to be reintegrated at a second stage, in light of new knowledge (which is produced by interactions between methods, previous explanations and data). However, epistemic iteration does not explicitly pay attention to something we consider important in explanatory integration, which is the horizontal movement of explanatory resources between different fields or research programmes (Brigandt, 2010; see Section 4). We contrast this horizontal dimension to the vertical 'evolution' of a single stream of research, which is how we perceive epistemic iteration to work. We will take up this issue again in Section Four after we have given some exemplification of all three dimensions of integration.

3. Exemplifying integration

It would be convenient if the actual process of integration as practised in molecular systems biology could be separated into the three categories we have identified, with perhaps a little overlap between them (as in a Venn diagram of minimally overlapping circles). However, not only have we left out potential spheres of integration (e.g., disciplinary — see Section 4), but also, the dynamic interactions between these three types are barely captured by such a representation. To provide some substantiation of what we mean by integration, and how a single research project may encompass all three types of integration, we will discuss a number of examples, which range from specific discoveries to the creation of new fields within today's molecular life sciences. We have selected examples that may not be labelled 'Systems Biology' as such (see Footnote 4), but which involve a range of integrative activities in regard to the analysis of molecular systems. All of them involve major conceptual revisions and technological achievements, and none of these changes could have been anticipated even 10 years earlier.

3.1. Vesicle transport in mathematical cell biology

We focus in this section on a novel understanding of a cellular transport process that was achieved by using a mathematical model to interrogate data obtained from experimental and observational methods (Schuster et al. 2011). The system under scrutiny involved the transport of vesicles (endosomes) along microtubules in a fungal hypha. This is a bidirectional transport system, in which the endosomes travel to the end of the microtubule where they are picked up by a dynein protein and shunted back along the microtubule. When the endosomes 'flip' at the end of the microtubule, only a tiny number of them fall off the track (<2%). A lot of dynein is required to make sure this happens successfully, and the process has seemed to be an obviously deterministic one that relies on strong regulation to ensure sufficient protein concentrations and effective flipping. But by exploring this process mathematically, on the basis of the intuition that models of 'particle hopping' in physical systems such as traffic flows (e.g., Nagel, 1996; see also Chowdhury et al., 2005) could be transferred to the analysis of cellular processes, the research team was able to establish that despite the regularity of the process, the dynein abundance was stochastic. The mathematical modelling in this study reinterpreted the in vivo visualization of the protein motor process, and thus added to the impact of the model's predictions about the relationships of stochastic and controlled movement of endosomes along microtubules (Ashwin et al., 2010; Schuster et al., 2011; Brown, 2010). For the biologists involved, this was a huge surprise. Although experimental biologists sometimes think (according to one of the Frontiers meeting participants; see also Kirkwood et al., 2003) 'modelling is boring — it's just a process of making data fit models', in a situation where strong regularities are not explained deterministically, a greatly revised perception of the effectiveness of modelling can arise.

This example illustrates all three aspects of integration. Quantitative data from biochemical assays and live imaging (gained from genetically engineering fluorescent proteins into cells, which are subsequently observed by solid-state laser microscopy) were reconceptualized by a model imported from the physical sciences. This import, suitably adapted, generated testable hypotheses and led to novel biological insight. The research did not involve enormous datasets, but was concerned with locally produced data (e.g., visualization and biochemical data) that had to be made comparable before they could become the basis of the mathematical modelling. The model also generated many new questions about transport processes in cells and how they might be explained. Of importance to the general context of the inquiry was that the initial stimulus of this project was not a well defined hypothesis, but an exploratory 'what if' question: 'What if endosome transport can be understood through models used to explain particle hopping?' We will come back to this question of exploratory question-driven science, as well as the process of transferring explanatory resources from one field to another, in our discussion in Section Four.

To clarify further how integration works on a broader scale, we will turn to two other areas of investigation in which plastic research programmes appear to have created new fields: noise biology, and evolutionary systems biology. These are good examples for this stage of our discussion because they are not confined to a particular finding or research project, and they thus illustrate more broadly how integration is achieved in communities of scientists who are not necessarily connected by disciplinary training or even (at the beginning) common research aims.

3.2. Noise biology

Noise biology provides a useful general example of integration because the whole field, with its rapidly rising numbers of publications, can be seen as a direct product of innovative technologies producing new data that are used to raise previously unasked questions about biological phenomena. Noise in biological systems refers to the endogenously produced stochastic fluctuations in the molecular events that occur in cellular processes (Simpson et al., 2009; Arkin et al. 1998; Arriaga, 2009). Measurements at one level (e.g., protein abundance) may not have any observable correlation with the mRNA levels of the same gene (Taniguchi et al., 2010). These phenomena have tended to be poorly understood both because they were overwhelmed by the general concept of genomically identical cells in populations, and because in the main, single-cell fluctuations could not be observed until the advent of quantitative single-cell methods and approaches (Huang, 2009; Altschuler and Wu, 2010; Paulsson, 2004; however, see Spudich and Koshland, 1976⁴). The technologies that contributed to the creation of noise biology include flow cytometry (automated cell sorting and counting on the basis of particular characteristics) and time-lapse fluorescence microscopy (automated image acquisition of single cells over specific time periods). As these tools developed, they made visible a new realm of phenomena.

The very ability to detect and measure noise, in combination with technologies involving the synthetic construction of biological phenomena, has created a whole new field that appears to be saturated by integrative practices (Simpson et al., 2004; McCullagh et al., 2009; Eldar and Elowitz, 2010). Large quantities of new data involving single-cell measurements have to be standardized and made useable for reanalysis by other methods. Noise biology lies at the confluence of a range of methods, so much so that it could be considered the poster-child of methodological integration. It combines a variety of observational and experimental approaches with extensive mathematical modelling. These syntheses may lead to the constructions of material models manifesting various noisy behaviours. The production of new data and analyses requires new explanatory resources, which were partly found in engineering fields (Elowitz et

⁴ As early as 1976 John Spudich and Daniel Koshland undertook an investigation of 'non-genetic individuality' in genetically identical cells, but it took until the early 2000s for such phenomena to be investigated systematically.

al., 2002). The integration of mathematical accounts of noise and experimental exploration of the processes involved has produced a massive expansion of explanations of noisy phenomena (Raj and van Oudenaarden, 2008; Kaern et al., 2006).

In synthetic biology noise comes into its own as a subject of study. Because its existence can cause synthesized devices to fail or — even worse — develop unpredicted emergent properties, numerous studies in synthetic biology have deliberately investigated noisy phenomena and categorized types of noise according to its sources and effects (Swain et al., 2002; Elowitz et al., 2002; Blake et al., 2003). They have modelled, controlled, and provoked noise in order to learn more about it (e.g., Rosenfeld et al., 2005; Guido et al., 2006; Warmflash and Dinner, 2008; Munsky et al., 2009). Synthetic devices, such as genetic switches and oscillators, have been built on the basis of noisy processes (e.g., Hasty et al. 2000; Becksei and Serrano, 2001). Engineering approaches have thus been combined with microscopy and flow cytometry to devise even better ways of measuring noise and its effects, and have thereby produced integrated and novel explanations (Locke and Elowitz, 2009). As a consequence, mechanistic understanding is now emerging of how noise affects or even effects developmental processes and robustness to environmental perturbation (Elgar et al., 2007; Çağatay et al., 2009; Hasty et al., 2002; Loettgers, 2009; Love, 2009).

Noise biology generates insight *because* it integrates modelling practices (aimed at understanding the stochastic behaviour of molecules in dynamic sub-cellular systems) with experimental biology, and with new observational technologies that can take high-throughput and high-resolution measurements. When these methods are synthesized in synthetic biological constructs, another level of methodological and explanatory integration is achieved, when material models are combined with mathematical models (Loettgers and Knuutila, 2010). This is why integration — as seen in noise biology — is not the same as triangulation, which is when different approaches are used to confirm the findings of a previous line of inquiry. To be regarded as integration, methods embodying different epistemic angles and data from other domains need to be brought together. In the process of doing so, new explanatory syntheses may be generated that change how biologists think about cellular and other processes. In noise biology, these include revelations about the biological function of noise (both physiologically and evolutionarily), and even very basic assumptions about biochemical accuracy and the suppressibility of noise (Letsas et al., 2010). Simultaneously, new knowledge about tools and how they can be used has been developed alongside the creation of niches for technological innovation, and these developments have pushed the field into associated areas of inquiry. We will explore this issue of technology development further in Section Four. It is important to note that the future is wide open for the new field of noise biology, as it generates further questions about which processes are truly deterministic and only appear stochastic, and even whether noise as a biological phenomenon can

be 'translated' into clinical and industrial applications (McCullagh et al., 2009; Zernicka-Goetz and Huang, 2010).

3.3. Evolutionary systems biology

Evolutionary systems biology has taken some time to develop out of standard systems biology. In a 2007 book providing a wide-ranging introduction to philosophical issues in systems biology, the editors pointed out that 'systems biology tries to understand life as it is now, ... [and] does not yet aim at explaining the evolution of biological systems' (Boogerd et al. 2007, p. 325). The early wave of systems biology took evolution for granted as the background to efforts aimed at the understanding of particular cell functions. Distinguishing evolutionary analyses from functional ones has often been done on the basis of Ernst Mayr's (1961) distinction between proximate and ultimate explanations. According to this division of explanatory labour, systems biology focuses on proximate explanation, requiring only explanations of proximate processes such as physiology rather than ultimate evolutionary explanations. Recently, however, evolutionary explanations have been integrated into systems biological explanations of cellular processes. This has usually been achieved by modelling approaches that are combined with experimental work in order to synthesize a range of molecular datasets. These efforts have been carried out not to contradict any division between functional and evolutionary inquiry. They have been conducted because of the realization that the way evolutionary processes are explained in regard to organisms and genes can be productively integrated into the fields that explain mechanistically and mathematically molecular processes such as cell signalling and network interactions.

A useful illustration of evolutionary systems biology⁵ (ESB) comes from Mark Isalan's group in Barcelona. In Isalan's project of rewiring gene networks in bacteria, he showed different experimental strategies can be combined to generate radically novel insights into network behaviour, network evolution, and even how to do network biology (i.e., by diagnosing standard 'arrow' diagrams as inadequate representations and by showing that synthetic constructs with modified circuitry are deep sources of new knowledge). Rather than being driven by a specific hypothesis, Isalan's work began with a broad exploratory question of 'what's going on here, and what happens when we construct things differently?' (Isalan et al., 2008). He and his group mass-produced new network connections in *E. coli* in order to see what would happen to the organisms with the rewired circuitry. As well as testing the viability of a large number of new circuits (almost 600), the study was able to shed light on some of the key system-level properties. To everyone's surprise, the cells displayed high robustness,

⁵ As noted earlier in the text (e.g., Section 2) we are talking about systems biology as the investigation of dynamic large-scale systems of molecules, rather than Systems Biology, which invariably entails mathematical modelling. The Isalan group's research might also usefully be conceived as 'high-throughput synthetic biology'.

tolerating almost all these 'unnatural' circuits, even when the circuits involved heavily connected hub genes. On the other hand, modularity, a key concept in systems and synthetic biology, was found to be a somewhat misleading heuristic because of how small 'modules' could be so greatly affected by the rest of the network context.

In the process of answering these exploratory questions, the research project employed a mixture of methods, the data from which were tightly integrated via a dynamic evolutionary-systems explanation involving robustness. Circuit design (using a DNA shuffling process to explore the 'design space' of possible promoter and transcription factor combinations), growth analysis, high-throughput expression assays and selection experiments were combined in numerous rounds of inquiry in order to gain insight into the dynamics of networks and how gene change is tolerated, suppressed and, occasionally, taken advantage of. These studies were able not only to generate data about systematically redesigned genetic circuits, but also to challenge conventional mechanistic explanations of the evolving networks and organisms (Isalan et al., 2008; Bennett and Hasty, 2008; Bashor et al., 2010). But as Isalan and a colleague say in another paper, 'the interplay of modelling and experiments takes the research much further than either approach on its own' (Michalodimitrakis and Isalan, 2008, p. 28). Their work remained more strictly experimental, however, and was not combined with mathematical analyses. Much ESB does integrate mathematical approaches with experimentation, and in the process, produces a more expansive explanatory integration.

A major strand of ESB explores questions of how different evolutionary factors (e.g., antagonistic co-evolution, fluctuating environments, population dynamics) have affected network structure and dynamics, and why certain alternative structures have not evolved and yet are theoretically viable. Researchers in this area hope to realize biological design principles (basic organizational templates and functional characteristics of living processes) that will be valuable for an engineering perspective. However, being able to do so depends on integrating not only data and methods but also explanatory strategies in order to produce novel insights into evolution at the molecular level. How ESB does this is by transferring explanations and methods from one domain of inquiry to another domain, with the overall outcome being the production of a new integrated field.

Evolutionary biology is usually based on data compiled at the organismal and gene level. In ESB, the explanations and some of the methods of evolutionary biology are transferred into the domain of phenomena such as signalling networks and genetic circuitry. While this transfer and subsequent integration can produce evolutionary explanations of networks that substantiate standard selectionist accounts (e.g., Peter and Davidson, 2011), it can also take the inquiry into less conventional theoretical territory and produce surprising findings. One of the novel insights that is gaining considerable ground in ESB is the notion of non-adaptive complexity. Rather than presuming that complex biological

structures are selected for, this line of inquiry asks: 'what if neutral evolution produced these complex structures?' (Sancar, 2008; Gray et al., 2010; Lynch, 2007; Goldstein and Soyer, 2008). One insight produced by transferring neutralist evolutionary hypotheses into ESB is the finding that complexity-decreasing mutations are more deleterious than those that increase the complexity of networks (Soyer and Bonhoeffer, 2006). Networks are driven to high levels of complexity not because this is selected functionally but because of the fact that it is less difficult to go one way (in the direction of increasing complexity) than the other (decreasing complexity).

The very act of integrating evolutionary knowledge and practice within the realm of molecular systems biology has led to a new regime of inquiry being formed. Transfers of tools, methods and explanations from one domain of inquiry to another are important means of producing integration, and we think the transferability of such systems of research is likely to be driving molecular systems biology into new perspectives on biological processes as well as the creation of new fields and new accounts of scientific methodology. Wider factors, such as transferability, are linked to other contributions to integration, such as exploratory questioning and technological innovation. Gaining a bigger picture of integration requires at least a brief overview of this context.

4. Integration in context

Integration does not occur in isolation. As we noted in Section One, the classification and discussion of DD research was a general stimulus for the development of a call for more integrative practices in the molecular life sciences. While understanding DD practices (or more usefully, data-intensive research — see Leonelli, this issue) might help to contextualize the rise of an emphasis on integration, we see other factors that contribute to integration and how it works. In other words, we think the context of integration also includes some of its mechanisms. Amongst them are the factors that have already appeared in the examples above: exploratory questioning, technological innovation, and the transfer of existing knowledge-producing systems into a new domain of phenomena. We suggest that these factors are of more importance in understanding integration and the dynamics of contemporary molecular life sciences than talking about 'driving processes' being data gathering or hypothesis testing.

Exploratory questions have come up a few times in our examples in Section Three. They involve a general inquiry into a biological system rather than a closely focused hypothesis, although hypotheses are certainly developed in the process of making exploratory inquiries. Because exploratory questions are more open-ended than hypotheses, they tend to sustain and help integrate research activities, by inviting the construction of broader explanatory frameworks and the import and synthesis of appropriate methods, data and tools. One way to

conceive of how such broad questioning contributes to molecular systems biology is to think of it as central to the production of what could be called 'integrative systems'. The more familiar term for philosophers and historians of science is 'experimental systems', which is Hans-Jörg Rheinberger's term for describing 'systems of manipulation designed to give unknown answers to questions that experimentalists are not yet clearly able to ask' (1997a, p. 28). These systems of inquiry are 'vehicles for materializing questions', and do not have pre-defined boundaries or endpoints (Rheinberger, 1997a, p. 28). However, the term 'experimental systems' tends to conjure up for most scientists assumptions of specific hypotheses or series of hypotheses, and the associated connotation that such inquiries have limited freedom to move between different lines of inquiry. In order to describe the openness and question-driven nature of such inquiry, we suggest that these systems of practice, knowledge and inquiry could be called *integrative systems*.

An integrative system is one that combines a range of approaches for the purpose of exploring a general question about a biological system. Integration occurs within such knowledge-producing systems because the combination of different methods, data and explanations enables the multi-dimensional exploration of the overarching question of the inquiry. As this question generates answers and more precise hypotheses, the integrative system may become more tightly fused, such that particular methods and datasets are prescribed, along with specific explanatory frameworks. Despite the fact such inquiries are stimulated and structured by general exploratory questions, we do not think integration is a characteristic solely of exploratory (rather than more strictly HD or DD) work. Our broader concept of integrative system encompasses not only methods, data and explanations but also extensive data-gathering efforts as well as precise hypothesis testing.

Deeply embedded in such integrative systems of inquiry are the technologies that constitute methods and enable data gathering and analysis. It is obvious to many commentators that science is at least as technology-driven (TD) as it is HD or DD. Biochemist William Bains, for example, has argued that technology is the force behind many breakthroughs in the life sciences (Bains, 2008). 'Technology speculates that something can be done, and then tries it', he says (2008, p. 178), and such speculations (assumed to be false until shown otherwise, thus inverting the standard logic of hypothesis-driven science) lie at the heart of the success of much of today's systems biology (and is a notable characteristic of synthetic biology). Even if scientists talk about TD science, they are perfectly clear that scientific practice is driven and structured by many other considerations. However, we raise the issue of technology because of how deeply integrated it is in everything we have had to say about methodological, data and explanatory integration. A number of historians and philosophers of science have investigated the role of scientific technology and instrumentation in great detail (e.g., Galison, 1997; Pitt, 2000; Ihde, 2004). From their perspectives and ours, technology does not just function as a tool but — as our examples show — can also be

constitutive of the research question, as in noise biology. And as technologies deployed in one field are brought into other fields, a variety of other resources are transferred and transformed along with them, including conceptual frameworks and data preferences.

We have mentioned a few times already the importance of transferability, in which aspects of one research programme can be brought into another. It is potentially useful to think of this process as *translation*, which is often discussed as the movement of scientific findings from laboratories to the clinic or industry. It is standard to presume that research is organized differently in these applied settings, which are distinguished from 'basic' science laboratories (Petsko, 2010). We think that it might be much more useful to consider translation in light of the transfer and integration of methods, data and explanations, in the way we have shown to be occurring in, for example, the development of ESB. We view translation not simply as a matter of knowledge moving from context A to context B, but of systems of inquiry making transformative shifts between research domains. Rheinberger calls this, in the context of experimental systems, 'hybridization' (1997b, p. S250), but we think it can be usefully understood as the integration of aspects of already developed systems of knowledge production within new research areas. Such horizontal shifts can go in many directions for many different reasons, but are unlikely to be captured by descriptions that emphasize simple transitions 'from basic to applied'. In fact, translation can be understood as the ultimate interpretation of integration because it addresses the question of how basic science and application are integrated in more complex ways than presently modelled.

The general context of integration is thus multifaceted and multilevel. Our examples, all success stories because of the way they have transformed understandings of particular biological systems, do not mean that the problems of achieving integration are not clearly and repeatedly recognized. Most systems biologists realize there is a long way to go in order to achieve truly effective integration in a systematic way. Included amongst the big issues that are the subjects of ongoing discussion (some of it at the meeting mentioned in Section 2) are numerous questions about how to implement integration more effectively. From a practical point of view, data integration is the most pressing challenge. Because experiments, for example, are not easily standardized (the biological system can vary, as can procedures and data analysis), they can produce very specific results that are hard to model mathematically since the aim of doing so is to produce something more generalizable (Schilling et al. 2008). Connecting different levels and scales of data is still difficult, and failure to achieve it can inhibit methodological and explanatory integration.

Another potential inhibitor of integration, and particularly explanatory integration, is disciplinarity. Disciplinary integration has been a topic of philosophical discussion (Burian, 1993; van der Steen, 1993). We have not given it the same attention in our analysis of integration as we did methods, data and explanation

because we see disciplines as forming some of the context of integration. Although shorthand accounts of systems biology often talk about integrating disciplines, this usually boils down to the three sets of integrative practices we outlined in Section Three. Multidisciplinary capacities certainly inform and even guide integration, but are *conditions* for integration rather than integration itself. If approaches and researchers are too far apart in terms of perspective, integration is difficult to achieve in any meaningful way. But concomitantly, if approaches and backgrounds are too close together, then what is achieved may not address the biological system or research question in a way that takes advantage of large bodies of data, capacities to combine methods and the drive for multilevel predictive explanations. Integration is often, therefore, presented normatively, as something that *should* be done because of the aim to generate dynamic insight into interactive biological systems. The requisite multidisciplinary is sometimes discussed in terms of having a full systems-biological approach exhibited in one person, so that multiple approaches can be carried out by that one researcher (Calvert and Fujimura, 2010). Increased access to a variety of technologies and datasets allows scientists to perform not just hypothesis- and data-driven research within the same lab but also to combine both approaches in the same project. But for many biologists, whether their background is mathematical or experimental, optimal individuals are those who have trained in one discipline and learned to work with other disciplines. Mono-disciplinary training is still considered by many scientists to be the best way to avoid the production of undertrained but multidisciplinary researchers (Eddy, 2005; Berg, 2005). So far, no empirical studies are available on the efficacy of either perspective in regard to molecular systems biology, and both policy and practice are based on 'gut feeling' and anecdotal evidence.

Each one of these contextual issues raises major epistemic and methodological questions. The very fact they are being discussed indicates to us that systems biology is open to different ways of thinking about science and how it can be done. Philosophers of science can benefit from such discussions and from analysing the abundance of examples of integration in molecular systems biology. Whether or not this will give rise to a normative philosophy of the molecular life sciences is a different question, of course. Urging specific combinations of approaches for their own sake, as part of a formula of how science should be practised, do not seem to be indicated by our examples. But it is clear that integration is being treated normatively, as something molecular systems biologists should do if genuinely system-oriented biology is to be accomplished. In this sense, integration is contributing to the reconfiguration of accounts of how science works.

5. Conclusions

Integration, as depicted in our examination of molecular systems biology, has become a multidimensional activity that is qualitatively different from earlier

discussions of integration (whether discussed philosophically or scientifically). Large-scale datasets, high-throughput and high-resolution methods and technologies, the consolidation of broad community-based efforts to predict and control biological processes, and institutional encouragement (including that of funding bodies) have all made integration a contemporary focus of practice. It has become something that happens not just occasionally or accidentally, but an activity that must be aimed at systematically and achieved consistently in molecular systems biology. While we have emphasized examples in which integration has underpinned the creation of novel insight into biological systems and exploratory research questions, integration is also the meta-methodological thread connecting many different practices in molecular systems biology. In this respect, integration can be conceived as the basis for a more developed account of scientific practice, in which strict hypothesis testing or 'mere' data gathering are framed as aspects of an integrating dynamic of scientific practice.

Examining systems biology from a pragmatic but philosophical point of view suggests that we need a more multi-dimensional account of different kinds of practice and how these are synthesized to produce new biological knowledge and capabilities. Philosophy of science has traditionally focused more on theory, but the new wave of philosophical approaches to scientific practice has paid much more attention to experiment, modelling and simulation (e.g., Galison, 1988; Lenoir, 1988; Weber, 2005; Creager et al., 2007; Winsberg, 2009). In contemporary molecular life sciences, it now seems something even broader is needed — not just studies of experiments in addition to theories, but of the very ways in which knowledge-making is approached, justified and revised. Examination of integration is one tack to take in achieving a down-to-earth philosophical view of contemporary biology. The important point we take from our analysis of molecular systems biology is that new ways of thinking about scientific practice are emerging, and that discussions about these developments are likely to provide valuable opportunities for philosophers of science to examine their own conceptions of science and even philosophy of science itself.

6. Acknowledgements

We thank our referees and editor for the exceedingly helpful comments that led to major revisions of our initial draft. MAO's research is funded by the Australian Research Council in the form of a Future Fellowship held at the University of Sydney. Her initial research for this paper was done with funding from the Economic and Social Research Council through the programme of research at Egenis, at the University of Exeter. OSS acknowledges the support of University of Exeter Science Strategy.

7. References

- Aderem, A. (2005). Systems biology: Its practice and challenges. *Cell*, 121, 511-513.
- Aebersold, R., Hood, L. E., & Watts, J.D. (2000). Equipping scientists for the new biology. *Nature Biotechnology*, 18, 359.
- Allen, J. F. (2001). Bioinformatics and discovery: induction beckons again. *BioEssays*, 23, 104–107.
- Altschuler, S. J., and Wu, L. F. (2010). Cellular heterogeneity: Do differences make a difference? *Cell*, 141, 559-563.
- Arkin, A., Ross, J., and McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149, 1633-1648.
- Arkin, A. P., and Schaffer, D. V. (2011). Network news: Innovations in 21st century systems biology. *Cell*, 144, 844-849.
- Arriaga, E. A. (2009). Determining biological noise via single cell analysis. *Analytical and Bioanalytical Chemistry*, 393, 73-80.
- Ashwin, P., Lin, C., and Steinberg, G. (2010). Queueing induced by bidirectional motor motion near the end of a microtubule. *Physical Review*, 82, 051907.
- Auffray, C., Imbeaud, S., Roux-Rouquié, M., & Hood, L. (2003). From functional genomics to systems biology: Concepts and practices. *Comptes Rendus Biologies*, 326, 879-892.
- Bains, W. (2008). Technological speculations and science. *Bioscience Hypotheses*, 1, 177-178.
- Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. (2010). Rewiring cells: Synthetic biology as a tool to interrogate the organizational principles of living systems. *Annual Review of Biophysics*, 39, 515-537.
- Bechtel, W. (1993). Integrating sciences by creating new disciplines: The case of cell biology. *Biology and Philosophy*, 8, 277-299.
- Becskei, A., Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405, 590-593.
- Bennett, M.R., and Hasty, J. (2008). Genome rewired. *Nature*, 452, 824-825.

Berg, H. (2005). Q & A. *Current Biology*, 15, R189-R190.

Blake, W. J., and Isaacs, F. J. (2004). Synthetic biology evolves. *Trends in Biotechnology*, 22, 321-324.

Blake, W. J., Kaern, M., Cantor, C. R., Collins, J. J. (2003) Noise in eukaryotic gene expression. *Nature*, 422, 633-637.

Boehm, B. W. (1988). A spiral model of software development and enhancement. *Computer*, 21, 61-72.

Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J.-H., and Westerhoff, H. V. (Eds.), *Systems biology: Philosophical foundations*. Amsterdam: Elsevier.

Brent, R. (2000). Genomic biology. *Cell*, 100,169-83.

Brigandt, I. (2010). Beyond reductionism and pluralism: Toward an epistemology of explanatory integration. *Erkenntnis*, 73, 295-311.

Brown, A. J. P. (2010). Microbial signalling and systems biology. *Genome Biology*, 11, 302, <http://genomebiology.com/2010/11/5/302>

Bruggeman, F. J., & Westerhoff, H. V. (2006). The nature of systems biology. *Trends in Microbiology*, 15, 45-50.

Burian, R. M. (1993). Unification and coherence as methodological objectives in the biological sciences. *Biology and Philosophy*, 8, 301-308.

Çağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J., and Süel, G. M. (2009) Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139, 512-522.

Cali, A., Calvanese, D., de Giacomo, G., and Lenzerini, M. (2003). Data integration under integrity constraints. *Information Systems*, 29, 147-163.

Calvanese, D., and de Giacomo, G. (2005). Data integration: A logic-based perspective. *AI Magazine*, 26, 59-70.

Calvert, J., and 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.

Cary, M. P., Bader, G. D., and Sander, C. (2005). Pathway information for systems biology. *FEBS Letters*, S79, 1815-1820.

O'Malley MA, Soyer OS (forthcoming, 2011). The roles of integration in molecular systems biology. *Studies in History and Philosophy of the Biological and Biomedical Sciences*.

Cat, J. (2007). The unity of science. *Stanford Encyclopedia of Philosophy*.
<http://plato.stanford.edu/entries/scientific-unity/>

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.

Chaudhury, D., Schadschneider, A., and Nishinari, K. (2005). Physics of transport and traffic phenomena in biology: From molecular motors and cells to organisms. *Physics of Life Reviews*, 2, doi: 10.1016/j.plrev.2005.09.001

Chen, W. W., Niepel, M., and Sorger, P. K. (2010). Classic and contemporary approaches to modeling biochemical reactions. *Genes and Development*, 24, 1861-1875.

Creager, A. N. H., Lunbeck, E., & Wise, M. N. (2007). *Science without laws: Model systems, cases, exemplary narratives*. Durham NC: Duke University Press.

Darden, L. (1991). *Theory change in science*. New York: Oxford University Press.

Darden, L., and Maull, N. L. (1977). Interfield theories. *Philosophy of Science*, 44, 43-64.

Eddy, S. R. (2005). 'Antedisciplinary' science. *PLoS Computational Biology*, 1 (1), e6. doi:10.1371/journal.pcbi.0010006

Eldar, A., and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, 467, 167-173.

Elliott, K. (2004). Error as means to discovery. *Philosophy of Science*, 71, 1-24.

Elowitz, M.B., Levine, A.J., Siggia, E. D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science*, 297, 1183-1186.

Endy, D. (2005). Foundations for engineering biology. *Nature*, 438, 449-453.

Evans, J., and Rzhetsky, A. (2010). Machine science. *Science*, 329, 399.

Finkelstein, A., Hetherington, J., Li, L., Margoninski, O., Saffrey, P., Seymour, R., and Warner, A. (2004). Computational challenges of systems biology. *Computer*, 37 (4), 26-33.

Fox, P., and Hendler, J. (2011). Changing the equation on scientific data visualization. *Science*, 331, 705-708.

Galison, P. (1988). History, philosophy, and the central metaphor. *Science in Context*, 2, 197-212.

Galison, P. (1997). *Image and logic: A material culture of microphysics*. Chicago: University of Chicago Press.

Ge, H., Walhout, A. M. M., and Vidal, M. (2003). Integrating 'omic' information: A bridge between genomics and systems biology. *Trends in Genetics*, 19, 551-560.

Goldstein, R. A., and Soyer, O. S. (2008). Evolution of taxis responses in virtual bacteria: Non-adaptive dynamics. *PLoS Computational Biology*, 5 (4), e10000084. Doi:10.1371/journal.pcbi.10000084

Good, R. J. (1999). Why are chemists turned off by philosophy of science? *Foundations of Chemistry*, 1, 65-95.

Grantham, T. A. (2004). Conceptualizing the (dis)unity of science. *Philosophy of Science*, 71, 133-155.

Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., and Doolittle, W. F. (2010). Irremediable complexity? *Science*, 330, 920-921.

Gudelj, I., Weitz, J. S., Ferenci, T., Horner-Devine, M. C., Marx, C. J., Meyer, J. R., and Forde, S. E. (2010). An integrative approach to understanding microbial diversity: From intracellular mechanisms to community structure. *Ecology Letters*, 13, 1073–1084.

Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science*, 296, 1466-1470.

Gunawardena, J. (2010). Biological systems theory. *Science*, 328, 581-582.

Hasty, J., McMillen, D., and Collins, J. J. (2002). Engineered gene circuits. *Nature*, 420, 224-230.

Hasty, J., Pradines, J., Dolnik, M., and Collins, J. J. (2000). Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences USA*, 97, 2075-2080.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: More than just noise. *Development*, 136, 3853-3862.

Hwang, D., Rust, A. G., Ramsey, S., Smith, J. J., Leslie, D. M., Weston, A. D., de Atauri, P., Aitchison, J. D., Hood, L., Siegel, A. F., and Bolouri, H. (2005). A data integration methodology for systems biology. *Proceedings of the National Academy of Sciences USA*, 48, 17296-17301.

Ideker, T., Bafna, V., and Lemberger, T. (2007). Integrating scientific cultures. *Molecular Systems Biology*, 3, 105: doi: 10.1038/msb4100145

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics*, 2, 343-372.

Ihde, D. (2004). Has the philosophy of technology arrived? A state-of-the-art review. *Philosophy of Science*, 71, 117-131.

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., and Serrano, L. (2008). Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, 452, 840-845.

Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6, 451-464.

Kell, D. B., and Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays*, 26, 99-105.

King, R. D., Liakata, M., Lu, C., Oliver, S. G., and Soldatova, L. N. (2011). On the formalization and reuse of scientific research. *Journal of the Royal Society Interface*, doi: 10.1098/rsif.2011.0029

Kirkwood, T. B. L., Boys, R. J., Gillespie, C. S., Proctor, C. J., Shanley, D. P., and Wilkinson, D. J. (2003). Towards an e-biology of ageing: Integrating theory and data. *Nature Reviews Molecular Cell Biology*, 4, 243-249.

Kitano, H. (2002). Computational systems biology. *Nature*, 420, 206-210.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48, 507-531.

Knuttila, T. T., and Loettgers, A. (2010). Modeling and experimenting: The combinatorial strategy in synthetic biology. Presented at Philosophy of Scientific Experimentation: A Challenge to Philosophy of Science (October 15-17). Pittsburgh: Center for Philosophy of Science. Draft text available at PhilSci Archive, <http://philsci-archive.pitt.edu/>

Lenoir, T. (1988). Practice, reason, context: The dialogue between theory and experiment. *Science in Context*, 2, 3-22.

Lerman, J., and Palsson, B. O. (2010). Topping off a multiscale balancing act. *Science*, 330, 1058-1059.

Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the twenty-first ACM SIGACT-SIGMOD-SIGART symposium on principles of database systems* (pp. 233-246). NY: ACM.

Leonelli, S. (2008). Bio-ontologies as tools for integration in biology. *Biological Theory*, 3 (1), 7-11.

Lestas, I., Vinnicombe, G., and Paulsson, J. (2010). Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467, 174-178.

Locke, J. C. W., and Elowitz, M. B. (2009). Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology*, 7, 383-392.

Loettgers, A. (2009). Synthetic biology and the emergence of a dual meaning of noise. *Biological Theory*, 4 (4), 340-356.

Love, A.C. (2009). Investigating biological information with noise. Presentation at the *Biennial Meeting of the International Society for the History, Philosophy, and Social Studies of Biology*, Brisbane (July). www.ishpssb2009.org/program.html

Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*, 8, 803-813.

Mäki, U. (2001). Explanatory unification: Double and doubtful. *Philosophy of the Social Sciences*, 31, 488-506.

Maull, N. L. (1977). Unifying science without reduction. *Studies in the History and Philosophy of the Life Sciences*, 8, 143-162.

Mayr, E. (1961). Cause and effect in biology. *Science*, 134, 1501-1506.

McCullagah, E., Farlow, C., Girard, J., Lipinski-Kruszka, J., Lu, D., Noriega, T., Rollins, G., Spitzer, R., Todhunter, M., and El-Samad, H. (2009). Not all quiet on the noise front. *Nature Chemical Biology*, 5, 699-704.

Michalodimitrakis, K., and Isalan, M. (2008). Engineering prokaryotic gene circuits. *FEMS Microbiology Reviews*, 33, 27-37.

Mitchell, S. D. (2004). Why integrative pluralism? *E:CO*, 6, 81-91.

Munsky, B., Trinh, B., and Khammas, M. (2009). Listening to the noise: Random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5, 318 doi:10.1038/msb.2009.75

Nagel, K. (1996). Particle hopping models and traffic flow theory. *Physical Review E*, 53, 4655-4672.

Ng, A., Bursteinas, B., Gao, Q., Mollison, E., and Avelebil, M. (2006). Resources from integrative systems biology: From data through databases to networks and dynamic systems models. *Briefings in Bioinformatics*, 7, 318-330.

Nickles, T. (1997). A multi-pass conception of scientific inquiry. *Danish Yearbook of Philosophy*, 32, 11-44.

Nielsen, J. (1993). Iterative user- interface design. *Computer*, 26, 32-41.

O'Malley, M. A., Elliott, K. C., Haufe, C., & Burian, R. M. (2009). Philosophies of funding. *Cell*, 138, 611-615.

Oppenheim, P., and Putnam, H. (1958). Unity of science as a working hypothesis. In Feigl, H., Scriven, M., and Maxwell, G. (Eds.), *Minnesota studies in philosophy of science*, Vol. 2 (pp. 3-36). Minneapolis: Minnesota University Press.

Padian, K. (2001). Cross-testing adaptive hypotheses: Phylogenetic analysis and the origin of bird flight. *American Zoologist*, 41, 598-607.

Palsson, B. Ø. (2006). *Systems biology: Properties of reconstructed networks*. Cambridge: Cambridge University Press.

Patel, M., and Nagi, S. (2010). *The role of model integration in complex systems modelling: An example from cancer biology*. Berlin: Springer.

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, 427, 415-418.

Peter, I. S., and Davidson, E. H. (2011). Evolution of gene regulatory networks: Controlling body plan development. *Cell*, 144, 970-985.

Petsko, G. (2010). Lost in translation. *Genome Biology*, 11:107. doi:10.1186/gb-2010-11-2-107

Philippi, S., and Köhler, J. (2006). Addressing the problems with life-science databases for traditional users and systems biology. *Nature Reviews Genetics*, 7, 482-488.

Pitt, J. C. (2000). *Thinking about technology: Foundations of the philosophy of technology*. NY: Seven Bridges.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135, 216-226.

Rheinberger, H.-J. (1997a). *Toward a history of epistemic things: Synthesizing proteins in the test tube*. Stanford, CA: Stanford University Press.

Rheinberger, H.-J. (1997b). Experimental complexity in biology: Some historical and epistemological remarks. *Philosophy of Science*, 64 (Supplement), S245-S254.

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science*, 307, 1962-1965.

Sancar, A. (2008). The intelligent clock and the Rube Goldberg clock. *Nature Structural & Molecular Biology*, 15, 23-24.

Schilling, M., Pfeifer, A. C., Bohl, S., and Klingmüller, U. (2008). Standardizing experimental protocols. *Current Opinion in Biotechnology*, 19, 354-359.

Schmid, A., and Blank, L. M. (2010). Hypothesis-driven omics integration. *Nature Chemical Biology*, 6, 485-487.

Schuster, M., Kilaru, S., Ashwin, P., Lin, C., Severs, N. J., Steinberg, G. (2011). Controlled and stochastic retention concentrates dynein at microtubule ends to keep endosomes on track. *EMBO Journal*, 30, 652-664.

Simpson, M. L., Cox, C. D., Allen, M. S., McCollum, J. M., Dar, R. D., Karig, D. K., and Cooke, J. F. (2009). Noise in biological circuits. *WIREs Nanomedicine and Nanobiotechnology*, 1, 214-225.

Smolke, C. D., and Silver, P. A. (2011). Informing biological design by integration of systems and synthetic biology. *Cell*, 144, 855-859.

Soyer, O.S., and Bonhoeffer, S. (2006). Evolution of complexity in signaling pathways. *Proceedings of the National Academy of Sciences USA*, 103, 16337-16342.

Spudich, J. L., and Koshland, D. E. Jr. (1976). Non-genetic individuality: Chance in the single cell. *Nature*, 262, 467-471.

Star, S. L. (1986). Triangulating clinical and basic research: British localizationists, 1870-1906. *History of Science*, 24, 29-48.

van der Steen, W. J. (1993). Towards disciplinary disintegration in biology. *Biology and Philosophy*, 8, 259-275.

Strasser, B. J. (2008). GenBank — Natural history in the 21st century? *Science*, 322, 537-538.

O'Malley MA, Soyer OS (forthcoming, 2011). The roles of integration in molecular systems biology. *Studies in History and Philosophy of the Biological and Biomedical Sciences*.

Sullivan, D. E., Gabbard, J. L. Jr., Shukla, M., and Sobral, B. (2010). Data integration for dynamic and sustainable systems biology resources: Challenges and lessons learned. *Chemistry and Biodiversity*, 7, 1124-1141.

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences USA*, 99, 12795-800.

Taniguchi, Y., Choi, P. J., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329, 533-538.

Triplet, T., and Butler, G. (2011). Systems biology warehousing: Challenges and strategies toward effective data integration. In *DBKDA 2011: The third international conference on advances in databases, knowledge, and data applications* (pp. 34-40). St. Maarten, The Netherlands Antilles: IARIA.

Understanding Science. (2010). Reviewing test results.
undsci.berkeley.edu/article/0_0_0/howscienceworks_10

Vidal, M., Cusick, M. E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell*, 144, 986-998.

Warmflash, A., and Dinner, A. R. (2008). Signatures of combinatorial regulation in intrinsic biological noise. *Proceedings of the National Academy of Sciences USA*, 105, 17262-17267.

Weber, M. (2005). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.

Westerhoff, H. V., and Kell, D. B. (2007). The methodologies of systems biology. In Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J.-H., and Westerhoff, H. V. (Eds.), *Systems biology: Philosophical foundations* (pp. 23-70). Amsterdam: Elsevier.

Weston, A. D., and Hood, L. 2004. Systems biology, proteomics, and the future of health care: Toward predictive, preventative, and personalized medicine. *Journal of Proteome Research*, 3, 179-196.

White, M. (2010). Systems biology has become meaningless. Science 2.0, http://www.science20.com/adaptive_complexity/systems_biology_has_become_meaningless (accessed November 7th, 2010).

Wilson, E. O. (1998). Consilience among the great branches of learning. *Daedalus*, 127 (1), 131-149.

O'Malley MA, Soyer OS (forthcoming, 2011). The roles of integration in molecular systems biology. *Studies in History and Philosophy of the Biological and Biomedical Sciences*.

Wimsatt, W. (2007a [1981]). Robustness, reliability, and overdetermination. In Wimsatt, W., *Re-engineering philosophy for limited beings: Piecewise approximations to reality* (pp. 43-74). Cambridge, MA: Harvard University Press.

Wimsatt, W. (2007b [1987]). False models as means to truer theories. In Wimsatt, W., *Re-engineering philosophy for limited beings: Piecewise approximations to reality* (pp. 94-132). Cambridge, MA: Harvard University Press.

Winsberg, E. (2009). Computer simulation and the philosophy of science. *Philosophy Compass*, 4, 835-845.

Zernicka-Goetz, M., and Huang, S. (2010). Stochasticity versus determinism in development: A false dichotomy? *Nature Reviews Genetics*, 11, 743-744.