Molecules, Cells and Minds

Aspects of Bioscientific Explanation

Submitted by Alexander Powell, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Philosophy, December 2009.

I certify that all material which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Alexander Powell

.....

Acknowledgements

The work described in this thesis in part builds on earlier experience, and I should like first to acknowledge several long-standing debts. One is to Hilary Muirhead, who was notably generous with her time when I was undertaking my final year biochemistry project at Bristol. At Oxford I benefited greatly from (even if I could have done rather more to deserve) the support of Janos Hajdu and Chris Dobson, while Andrew Martin and Dan Raleigh were munificent with their time and expertise.

As for the research reported here, a key enabling condition was John Dupré's willingness to take me on as a research student. I am extremely grateful to him for that, and for the patient and subtle way in which he facilitated my discovery of the limitations of various of my philosophical thoughts and assumptions. Warm thanks go to Cheryl Sutton for performing administrative magic in order to make things feasible. Sabina Leonelli organized the viva with characteristic brio, and I am grateful to her for much besides her inspiring intellectual breadth and generosity. Maureen O'Malley provided encouragement and vital critical opinion on several chapter drafts, as well as on my various other productions over the past four years, for which a big thank you. Jane Calvert and Staffan Mueller-Wille did much to make my time at Egenis rewarding. I am especially grateful to Jane for several enjoyable collaborations and for her moral support at various times, while Steve Hughes was an assiduous mentor and a reassuring presence.

Beyond that the debts start to proliferate to such a degree that it becomes tempting to resort to catch-all expressions of gratitude. For their company and conversation, however, I cannot resist singling out Daniele Carrieri, Jonathan Davies (who was good enough to read a late draft of the entire thesis), Mattia Gallotti, Katie Kendig (who steered me away from several philosophical tar pits), Maren Klotz, Michiru Nagatsu and Sally Wasmuth. The kindnesses of my fellow 2005-ers Kate Getliffe and Jean Harrington have been legion, and frankly they deserve medals for putting up with my puns and musings. Doctoral life would have been much the worse without them. Special thanks also go to Pierre-Olivier Méthot. In addition to making many pertinent suggestions regarding several chapters, his general soundness and joie de vivre helped to make the write-up stage considerably less burdensome than it otherwise would have been. Exploring obscure corners of Devon with him and the Bootlegs was a highlight of 2009. Another important stimulus has been the splendid hospitality of Tom Gaertner and Nadine Schaefer. When on occasion I felt delighted enough by philosophical reflection the excellence of their company and their cuisine were very reliable restoratives.

Implicit in this thesis is the idea that words can only get one so far. When it comes to family, where the debts are deepest, they become especially inadequate. (Although it is straightforwardly the case that Jo Anson did wonders for the bibliography.) Suffice it to say that without the steadfast support and encouragement of Jo, Mike, and my parents Pam and Chris none of this would have been even vaguely possible. This thesis is therefore dedicated to them and to Sarah, with love and thanks.

Molecules, Cells and Minds: Aspects of Bioscientific Explanation

Abstract

In this thesis I examine a number of topics that bear on explanation and understanding in molecular and cell biology, in order to shed new light on explanatory practice in those areas and to find novel angles from which to approach relevant philosophical debates. The topics I look at include mechanism, emergence, cellular complexity, and the informational role of the genome. I develop a perspective that stresses the intimacy of the relations between ontology and epistemology. Whether a phenomenon looks mechanistic, or complex, or indeed emergent, is largely an epistemic matter, yet has an objective basis in features of the world.

After reviewing several concepts of mechanism I consider the influential recent account of Machamer, Darden and Craver (MDC). That account makes interesting proposals concerning the relationship between mechanistic explanation and intelligibility, which are consistent with the results of the investigation I undertake into the science surrounding protein folding. In relation to a number of other issues pertaining to biological systems I conclude that the MDC account is insufficiently nuanced, however, leading me to outline an alternative approach to mechanism. This emphasizes the importance of structure—function relations and addresses issues raised by reflection on the nature of cellular complexity. These include the distinction between structure and process and the different possible bases on which system organization may be maintained.

The account I give of emergence construes the phenomenon in terms of psychological deficit: phenomena are emergent when we lack the capacity to trace through and model their causal structures using our cognitive schemas. I conclude by developing these ideas into a preliminary and partial account of explanation and understanding. This aspires to cover the significant fraction of work in molecular and cell biology that correlates biological structures, processes and functions by visualizing phenomena and making them imaginable.

(298 words)

List of Contents

List of Figures

Author's declaration

Four sentences in Chapter 1, dealing with the Oppenheim-Putnam view of the ontological stratification and unity of science (page 13), are taken from Powell and Dupré 2009 (listed in the Bibliography).

Chapter 6 is based on the talk on 'Emergence, imaginability, and causal schemas' I delivered at the meeting of the International Society for the History, Philosophy, and Social Studies of Biology held at Exeter University in July 2007. Several passages of that chapter also appear in Powell and Dupré 2009.

"There are today only a few philosophers of science who would defend any major logical empiricist doctrines in anything like their original form. Yet the legacy lives on in assumptions held even by those most opposed to the logical empiricist account of science. One is the widespread presumption that scientific theories are essentially linguistic entities – maybe not formal axiomatic systems, but nevertheless something like sets of statements. A second is the presumption that the job of the philosophy of science is at least to 'explicate,' even if no longer to justify, the rules of rational belief and action that are supposed somehow to 'govern' scientific activity. Of course, some philosophers have challenged even these presumptions. A cognitive approach to the study of science, as I understand it, rejects both."

- Ronald Giere [1]

"[W]e are ignorant how far the mechanical mode of explanation possible for us may penetrate."

– Immanuel Kant [2]

"In general, we're least aware of what our minds do best."

– Marvin Minsky [3]

[1] Giere (1988), p.28.
[2] Kant (1790/2007), p.243
[3] Minsky (1987), p.29.

1. The logical empiricist legacy

Introduction

In this thesis I investigate the nature of scientific explanation and understanding. My principal aim, however, is not to develop a grand theory of explanation to rival or supplant those devised and debated by philosophers of science over the past half century or so. Rather I propose to investigate a number of related topics that are salient to explanation and understanding in molecular and cell biology. In this way I hope to be able shed new light on aspects of explanatory practice in these areas of biology, and reciprocally find novel ways of looking at philosophical accounts of explanation and understanding.

Patterns of explanation in molecular and cell biology depend strongly on our knowledge of, and the assumptions we make about, the character of molecular and cellular systems. Hence much of the thesis consists of reflection on ontological issues. An important aspect of my position, however, is that to develop a philosophically satisfactory picture we sometimes need to take our own psychological capacities and dispositions into account. I therefore stress the intimacy of the relations between ontology and epistemology, and the importance of epistemic considerations generally for getting to grips with a number of central philosophical concepts. One such is causation, a traditional view about which could be expressed sloganistically by saying that things really exist if they have causal powers and so have causal effects. Thus causation and ontology are regarded as tightly intertwined, and things dignified with causal powers are considered to be unproblematically real.[4] I see causation as a metaphysically loaded concept, however, and correspondingly regard the ontological distinctions we make as often being more psychologically contingent than is perhaps generally acknowledged. A view to which I am sympathetic is that we interpret and understand phenomena in the world by paralleling them cognitively using schemas and entailment structures of various kinds (Sloman 2005).[5]

[4] Hacking (1983), for example, talks about the reality of entities capable of playing instrumental causal roles in experimental interventions in phenomena.

[5] And increasingly via the epistemic prosthesis of computer simulation, as I describe in Chapter 3 in relation to protein folding.

Sometimes these are supported by very little besides raw belief.[6] A lack of fit between our schemas and events in the world impels us to develop better schemas, from which the progressive character of science results.

This overall orientation shapes, as well as reflects, the views I develop in the following chapters about a range of topics relating to biological explanation and understanding. The principle novel elements of my treatment are:

1) a perspective on mechanism in biology that cleanly separates ontological from epistemological factors and which emphasizes the importance of structure—function relations (Chapters 2, 4 and 5);
2) a review of protein folding science and discussion of some of the epistemological issues surrounding molecular dynamics simulation methods (Chapter 3);
3) a discussion of biocomplexity that highlights the 'high bandwidth' nature of cellular causality and the epistemological significance of non-equilibrium behaviour (Chapters 4 and 5);
4) a jointly ontological and epistemological account of emergence framed in terms of cognitive constraint and schema deficit (Chapter 6);
5) an attempt to understand genetics- and genomics-related information talk in terms of the distinctive character of DNA and the processes in which it participates, allied to our function-attributing tendencies (Chapter 7).

I should mention too what I do not cover. The major omission is evolution, about which I say almost nothing beyond some brief passages in Chapter 7 relating to information storage in the genome. This is because my primary interest, which I share with Boogerd et al. (2007, pp.325-326), is in how micro-scale biological systems such as cells work in a proximate sense, rather than how it is that such systems have come to exist at all. Thus much of the thesis – Chapters 2 to 5 inclusive – consists of an investigation of mechanistic concepts and their relevance to molecular and cell biological explanation and understanding.

To set the scene I first outline in the remainder of this chapter the philosophical context from which contemporary interest in those concepts has arisen. I begin by discussing the concept of reduction, which historically has played a key structuring role in

[6] For example, in England even as late as the seventeenth century the causal powers of witches were sometimes held to account for the occurrence of particular events (Thomas 1971).

philosophy of science and philosophy of biology. The rise of the latter can be viewed in part as a response to the failure of the logical empiricist programme to articulate an account of explanation capable of illuminating biological thought and practice. This was because logical empiricism focused on formal theories and their relationships, whereas in many areas of biology formal theories – qua mathematically specifiable quantitative relationships of wide applicability – are thin on the ground. In formalist views of science that emphasize theories and theoretical relationships, reduction of one theory to another represents one sense of explanation. The idea is that a theory A explains another theory B if the formal elements of theory B can be systematically related to those of theory A, and if in some set of circumstances the formal picture (in terms of the values of variables etc.) provided by theory A looks the same as that of theory B. If the range of circumstances covered by theory A is greater than that of theory B then theory A is considered more fundamental than theory B. Moreover the fact that theory B takes the form it does is explained by the way in which the formal picture offered by theory A adopts the form it does in the specific circumstances covered by theory B. Other senses of reduction are less directly connected with the concept of explanation, but nonetheless enter into discussions of explanation in a variety of ways.

From reduction I move on to more overt conceptions of explanation itself, starting with the view that explanations amount logically to arguments. Hempel's deductive-nomological (D-N) model of explanation is the best-known and most influential expression of this idea. I also look briefly at unification-based models of explanation such as those articulated by Friedman and Kitcher as a response to defects of the D-N model. What these different models of explanation share – besides seeing explanation in terms of relations amongst linguistic terms – is the deliberate avoidance of causation as a concept around which to build an account. It is this eschewal of causation – on the grounds that the concept is deemed too metaphysically laden to serve as a philosophically satisfactory basis for explicating explanation – that gives rise to difficulties. I then quickly survey a variety of accounts of causation and causal explanation, including the 'causal-mechanical' account associated with Salmon and Dowe. My treatment of all these different perspectives is highly compressed, there being insufficient space to do justice to the complexity of the relevant debates when my major focus is elsewhere. (In any case they have been covered in considerable depth by many other authors.) However, it serves as pertinent philosophical background against which to view the issues that concern me, and provides at least a partial account of how it is that concepts of mechanism (about which I shall have much to say) have been the focus of so much recent philosophical activity.

The received view and aftermath

The formalist programme promoted by the logical positivists in the first half of the twentieth century led to a view of science and the aims of philosophy of science that continues to be reflected in the shape of many philosophical debates, even if often only in the negative sense of providing a set of positions to which contemporary alternatives can be seen as a reaction. The 'received view' to which logical positivism gave rise (Suppe 1974, 2000), under the banner of logical empiricism, might be characterized as asserting that the central concern of philosophy of science is to formalize the relations between observed phenomena and scientific theories qua mathematical objects. Theories gain legitimacy to the extent that they can be shown to connect in formal terms (via 'correspondence rules') with 'observation sentences', or linguistic codifications of perceived phenomena. There are several points to note here. First, a strong distinction is drawn between a realm of observation and a realm of theory. Secondly, there is the sense in which observations are epistemically prior to theory – hence the received view might be said to be anti-realist about the objects of theory. And thirdly there is the importance attached to language and mathematical logic, allied to the implicit view that there is an intimate relationship between the two.[7]

Underlying the logical empiricist programme was the belief that formal structures representing the relationships between scientific observations and theoretical statements and generalizations could collectively provide an objective view of how science works without making reference to metaphysically problematic notions such as causation. The aim of articulating formal relations between observations and scientific theories was more congenial to the mathematical codifications of knowledge possible (indeed prevalent) in physics than to than the looser, less mathematically systematized conceptual structures found in other ('softer') disciplines. This bias towards formal theoretical structures, supposedly amenable to interpretation in terms of the positivist model of science, encouraged the idea of a disciplinary hierarchy. Physics was the foundation, and it was towards the condition of physics that all other disciplines should aspire. There was thus a normative dimension to the positivist view (Suppe 2000, S104).

[7] The difficulties that researchers in artificial intelligence have encountered in their attempts to develop computer systems capable of natural language processing and understanding might, however, argue for caution about regarding language purely as a logically driven symbolic system.

In order to extend the positivist model to incorporate all of science it was necessary then to show how the constitutive ideas of other disciplines are related to the laws of physics that the positivist model might appear to deal with satisfactorily. The classic formulation here is that of Oppenheim and Putnam (1958). They conceived of nature as being constituted by a hierarchy of objects which, in turn, defined a hierarchy of distinct sciences. At each level above the root level the objects are structures composed of objects from the next lower level. Thus elementary particles combine to form atoms, and atoms combine to form molecules; so the hierarchy ascends through living cells, multicellular organisms, and social groups. The sciences are individuated on the basis of the ontological level with which they deal, and reduction, on this model, consists in relating laws pertaining to the objects at one level with those of the next lower level via 'bridge principles' that identify the objects at any level with the set of lower level objects of which they are composed. The 'layer cake' reduction of the Oppenheim-Putnam model is thus a composite involving the conflation of two distinct ideas, one to do with inter-theoretic relations and another to do with mereology or the relations between wholes and their parts.

This view – a working hypothesis – that the world is ontologically stratified and unified by formal relations between its layers is problematic, not least for its requirement for the existence of theories of a kind capable of entering into deductive relations. In contexts beyond physics, and certainly in biology, mathematically lawful codifications of knowledge are in short supply. And in what sense exactly could the Oppenheim-Putnam scheme and its machinery of intertheoretic reductions be said to show how science 'works'? It is not obvious how the positivist project of 'rational reconstruction' relates to events and processes that actually play out in scientists' minds or that get enacted within scientific or wider society.[8]

Reduction

Reduction is at the heart of the Oppenheim-Putnam worldview, but what was meant by the term? How is the reduction relation to be characterized? For logical empiricist philosophers of science the question amounted to asking how reduction could be formally articulated given a syntactically structured conception of science. Thus the

[8] For Carnap this did not represent a valid concern (Carnap 1955). An account of why this should have been so could perhaps be framed in terms of how (1) the logical positivist programme was a principled attempt to purge philosophy of metaphysical excess, with logical abstraction and verificationism its main tools for accomplishing this, and how (2) the brain and its relations to the mind were still for the most part empirically inaccessible terra incognita.

important sense of the concept was seen as being to do with relations between formal scientific theories, and other possibilities were relatively neglected. The account of Nagel (1961) is the canonical expression of the logical empiricist view. His starting point is to note that reduction is about connections between different domains:

> Reduction, in the sense in which the word is here employed, is the explanation of a theory or a set of experimental laws established in one area of enquiry, by a theory usually though not invariably formulated for some other domain.
>
> (Nagel 1961/1979, p.338)

Excluded as unproblematic ('part of the normal development of science') are cases of what Nagel refers to as 'homogeneous' reduction. These, as he puts it, establish deductive relations between two sets of statements that employ a common vocabulary. In other words, the laws of the reduced theory employ terms that occur in vocabulary of the reducing theory. The interesting cases of reduction identified by the definition given above are different in that the domain in which the reducing theory was developed, or in Nagel's terms the 'primary science', 'seems to wipe out familiar distinctions as spurious, and appears to maintain that what are prima facie indisputably different traits of things are really identical' (p.340). Thus an aspect of reduction so construed is unification: a certain mode of description is shown to cover a set of cases thought previously to require distinct modes of description. But such reductions go further, for the identity Nagel cites is combined with an asymmetry of epistemic value: the reducing theory is held to be more fundamental than the reduced theory.

Schaffner (1967) describes Nagel's model of reduction as *direct*, in that the model asserts that theories are related by establishing correspondences between their terms, using so-called bridge principles where necessary. This contrasts with *indirect* accounts, such as that of Kemeny and Oppenheim (1956), in which there need be no formally expressible relationship between the terms of the two theories involved in a reduction. Rather, one theory T1 is said to reduce another theory T2 if all the observations that can be explained under T2 can be explained under T1, and if T1 makes additional predictions that are confirmed and which are inexplicable under T2. This indirect connection via observational overlap makes for greater flexibility, and associates the notion of fundamentality with explanatory range.[9] However, indirect reduction of this sort seems to be more about the

[9] The classic example is the subsumption of Newtonian physics by relativistic accounts. The latter continue to make accurate empirical predictions under conditions in which Newtonian theories break down, although within particular bounds similar predictions are made by both sets of theories.

criteria by which one scientific theory replaces another, diachronically, than about the kinds of case of particular interest here which involving competing claims about the terms by which phenomena in some domain are best to be accounted for (Sarkar 1998, p. 27).

A number of philosophers have developed taxonomies of reduction concepts that recognise broad families of variants. Van Gulick (2007) sees the major division as being between ontological and representational kinds of reduction, whilst others express more or less the same distinction by contrasting the former with epistemological reduction. There are substantial overlaps between these two categories, however, as Sarkar (1998) has noted: the kinds of object we suppose make up the world and the possibilities of and limits on the knowledge we can obtain by presuming their existence are not independent. Intertheoretic reduction is a case in point. Both ontological and epistemic components are involved, since theoretical terms are typically associated with objects we believe to be substantially realized in the world, and the structure of theories and the ways they work reflect and constrain our knowledge-making capacities.[10] In the Oppenheim-Putnam view intertheoretic reduction is conjoined with a particular ontological picture. This yields the mereological sense of reduction, which is about the ways in which wholes relate to parts. This sense of reduction is potentially problematic when applied to biological phenomena, since cells and organisms are not compositionally static even if they manifest relatively stable structure at a variety of scales. Such structure is typically maintained through processes that involve the constant turnover of matter, sometimes on surprisingly short timescales, and biological systems are in this respect very different from classical macroscopic mechanisms and machines (Dupré 2008) – as I discuss at length in Chapters 2, 4 and 5.

The idea that the properties of a whole are determined by the properties of its parts brings in issues to do with the direction of causation. The microdeterminist thesis is that causation runs exclusively from the micro to the macro, and indeed this can appear to be a natural view to take when one thinks about how the force field associated with a charged particle radiates outwards, say. But forces can converge too, and sometimes what happens at some delimited region is determined by the way in which multiple influences from different regions sum and interact. Hence there seem to be grounds for saying that

[10] No doubt a high degree of mutual reinforcement across the ontological/epistemic divide often occurs, the descriptive adequacy or predictive power of a theory providing grounds for believing in the reality of objects appealed to by its terms. Striking examples arise in high-energy physics of objects being directly associated with terms found to be necessary elements of empiricially satisfactory (non-disconfirmed) theoretical frameworks: particles have been sought and found on exactly this basis. (Indeed, the Large Hadron Collider owes its existence to the desire to confirm the existence of the postulated Higgs' boson.)

causation may also operate downwards – or perhaps we should say inwards – from the macro to the micro (see Chapter 6).

The mereological sense of reduction pertaining to material composition and part-whole relationships is related to another variant. This is the concept of methodological reduction, which approximates on one reading to the idea of material analysis.[11] This sense is highly pertinent to present concerns, since a notable characteristic of molecular and cell biological work has been the way in which the systems of interest have been investigated through a strategy of decomposition. Indeed the early development of molecular biology went hand-in-hand with the development of analytical techniques to support such a strategy, such as centrifugation, electrophoresis and chromatography (Kay 1993; Morange 2000). An important part of the debate about molecular biology and reduction is to do with the limits that methodological reduction qua material decomposition, and the apparent involvement of certain mereological, ontological and causal assumptions, might set on our explanatory powers. We can think of analysis as providing raw materials with which to synthesize explanations, and this work has a large cognitive component. An important question then is whether the raw materials so obtained provide a sufficient basis for answering all of the questions that interest biologists.

Reduction and explanation

Even a somewhat cursory survey such as the preceding brings to light the diversity of issues and dimensions implicated in different senses of reduction, including those of scale, stability, containment and causation implicit in mereological senses and the theme of epistemic fundamentality that is perhaps ubiquitous. One very rough-and-ready interpretation of reduction is that it relates to the terms by which something may best be explained. To be excessively reductionist, in this crude sense, is to have an unduly narrow – simplistic – view of what it is that needs to be taken into account in order to explain something adequately. In the limiting case there is what Dupré describes as unifactorial explanation, in which a range of phenomena is accounted for in terms of a single factor (Dupré 1993, p.87). He gives the example of Marxist theories that hold that all social phenomena are to be explained in terms of economics. Here it is clear that there is a connection with what is perhaps the most basic sense of 'to reduce', which is to diminish in number or scale. In the context of explanation what is being diminished is the number of

[11] A more straightforward sense of methodological reductionism is just the excessive reliance on one particular experimental (or theoretical) methodology.

explanatory factors. A conflation with this quite basic sense probably enters unbidden into many of the instances in which we speak of reduction – even in moderately sophisticated scientific or philosophical contexts.

Reduction in the positivist theoretical sense of Kemeny and Oppenheim or Oppenheim and Putnam can be connected with a particular view of explanation. Associated with this view is the idea that observation statements can be linked formally to theoretical statements, and that observation-proximate theoretical statements will be found to be subsumed under a smaller number of generalizations, which will in turn be subsumed under a still smaller number of yet more abstract generalizations, and so on. In other words it envisages a logic of science that can be explicated via the notion of reduction qua theoretical subsumption, establishing a theoretical unity as well as a hierarchy of abstraction. (The latter notion suggests something akin to – but not I think quite the same as – Feigl's well-known 1970 diagram showing how formal postulates are linked to the 'soil' of observation via concepts of various kinds (see Godfrey-Smith 2003, p.35).) Had the logical empiricist programme succeeded, the observations we make of natural phenomena might have been shown in this manner to be logically consistent with a small number of highly abstract principles, and this logical consistency could perhaps have been said to account for, or explain, the observed phenomena. However, this view of explanation, even if something like it can be postulated to have been implicit in the logical empiricist programme, was not explicitly articulated.[12] The reason, presumably, was that it would have run counter to positivist empiricism and anti-realism about theories (Godfrey-Smith 2003, pp.34-36).

Explanation

Reduction as theoretical subsumption played a major part in the logical empiricist picture of science, but it was not seen as providing a complete account of explanation. Far more influential in that respect was the deductive-nomological (D-N) or 'covering law' model of explanation (Hempel and Oppenheim 1948; Hempel 1965), described by Salmon as 'the fountainhead from which almost everything done subsequently on philosophical problems of scientific explanation flows' (Salmon 1998, p.68). The key features of the account, according to which explanations are construed as logical arguments, are

[12] Godfrey-Smith describes unification – presumably along lines similar to those outlined above in terms of a hierarchy of formal relations – as an 'unofficial' account of explanation within logical empiricism (2003, p.196).

highlighted by the name: the relation between the premises (or explanans) and the conclusion E (explanandum) is deductive, and the premises or explanans contain at least one law-like ('nomological') generalization (L_i) in addition to a set of statements of antecedent conditions (C_j):

L_1{, L_2, ... L_i}
C_1{, C_2, ... C_j}

E

Woodward (2003b) gives the example of explaining the future position of Mars (the explanandum). The laws in this case would include Newton's laws of motion and his gravitational inverse square law, whilst the mass of the sun, the mass of Mars, and the position and velocity of each, would be provided as antecedent conditions. (If we were numerically computing the position of Mars we might want to add the other planets of the solar system to the model to improve the accuracy of the prediction.)

The explanatory weight in the D-N model rests on the properties of laws as distinct from accidental generalizations, and on the deductive assurance that the truth of the explanans entails the truth of the explanandum.[13] Hempel gives as an example of non-nomological (i.e. accidental) generalization 'All members of the Greensbury School Board for 1964 are bald' (Hempel 1965, p.339). If this generalization were combined with the statement 'Harry Smith is a member of the Greensbury School Board for 1964', we would not thereby derive an explanation for the fact that Harry Smith is bald (if he is).[14]

Does the D-N model describe necessary conditions for explanation? Certain singular causal explanations show that it does not. For example, Scriven's example that it was the impact of my knee on the desk that caused the tipping over of the inkwell appears to be explanatory in the absence of any law or generalization (Woodward 2003b). Hempel's response was that this kind of narrative statement incorporates implicit law-like statements, and hence it ought to be regarded as 'explanation-sketch' (Hempel 1965, p.423). An alternative possibility, however, is that statements like this show that explanation can often be related to our ability to imagine phenomena (an idea to which I return in due course).

[13] The D-N model was later extended to cover cases where deduction is possible on the basis of laws that are statistical in character rather than fully deterministic (deductive-statistical explanation), and cases in which individual events are subsumed under statistical laws (inductive-statistical explanation).
[14] For Hempel laws are to be seen in the Humean sense as exceptionless, counterfactual-supporting generalizations that capture regularities in phenomena (Woodward 2003b).

As well as counterexamples to the idea that the D-N model constitutes an account of the *necessary* conditions for explanation, inasmuch as laws seem inessential, cases can be devised that call into question whether it describes *sufficient* conditions for explanation. Famously there is Bromberger's example of the shadow cast by a flag-pole (Bromberger 1966). The mathematical function that correlates the position of the shadow with that of the sun relative to the flag-pole is symmetric: either variable can be used to derive the value of the other since they are systematically correlated. Yet intuitively we feel that it is derivation in only one direction that has explanatory value: the position of the shadow of the flag-pole does not seem to *explain* the position of the sun. What appears to be missing here that could overcome the symmetry problem is consideration of causal factors.

Friedman (1974) and then Kitcher (1981, 1989) developed alternative accounts of explanation in response to some of the problems of the D-N model.[15] Their accounts retain the idea that explanations depend on or are constituted by relations amongst sets of linguistic statements, but are distinguished by their appeal to global rather than local properties of such statement sets. The key idea is that explanatory power is located not in putative special properties of laws and deductive relations but rather in the capacity of certain patterns of argument to unify our beliefs. Kitcher (1981) showed that Friedman's account has defects that are not readily overcome. His own account is somewhat complex but depends on the idea that the set of scientific beliefs accepted at a particular time, K, is maximally unified by a set of argument patterns termed the explanatory store over K, E(K). Explanation becomes a relative matter that involves achieving a balance between the economy of a set of resources and their power, suggesting a possible connection with the epistemic asymmetry of reduction relations noted earlier. Objections to the Kitcher model have been raised on a number of grounds. Woodward argues, with Psillos, that it too suffers from problems in its capacity to handle the causal asymmetry of explanation (Woodward 2003b; Psillos 2002, pp.276-278). Gijsbers (2007) argues that the account is fatally flawed by the fact that it entails that every proposition can be explained by itself.

Causation

Both the D-N model and the unification accounts of explanation just mentioned eschew consideration of causal issues, and this creates significant difficulties. Part of the

[15] Salmon (1998) notes an additional problem with the D-N model: its inability to provide an account of explanations of general laws – despite the importance of theoretical subsumption by reduction for the logical empiricist programme (p.69).

trend away from logical empiricism and its metaphysical abstinence has been a greater willingness to frame explanation accounts in causal terms (Salmon 1989, pp.107-116). However, this comes with its own set of problems. Hume noted the absence of criteria by which we might distinguish unambiguously between genuinely causal and non-causal events. All we have to go on are the regularities of priority, contiguity and constant conjunction of events, and this seems an inadequate basis for causal understanding (Hume 1739/1978). Mackie made perhaps the definitive attempt to develop a workable regularity-based account, in which he introduced the notion of the INUS condition (Mackie 1974). The idea is that a cause can be regarded as being an ***i****nsufficient but* ***n****ecessary part of an* ***u****nnecessary but* ***s****ufficient* condition for an event. Despite the sophistication of this analysis there remain problems, such as its failure to capture fully the difference between genuine causes and joint effects of a common cause (Psillos 2002, p.90).

A recurrent and intuitively appealing idea is that efforts to explicate the notion of causation might succeed if they were framed to incorporate counterfactual conditionals, a possibility associated in particular with David Lewis.[16] The root notion here is that we say that X caused Y when there are grounds for supposing that if X had not occurred then Y would not have occurred. But again difficulties are encountered in particular cases, for example concerning causal over-determination and pre-emption (see Psillos 2002, pp.96-100 and Reiss 2007, pp.24-33), and the project looks ultimately unpromising.

Salmon attempted to provide a workable account of physical causation and causal-mechanical explanation. His idea was to distinguish between causal and non-causal events and processes in terms of the capacity to transmit information or a 'mark' (Salmon 1989, pp.107-111). A genuine causal process can do this because if it is modified at one time the modification persists without further intervention. For example, a light beam can transmit information because a modification made to the beam, e.g. by the insertion of a coloured filter, results in a persistent change. As Salmon puts it, causal processes 'provide the causal connections among events that happen at different times and places in the universe' (Salmon 1989, p.109). A causal connection or interaction is the intersection of several causal processes in which the 'processes are modified in the intersection in ways that persist beyond the point of intersection, even in the absence of further intersections' (Salmon 1998, p.71). Salmon's account also has a counterfactual aspect, in that a causal process need not actually transmit a mark; it is just that it has the potential to transmit a mark. Dowe later modified the account by reframing it in terms of the exchange of conserved physical

[16] The counterfactual aspect of causation was noted early on by Hume.

quantities (e.g. momentum), and argued that interactions are causal in virtue of the transmission of conserved quantities (with no counterfactual element) (Salmon 1998, chapter 16).

Even this account, however, proves inadequate as a basis for making sense of causal explanation in general. This is because causal explanations, even when they relate to the objects and phenomena of scientific study outside psychology, frequently appeal to factors besides the causal-mechanical interactions that actually occur. In the next chapter, for example, I note a criticism that has been made of a popular contemporary account of mechanism: it fails to encompass cases in which we explain phenomena causally on the basis of an absence of interaction. Sometimes things occur because they are allowed to happen, i.e. because nothing intervenes to prevent them occurring. And as debates about causation have made plain, objective phenomena such as spatiotemporal event correlations are often ambiguous or misleading about the underlying causal properties and powers of objects (e.g. Mackie 1974; Cartwright 1983 – especially Essay 1: 'Causal Laws and Effective Strategies'). Hence, pessimistically, we might well conclude that attempts to capture the objective nature of the causal relation in terms just of the patterns that occur, actually or counterfactually, between events in the world are destined to fail.

Despite all these difficulties we often *are* able to make causal judgements and ascriptions that serve well as a basis for explaining and predicting phenomena. How are our successes to be accounted for? An approach that figures less prominently in the philosophical causation literature than accounts of the kind just described is to see causation not solely in terms of objective properties of events and processes in the world but rather as something that in addition necessarily involves a significant psychological component. This somewhat Kantian approach would no doubt have been rejected by the logical empiricists as being metaphysically excessive, yet in the following chapters I aim to show its benefits. Specifically I shall make use of the idea (already mentioned) that we interpret phenomena using cognitive models and schemas of various kinds. This approach connects, I suggest, in an interesting way with the concept of mechanism that in recent years has come into vogue amongst philosophers of science.

Explanation and understanding

The tradition in philosophy of science, part of the legacy of positivism, has been to emphasize explanation over understanding. Moreover the tendency has been to concentrate on explanations as explanatory objects rather than on explanation as an act, i.e. as it pertains to the verb 'to explain' (with an exception being the illocutionary approach of Achinstein (1983)). There have been important exceptions to this privileging of explanation, however (e.g. Friedman 1974), and recently the topic of understanding has attracted renewed interest (de Regt and Leonelli 2009).

Part of the reason for the privileging of explanation has been the positivist aversion to psychology and sociology, and the associated belief that the ideas of those disciplines are inessential to explication of the logic of science. The concern has been to focus on what is objective rather than subjective, as borne out by Hempel when he contends that 'such expressions as 'realm of understanding' and 'comprehensible' do not belong to the vocabulary of logic, for they refer to psychological or pragmatic aspects of explanation' (Hempel 1965, 413), and then later when he says:

> Very broadly speaking, to explain something to a person is to make it plain and intelligible to him, to make him understand it. Thus construed, the word 'explanation' and its cognates are pragmatic terms: their use requires reference to the persons involved in the process of explaining. ... Explanation in this pragmatic sense is thus a relative notion: something can be significantly said to constitute an explanation in this sense only for this or that individual'.
>
> (Hempel 1965, pp.425-426)

But as Friedman notes in connection with these remarks, pragmatic can be taken to mean either psychological or subjective, and these are not the same. He argues that there can be a sense of scientific understanding that is both psychological ('pragmatic') *and* objective – and that the D-N model fails to show how explanations result in such understanding (Friedman 1974, pp.7-8).

De Regt (2009) notes the persistence of the objectivist tendency to seek accounts of explanation that make no appeal to psychology, and highlights enduring scepticism about the philosophical value of psychological interpretations of understanding, for example as evident in Trout (2002). He defends an account of understanding that challenges these objectivist positions by recognising the epistemic relevance of skills and

judgement. This account, which touches in various ways on what I say later, begins by distinguishing between three different senses of understanding:

FU: feeling of understanding – the phenomenal experience accompanying an explanation

UT: understanding a theory – being able to use a theory

UP: understanding a phenomenon – having an adequate explanation of the phenomenon

UP is the objective sense of understanding that is, as de Regt puts it, an 'essential epistemic aim of science' (p.26), and which may or may not be accompanied by **FU**. Regarding the latter de Regt agrees with Trout that this phenomenal sense of understanding, the possible accompaniment of **UP**, is epistemically irrelevant. The novel aspect of de Regt's argument centres on **UT**: he claims that **UT** is necessarily pragmatic, but is necessary for **UP**. Specifically what is required for objective understanding is pragmatic skill and judgement. The example given (deriving from Harold Brown) is the construction and checking of a logical proof. Each step in proof construction involves deciding which rule to apply at that step, and this decision-making is not itself guided by a set of rules; rather it is shaped by the application of a learnt skill (p.26). This means that scientific practice interweaves epistemic and pragmatic factors and that the latter have epistemic significance – contra the arguments of Hempel and others that they can and should be kept separate, and that the epistemic necessarily excludes the pragmatic.

How do these ideas relate to explanation and understanding? De Regt espouses Cartwright's simulacrum account of explanation, in which 'to explain a phenomenon is to construct a model that fits the phenomenon into a theory' (Cartwright 1983, p.17). Here a model is understood to be a mediating, idealized construction which combines aspects of theory and empirical information (Morgan and Morrison 1999). Models on this semantic view of explanation perform a similar role to bridge principles in the reduction accounts mentioned earlier, but are considerably more flexible in terms of structure and manner of coupling to the relata between which they mediate (Giere 1988). It is this flexibility and freedom of choice involved in their construction that makes models a point of entry into scientific practice for the skill and judgement to which de Regt draws attention. Specifically, attaining pragmatic understanding (**UT**) means exercising tacit skill and judgement, for example in relation to approximation, idealization or visualization, in order to construct perspicuous and epistemically potent models. This in turn requires that theories be

intelligible, where intelligibility is defined as 'the positive value that scientists attribute to the theoretical virtues [e.g. visualizability, simplicity] that facilitate the construction of models of the phenomena' (p.20). De Regt concludes on this basis that scientific understanding is not completely objective, i.e. independent of the subject. Two people may possess the same theories and background knowledge, and one may understand a phenomenon while the other does not. The differences lie in their relative skills in developing and working with models that effectively link theory and phenomena.

Explanation and understanding in biology

Aspects of De Regt's account of scientific understanding overlap significantly with my overall orientation, which accentuates the desirability of integrating internal (psychological, subjective) considerations with external (objective, ontic) ones. But at the same time it is unclear to what extent the account lends itself to biology, for it remains largely wedded to a positivist-derived concern with theory and the semantic accounts of theories and explanation to which that concern has given rise. To conclude this introductory chapter I now reflect briefly on biology's distinctive explanatory character. In the concluding chapter I reconsider de Regt's account in the light of what I say in the intervening chapters about ontological and epistemological issues in molecular and cell biology.

I have already drawn attention to one respect in which much of biology appears to differ from the physical sciences: its relative 'lawlessness' and corresponding lack of formal theories (Beatty 2006). Physics is largely composed of mathematically expressible theories involving law-like quantitative relationships amongst variables, and chemistry is largely defined by the desire to account for and exploit law-like patterns of molecular interconversion. In many branches of biology, however, there is a need for more complex, irregular, or context-dependent epistemological and ontological frameworks. In these more heterogeneous knowledge structures, some of which I discuss in subsequent chapters, lawful regularities figure more as the exception than the rule. How much significance should be attached to this difference? On the one hand it seems clear that philosophical accounts of explanation and understanding that are framed principally in terms of theories qua mathematically expressible generalizations will be of limited utility in many biological domains, including molecular and cell biology. On the other hand maybe it pays to be sceptical about the viability of drawing neat distinctions between the major scientific

disciplinary categories. Perhaps simple views that contrast biology *tout court* with physics, say, are liable to be not just simple but simplistic.[17] The fields of population biology, population ecology, population genetics and epidemiology, for example, are replete with (perhaps to the extent of being defined by) mathematical models, approximations and generalizations.

Another important characteristic of explanation and understanding in many branches of biology is the central importance of visualization techniques. Investigation of the cell, for example, arguably began in the seventeenth century with the microscopic investigations of Leeuwenhoek and Hooke.[18] Its subsequent development has been closely intertwined with the development and evolution of techniques for imaging cells and their contents, such as staining and electron microscopy. In recent years fluorescence-based techniques have become particularly important, especially in conjunction with possibilities for genetic manipulation (exemplified by the rapid adoption of green fluorescent protein (GFP) by cell biologists). At the molecular scale our knowledge of protein structure likewise depends on the 2D images and 3D models generated by specific techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, increasingly allied to the possibilities for computer modelling and visualization that I discuss in Chapter 3. Publications reflect this visual emphasis – to the extent that the journal *Cell* has investigated the use of graphical abstracts as a supplement to the traditional textual abstract.[19] Yet here again, as with laws, it is dangerous to generalize to all of biology. Structural characterization and visualization may be a fundamental aim of much work in molecular and cell biology, but visualization techniques arguably play a lesser role in population biology. By this I mean that while they still play a part, for example in representing and analysing the numerical datasets generated by mathematical models, the role of visual images here is secondary and instrumental, and obtaining them is not the principal goal or chief burden of the researcher. Figures in publications from these fields often take the form of graphs depicting quantitative relationships, as is common in publications in the 'harder' physical sciences (Smith et al. 2000).

A third characteristic of explanation and understanding in biology is the part played by functional concepts. It is generally agreed that a significant difference between biological explanations and explanations in the physical sciences is that the former tend to make

[17] See, for example, Elgin 2006 and Morgan 2009 on the idea that biology is not devoid of laws.
[18] Although it could probably not be said to have represented a distinct field of scientific enquiry until at least the mid-nineteenth century, with the advent of cell theory (Harris 1999).
[19] http://beta.cell.com/index.php/2009/07/article-of-the-future/ (last accessed 3 December 2009).

heavier use of functional terminology and concepts than the latter (see e.g. Hempel 1965, p.297; Ruse 2000). (I discuss this aspect of biological explanation in Chapter 7, in relation to 'information talk' and its application to genetic and genomic processes.) Why should this be so? One reason may be the more abstracted character of the physical sciences. The phenomena studied by physics and chemistry are often synthetically derived: the scientist constructs an elaborate experimental system which elicits or makes manifest a phenomenon that might not otherwise occur under normal terrestrial conditions[20]. The focus in physics is thus on the general properties of matter or forces believed to underpin the behaviour of specific physical systems and material configurations. These general properties are seen as exactly that – general (in the sense of universal) – and frequently they are revealed by and studied in non-natural systems. Sometimes they must be inferred from what can be observed on the basis only of a substantial body of theory, and acceptance of the inferences made is conditional on acceptance of that body of theory (which in turn is contingent on agreement about exactly what theories are relevant). In these circumstances, divorced from specific material configurations that might occur 'naturally' in the world (i.e. in the absence of artificial experimental constructions), and perhaps at some remove from direct observation, it is difficult to regard the ontological elements that are discerned or appealed to by physicists as being 'for' anything in particular. If they are functionless then (I conjecture) perhaps it is because they are context-free.

The upshot of these initial reflections is that explanation and understanding in specific areas of biology appear to have distinctive characteristics that are not common to all areas of biology and which are not necessarily shared with the physical sciences. Any account that aspires to provide a comprehensive view of scientific explanation and understanding will almost certainly have to take these characteristics into account. I shall argue that the requisite account-taking must combine what we know about biological systems and what to an increasing extent we know about ourselves as cognitive agents. (It has been suggested that one of our cognitive characteristics is a generalized bias towards simplicity (Chater 1999), a proposal which calls to mind the well-known epistemic dicta of

[20] An obvious objection here is not that this claim is wrong, but that it fails to distinguish physics from biology since increasingly research in the latter also depends on capacities to modify, intervene in or even synthesize novel systems, for example by applying genetic engineering techniques. However, there remains an important difference in that physics is concerned less with understanding specific material configurations than with elucidating relationships between abstract quantities and properties that may in principle be manifested by a diversity of physical systems. The relevant material configurations may be entirely absent from the known universe, or may at least not occur in a scientifically tractable form, and hence it is necessary for physicists to build colliders and so on in order to generate phenomena that reveal the physics in which they are interested.

Ockham[21] and Newton[22]. Perhaps such a tendency is what underlies some of the strands of reductionism that the complexity of biological systems is so often liable to wrong-foot.)

In the next chapter I focus on the concept of mechanism, which it has been argued represents an especially important explanatory concept in molecular and cell biology. Over subsequent chapters I will examine the relationship between mechanistic explanation, molecular and cellular complexity, and biological function, to illustrate how ontological, external and objective factors are interwoven with epistemic, internal and subjective ones. Additionally I aim to show some of the implications of this interweaving for the accounts we give of explanation and understanding. The following summary sets out the path I shall take.

Outline summary

In Chapter 2 I describe several senses of mechanism. Macroscopic artefacts such as clocks, engines and other mechanical contraptions ('machines') I take to be paradigm instances of a material sense of mechanism. I suggest that it is the neat alignment between structure and function in the generation of patterns of entailment that best captures what is distinctive and causally interesting about them. Another more general sense of mechanism concerns causal processes, a category which potentially includes singular and more-or-less unique causal phenomena. It is helpful to think of causal mechanisms in this second sense as being sometimes instantiated or implemented by material, machine-like, mechanisms in the first sense. I then discuss the popular and influential account of mechanisms and mechanistic explanation recently advanced by Machamer, Darden and Craver (MDC). Their account frames the concept of mechanism in terms of a dual ontology of activities and entities. Whilst it goes some way towards addressing problems in biological explanation to do with the relative ontic and epistemic status of structures and processes, it can be considered only a partial account. I argue that it lacks the means to distinguish satisfactorily between equilibrium and non-equilibrium structures and processes, and downplays the explanatory importance of functional attributions. However, the idea expressed by MDC that mechanism descriptions are explanatory because they make phenomena intelligible,

[21] Occam's Razor is the principle that *entia non sunt multiplicanda praeter necessitatem* ('entities must not be multiplied beyond necessity').
[22] Rule 1 of the 'Rules of Reasoning in Philosophy' listed in the *Principia* (second and third editions): 'We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances'.

and that they do this by 'showing how' they come about, is attractive, and one to which I return.

Chapter 3 examines the biologically fundamental phenomenon of protein folding, the spontaneous process by which a synthesized polypeptide chain adopts (in most cases) a sequence-specific three-dimensional structure. The phenomenon constitutes an interesting example of a process that is best regarded as schematically mechanistic, or as semi-mechanistic. Examination of the nature of protein folding and the methods used to investigate it highlights the importance of visualization for explanation and understanding, consistent with what MDC say about intelligibility. It also forms a useful background to the discussion of biological complexity that occupies Chapters 4 and 5. I do not provide a monolithic account of complexity but rather approach the topic from several directions, for it is in the nature of the biological complexity exhibited by the cell to resist subsumption under a single all-embracing account. However, it proves useful to pay attention to the diversity and biophysical specifics of cell processes. Doing so leads me to discuss structures versus processes as explanatory bases, as well as the nature of biological organization. The chapter closes by assimilating these reflections into a novel and somewhat epistemic perspective on mechanism.

The topic of emergence that arises in Chapter 4 in the context of cellular complexity is the subject of Chapter 6. I develop a perspective that stresses the intimate relation between the ontology and epistemology of emergent phenomena. My approach is compatible with several other recent accounts, but I argue that the epistemic neglect that is often a feature of those accounts is potentially more problematic than their proponents would have us believe. My negative account of emergence depends on the idea that (roughly speaking) we see phenomena as emergent when we lack causal schemas that allow us cognitively to model the entailment structures that underpin them. This offers as its corollary a corroboration of my preferred perspective on the metaphysics of causation. In addition it sheds further light on the perception of complexity and constraints on our understanding of complex phenomena.

Chapters 5 and 6 show how our understanding of phenomena in the world reflects our psychological natures. Function is an important explanatory concept in biology, but functions are not given to us as properties inherent in the objective nature of phenomena. Rather we attribute functions to structures and processes. On what basis do we do this? Frequently the answer is none too clear. I explore this theme in Chapter 7 by investigating

how the causal status of the genome has often been understood in informational terms. I am somewhat agnostic and non-prescriptive about such biological 'information talk', but think it interesting and useful to reflect on why informational properties have been attributed so readily in respect of genetic processes. I suggest that an under-appreciated view of the genome is as a data storage or memory structure that – necessarily, from an evolutionary point of view – partially encodes phenotypic details about organisms. This view goes well beyond what might reasonably be held to be 'in the phenomena', and teleological reasoning about the conditions required for evolution appears readily to play a part in its comprehension. Trying to see how material cycles of development and reproduction fit particular teleological schemas confirms the suspicion that biological explanation often depends on exactly the psychological factors that have been downplayed by the positivist philosophical tradition.

The thesis concludes with further reflection on the topics discussed and with a number of speculations concerning explanation and understanding, in biology and in general. In particular I synthesize what I say regarding mechanistic understanding to provide a rough sketch of how the sensory (and especially visually grounded) forms of understanding often sought in molecular and cell biology might work in cognitive terms. The overall philosophical burden of the thesis is that epistemological considerations are as important as ontological ones for our understanding of explanation and understanding. Getting to grips philosophically with those and related subjects requires us to combine internalist and externalist styles of thinking and give consideration to the total system of mind in the world. This is because the concepts involved in discussion of these subjects frequently pertain to the relationship between mind and world, rather than being localizable to just one or the other.

2. Concepts of mechanism

Introduction

> **mechanism** n. **1** a piece of machinery. **2** a process by which something takes place or is brought about. **3** [Philosophy] the doctrine that all natural phenomena allow mechanical explanation by physics and chemistry.
>
> *Concise Oxford English Dictionary* (Tenth Edition, Revised)

Much recent work on explanation, especially explanation in biology, has focused on the concept of mechanism. Very plausibly it is argued that biological explanation must be accounted for in terms other than the D-N model or any other framework that depends as heavily as it does on nomic regularities, for in biology these are in general few and far between. It is suggested that the concept of mechanism provides just the kind of philosophical resource we need. But what is the attraction of mechanism as a concept for making sense of explanation in biology? One might suppose its appeal to be connected with its seeming promise for relating causation to overtly material and spatial factors, thereby getting away from the syntactic or logical concerns that the logical empiricist tradition emphasized, say (as well as from the semantic accounts of theories to which it gave rise – see e.g. Thompson (1989) and Sloep and Van der Steen's (1991) review). In addition there is the fact that that 'mechanism talk' plays an important part in the explanations scientists give of phenomena, there being an abundance of references in the scientific literature to the mechanism for or of X. Protein scientists, for example, talk about 'the mechanism of protein folding', and in reflecting on their use of mechanistic language cell biologists have observed how 'those of us who work on cell signaling would be hard-pressed to avoid terms such as 'machinery' and 'mechanism'' (Mayer, Blinov and Loew 2009, p.81). These researchers go on to question the basis for this reliance on mechanism talk:

> The analogy between cell signaling and man-made machines is all-pervasive, frequently adopting the imagery of elaborate clockwork mechanisms or electronic circuit boards. This perception is undoubtedly shaped by what we know: the machines that we use in our everyday life and the ways that we describe such machines in diagrams or words. But is this really an accurate, or useful, description of the actual processes used by cells?
>
> (Mayer, Blinov and Loew 2009, p.81)

(Answering that question will occupy much of the next several chapters.) Mechanism talk is not confined to biology, however: cognitive psychologists and philosophers of mind also speak of mechanisms. For example (picking a book that happens to come readily to hand), Alvin Goldman notes on p.3 of his *Simulating Minds* that insect species 'feature sharp divisions of economic labor and intricate mechanisms of communication'.

Do these different usages have a common basis? The dictionary definitions quoted above – especially the second one – seem to pick out quite economically apparently different senses of the concept involved in such phrases.[23] But a number of questions arise in relation to those definitions and the intuitions to which they appeal. What distinguishes a piece of machinery from a process by which something takes place or is brought about – are the two senses related at all? If machines can be thought of as somehow representing a special or limiting case of the latter, then in what respect? (The third definition is rather different from the other two, and clearly taps into the extensive debates that have taken place concerning reduction. As I have already discussed major aspects of these I do not consider them here.) To begin with we can distinguish between material and causal senses of mechanism, with the latter being more general and more plastic. The two senses are related if one thinks of a mechanism as a causal structure, with some causal structures being realised materially.

Machines, if they are taken to be largely macroscopic and for the most part solid-state artefacts, are the classic exemplars of the material sense of mechanism. A key feature of the machine conception is the approximate alignment of structural and functional decompositions. After discussing this idea and some general issues concerning structures and functions I consider the much looser causal sense of mechanism involved in the second dictionary definition. Looseness notwithstanding, not all causal processes attract the designation of mechanism, and I outline some of the limitations under which the term is employed in this causal sense.

[23] My colleague Dan Nicholson has reminded me that in using the Concise Oxford Dictionary as a point of departure in this way I am following Michael Ruse, who structures his (2005) on exactly this basis. In his own PhD thesis Dan is also examining the concept of mechanism, but from a perspective that is more historical than mine. In addition he is, if I understand him correctly (on the basis of the several presentations I have heard him deliver), concerned to make an ontological case for the view that it is thoroughly mistaken to think of organisms in mechanistic terms. I am interested, on the other hand, in how and under what circumstances biological explanations, including those that invoke mechanism talk, work.

These initial explications of mechanism form the background against which I discuss recent philosophical articulations of the concept and discussions of mechanistic explanation in science, with the emphasis on biology. The foundations of these influential neo-mechanistic perspectives are papers by Machamer, Darden and Craver (2000) (which outlines what henceforth I shall refer to as the 'MDC account' of mechanism), Craver (2001), Glennan (1996, 2002), and the papers making up a 2005 special issue of *Studies in the History and Philosophy of Biological and Biomedical Science*. A number of other publications have appeared since 2005 which develop and amplify discussion of some of the issues on which those works dwell (e.g. Bechtel 2006), and the mechanism literature continues to expand rapidly. Unless otherwise stated, it can be assumed that when I talk about neo-mechanistic perspectives I am referring to the MDC account as the dominant exemplar.

I take one of the principal motivations underlying these new articulations of the concept of mechanism to be the desire to make sense of biological explanation in a way that does justice to the lability of biological structures, and to the fact that explanatory tasks consequently often focus on processes as much as on structures. Biological mechanisms, construed as the ontic targets of mechanistic explanation in biology, are thus rather different from the man-made structural assemblies we often think of as being quintessentially mechanistic. This basic orientation is one that I share, although I argue that there is more to be said about mechanism in biology than has been said thus far. In particular, I suggest that it is helpful to augment the new perspective with older ideas about mechanism that emphasize the importance of function.

One consequence of thinking about mechanism in structure—function terms is that it becomes easier to distinguish between different kinds of biological process. I see it as a limitation of the MDC account that it employs novel terminology in a manner that is sometimes rather ambiguous and vague. The paucity of discussion by MDC of the exact respects in which cell biological phenomena might *not* be amenable to mechanistic interpretation leaves it unclear just how significant the concept of mechanism is to understanding biological systems, irrespective of whether the concept is construed ontically (in terms of what such systems are like) or epistemically (as regards how we gain knowledge of them). An account of mechanism so flexible that it looks potentially capable of subsuming all molecular cell biological phenomena is probably less useful than a narrower account capable of distinguishing between different kinds of such phenomena. An account of the former sort is likely to achieve its universality by discarding valuable distinctions, whereas the latter kind of account is likely to say something important about the diversity

of the phenomena that interest us and (by implication) the methods by which we investigate and explain those phenomena.

One of the main tasks in this chapter is to interpret neo-mechanist positions regarding biological phenomena while keeping in mind the diversity, causal complexity, and explanatory challenges posed by such phenomena. I end up arguing that the concept of mechanism that MDC articulate is most promising not in relation to the features that have attracted most attention (concerning its ontological commitments) but rather as regards what it says about the epistemology of mechanistic explanation. Thinking about structure—function relationships may help to distinguish between mechanistic and non-mechanistic (or between more and less mechanistic) cellular processes, potentially providing the basis for a more comprehensive account of explanation in molecular and cell biology than we currently possess.

This chapter and the next three form an integrated series. In Chapter 3 I examine the phenomenon of protein folding while in Chapters 4 and 5 I describe the complexity of cellular phenomena, and argue that current mechanistic concepts provide an inadequate basis for making sense of some of the most significant features of such phenomena. Especially problematic is the relationship between structure and process. This leads me to develop a new perspective on mechanism that is, I argue, better able than the MDC account to deal with some of the distinctive characteristics of biological systems.

A material conception of mechanism: machines

A long tradition, going back at least to the mechanical philosophy we associate with Galileo and Descartes, identifies the concept of mechanism with the structural and functional characteristics of mechanical devices and machines such as clocks (Canguilhem 1952/2008).[24] This is the first sense of mechanism mentioned in the dictionary definition quoted at the start of this chapter. On this view systems are regarded as mechanistic to the extent that they share various features with such artefacts. These include being constituted of parts that are assembled into structures having definite and relatively stable spatial relationships, and when the relationships amongst parts do change it is in accordance with particular points, lines and planes of articulation. Typically such structures are readily represented diagrammatically:

[24] Much of what I say in this chapter is based on or inspired by Gregory (1981, Chapter 3) and Dupré (2008).

> A key property of such a structure [i.e. a 'bona fide manmade machine'] is that it can be described in terms of a parts list or blueprint for how those parts fit together. Any machines, from a can-opener to a computer chip to an Airbus, can be rendered in a diagram with sufficient detail that someone who has never seen one could make it from the component parts. Using the diagram, one could assemble any number of individual machines, each of which would be virtually identical in appearance and performance.
>
> (Mayer, Blinov and Loew 2009, p. 81)

The shape and shape stability of a machine's parts are critically important. Shape stability is attributable to the fact that – under the intended operating conditions of the mechanism – the parts are solid aggregations of matter. Shape stability matters because part shapes put very stringent spatial constraints on the relative motions that can occur between parts under standard operating conditions, and it is these structural constraints that define the action of the mechanism (qua pattern of configurational change). A simple illustration of the importance of shape stability is the fact that an extended cylinder situated in an appropriately sized cylindrical hole running through another object will rotate only if it is straight. A more complex example concerns a system comprising two cogs, for which it is intended that rotation of one of the cogs will cause the rotation of the other. One cog will mesh satisfactorily with another throughout its rotation only if its teeth are uniformly spaced in a circular arrangement, with the tooth spacing matching that of the teeth of the cog with which it is intended to mesh, and with its axis of rotation lying at the centre of the circle its teeth define, and so on.[25]

Associated with this conception of mechanism is the idea that parts have particular functions. If the overall mechanism has function F, then its components can be thought of as having sub-functions that contribute towards the accomplishment or performance of F. The overall structure of the mechanism is the result of the hierarchical assembly of a number of structural sub-systems, the overall function of the mechanism results from the combination of sub-functions, and structural sub-systems implement the sub-functions. In other words there is an alignment between the structural and functional decompositions of the mechanism. This alignment of structure and function is related to notions of system

[25] Actually, even this short description is not strictly correct. It is unnecessary for the teeth of the two cogs to share the same uniform spacing around the entire circumference, so long as the tooth spacing on one cog corresponds with that on the other at the point of engagement, for each rotational position (of each cog) – and so long as the circumference of one cog is an integral multiple of that of the other. And for the teeth of the two cogs to always engage the cogs need not even be circular, if the location of the axes of rotation are either positioned appropriately vis-à-vis rotational positions or can move laterally in a system that pulls the two axles together to maintain cog engagement at all times.

decomposability and modularity discussed by, for example, Simon (1996) (in relation to which see also Chapter 4).

The toy mechanical clock shown in Figure 1 illustrates these ideas. To understand how the clock works involves recognising and appreciating the roles of several functionally distinct structural sub-systems. First there is the sub-system comprising the mainspring, winder and driving wheel. Turning the winder places the spring under tension that is released by rotation of the driving wheel. Rotation of the driving wheel is communicated via a series of cogs called the driving train to the clock hands. The driving train can be thought of as a functionally discrete sub-system. The bell clapper arm is driven by the driving train in such a way that the clock chimes four times for each rotation of the hour hand. If this were all that the clock consisted of – the driving train, the hands and the bell/clapper sub-system – then the relevant hand rotations and chimings would occur, but at the wrong rates. To ensure that the clock keeps time requires the presence of another sub-system, the escapement mechanism. Another cog train, distinct from the driving train, couples the driving wheel to the escape wheel, the teeth of which engage with the anchor that is attached to the pendulum. The movement of the pendulum, the period of which is determined by its mass and length, causes the anchor to rock back and forth. With each rocking movement the escape wheel is permitted to rotate by one cog tooth (driven by the driving wheel's mainspring-powered rotation) as one of the anchor teeth lifts clear, before being arrested as the other anchor tooth engages. The escapement thus acts as a brake on the driving train, and places the hand rotations under the regulation of the pendulum.

Another more subtle feature should be noted. In order to ensure that the pendulum regulates the driving train, it must be able to exert a force at the escape wheel sufficient to brake the rotation of the driving wheel. In practice this is readily achieved, since the gearing is such as to favour the possibility, i.e. the escape wheel rotates much faster than the driving wheel. Whether this strikes the reader as readily intelligible will depend on a variety of factors, but practical experience with building and manipulating systems of gears using Lego or Meccano, for example, no doubt helps to develop the relevant intuitions. (I shall return to some of these psychological matters in due course.)

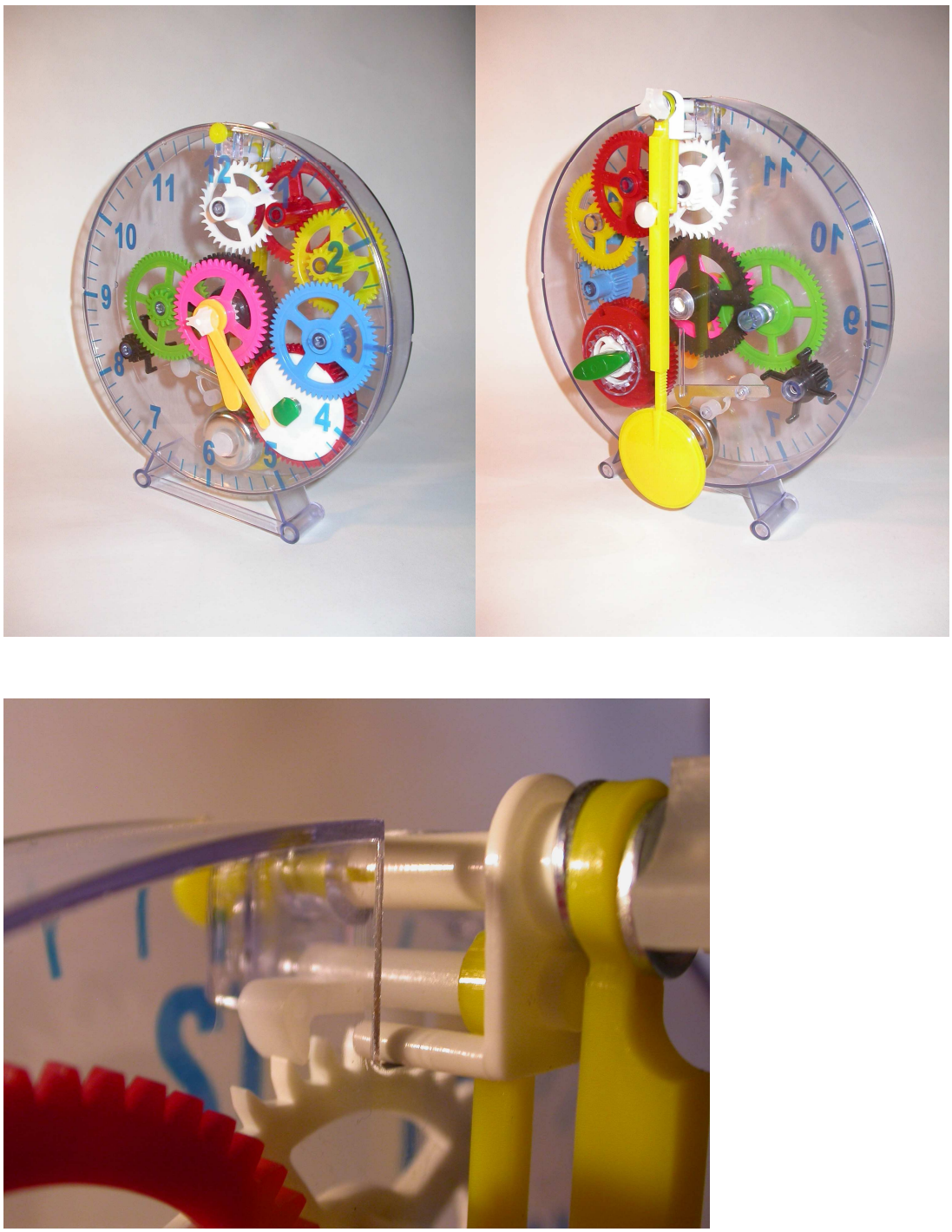


Figure 1 – Mechanical clock

The driving wheel is the lower of the two red cogs (with white front fascia). This turns the black cog, which turns the green cog, which in turn rotates both the pink cog that turns the hour hand and a smaller white cog (barely visible) that turns the minute hand. The driving wheel is regulated by the escapement mechanism, formed by the cog series blue – yellow – upper red – upper white. The last is coupled to, and regulated by, the pendulum via the anchor (lower picture). The bell is the silver dome behind dial numeral '6'.

First thoughts about structures and functions

The material, physical aspects of the characterization given above of machine-type mechanisms – that they are composed of solid-state parts arranged in particular ways defined by their shapes – provide the foundation for the more abstract claim that in such mechanisms the structural and functional decompositions are aligned. But given that this alignment too is a spatial matter – structural and functional boundaries are in rough spatial correspondence – why does the claim seem more abstract? It is, I suggest, because functional attribution is itself a somewhat problematic idea, raising not just the question of what is meant by a function, but also the issue of the basis on which we make functional attributions in respect of system components.[26] In contrast, structural concepts seem much more straightforward, at least in the context of machine-type solid-state mechanisms. This is presumably related to the visibility and stability of solid-state forms. Our perceptual capacities mean that we are well-equipped to recognize objects in the world by their shapes, colours, textures and so on, and we have strong aptitudes for recollecting, imagining, transforming and comparing the objects we recognize (Bruce et al. 2003). In addition we are capable of imagining both wholly new (as yet unseen in the world) objects and modified versions of objects we have experienced, and we can imagine acting on these imaginary objects in a wide variety of ways. Such imaginary thought is underpinned by the confidence with which we believe that, under the conditions under which we ourselves exist, solid-state objects are likely (all other things being equal, if they are not in tension, etc.) to retain their forms.[27]

Yet even in this structural sense matters are not as straightforward as they might seem. We experience ourselves as having little difficulty thinking about structures as three-dimensional objects, but three-dimensional structure must be derived from raw sensory stimuli. The processes by which this occurs depend on the interplay of immediate visual experience with prior visual experience, and on the integration of visual data with data from other senses such as touch. Perceptual illusions sometimes shed light on the automatic and subconscious bases on which such derivations are usually made (Gregory 1997). My purpose here, however, is not to discuss in laborious detail the psychology of visual perception and its role in the generation of our structural knowledge of the world. Suffice it to say that the fact that the structural side of the structure—function relation

[26] I pick out the issue of functional attribution for further analysis in Chapter 7.
[27] Later I discuss the idea that living systems are organized in ways that potentially make the drawing of parallels with the structures of artefacts somewhat problematic.

seems relatively straightforward is largely a reflection of the strength and automaticity of our combined perceptual and cognitive capacities. An upshot of this is that we tend to think of persistent structural phenomena as being unproblematically 'out there' in the world, and hence view them as more a matter of (objective) ontology than (subjective) epistemology.

On the functional side things are more difficult. What are functions, and how do we attribute functions to objects and systems (including biological ones)? These are enormous questions attended by a correspondingly voluminous literature, and I will not attempt to do them full justice here. As regards the first question Wouters (2003, 2005) provides useful summaries and discussion of some of the major philosophical options. Among the most prominent issues that arise in connection with the concept of function is the fact that it seems to make sense to attribute functions only in respect of certain kinds of object. We speak naturally of the function of a cup, or of the liver, but prima facie it seems odd to talk about the function of the moon[28]. And similarly pressing is the way in which functions are often not so much about what things straightforwardly do as about what they *ought* to do; hence something can operate incorrectly, or fail to perform its normal function. As Wouters notes, 'the main task of a theory of function is to explain how this norm arises in biological contexts' (2005, p.124). One option for explaining the normativity of function in the context of man-made artefacts is to appeal to the intentions of a designer, and it is suggested by some, e.g. Millikan (1984), that in the biological realm the role of designer can be fulfilled by natural selection. Subtly different from these 'selected effect' accounts are so-called 'consequence etiology' accounts of function (e.g. Wright 1973), that see both artefacts and organismic parts as being present because of their biologically advantageous consequences. Boorse (1976) frames his account in terms of goal-directedness: things have functions inasmuch as they contribute towards the attainment of particular goals. Meanwhile Cummins (1975) develops an account around the idea that functions are related to the causal roles objects play within particular systemic contexts.

Wouters himself argues that in attempting to derive an account of function relevant to biological contexts we should pay attention to scientific practice. He notes that biologists make extensive use of functional terminology, but rarely does such use relate to diachronic matters like evolutionary origins. This suggests that we might be well-advised to direct our

[28] But it would probably not seem so if we were to use the moon as a component of some astronomic contraption for achieving specific aims, e.g. for hurling a spacecraft out into space by harnessing the moon's gravitational field.

efforts towards explicating and utilizing something like a Cummins-style account, by attending to what it is for something to play a certain causal role within a system. This is the second of the four senses of function Wouters describes in his (2003) (his function$_2$) and is the kind of account of function that I shall adopt in what follows. At times, however, I will also have recourse to his first sense (his function$_1$), which is the idea of function as activity. Unless otherwise stated it can be assumed that I mean function as causal role.

One reason why functional attribution is problematic is that functions are in general underdetermined by structural and causal/mechanical facts. A car factory, for example, can be seen straightforwardly as having the function of making cars, of course. But for certain purposes it may make as much sense to see its function as being to provide jobs for large numbers of workers, and that is a role it does indeed also fulfil. (One can well imagine a politician, for example, seeing it in just such terms.) The functions of man-made artefacts often appear to be interest- or purpose-relative in this way, which perhaps bodes ill for attempts to provide an account of function solely in terms of the objective features of objects or processes in the world. Indeed maybe this subjective aspect of function attribution is an important feature and something to be incorporated into the account we give of functions and functional explanation, rather than something to be factored out or bypassed.

When we attribute a function to an artefactual object or process we are implicitly saying something about what could replace it without detriment to the fulfilment of some specific goal or ongoing purpose, or the occurrence of some activity. To continue with the car factory example, if our interest is in ensuring employment for large numbers of people, a functional alternative to the factory – albeit perhaps not a terribly satisfactory one – might be a shopping complex. If the goal is to consume lots of steel then the car factory could be replaced by a tractor factory, or a ship-yard. Ascribing a function to something also enables us to think about how to improve its performance relative to that function.

Discussions of function in biological contexts often focus on the heart as an example. It seems straightforwardly the case that the function of the heart is to pump blood. But the beating of a heart means that as well as pumping blood it produces sound. What is it that makes pumping blood the function of the heart and not the production of sound? Again, thinking in counterfactual replacement terms is helpful: it forces us to think about what activity it is that the heart provides for that means that its possessor is able to

live a biologically normal existence. If the heart were replaced by a device for producing heart sounds – perhaps by playing a recording – then this condition would not be met (unless the production of the sound happened to be accompanied by the pumping of blood as a side-effect). If the heart were replaced by an artificial device for pumping blood then the condition would be met, even if the device were silent. (By stipulating that an organism with a heart replacement is able to live a biologically normal existence, I mean that it would live the existence it would live in the absence of heart replacement, assuming that its original heart were typical of those of other members of its species.[29])

It might be felt that this counterfactual way of looking at function has little connection with a Cummins-style role-in-a-system interpretation. Yet there is common ground inasmuch as there is a focus on the causal possibilities that obtain in a given context. Function attribution can be seen as relating specific structures or processes to broad classes of counterfactual structures or processes, all of which would ensure that certain goals are capable of being met given a certain context. And goal accomplishment is about bringing things about, which relates to acts, activities and states of affairs (with connotations of Wouters' function$_1$). (But note that it can be a goal to ensure that things remain as they are, in which case what is to be brought about is continuity or maintenance of the status quo.) A goal can be accomplished if an event occurs by way of an act or if an activity is performed, perhaps on an ongoing basis, or if a state of affairs is brought into being or perpetuated. Part of what functional attributions do is allow us to think about causation in structurally (physically) non-specific terms. Knowing that something fulfils a particular function provides reassurance about the fulfilment of causal requirements within a particular context, whilst allowing for what could be called 'explanation with abstraction'. The parts of an artefact set structural and functional constraints on each other. If an artefact part P has function F, then the structures of the parts that P interacts with will set constraints on P's structure or any functional replacement for P. But recall that I said that by attributing functions we associate structures with classes of counterfactual structure. Many man-made systems are composed of structural and functional modules that interact via standardized interfaces of various sorts. A replacement module need only respect the relevant interface standards for system function to be maintained. Hierarchical organization means that if a module contains sub-modules, then a replacement module need not respect any particular organization in respect of its sub-modules, so long as it respects the standards relevant to interfacing with other modules. Two functionally equivalent artefacts

[29] There are no doubt still issues here that would merit further discussion in a fuller and more detailed treatment, but this gloss I believe serves present purposes adequately.

may have quite different internal architectures, provided that they end up presenting the same functional capacities to the host system or functional context.[30]

In general then I shall, as I say, adopt the view that functions equate to causal roles within systems. Typically the fulfilment of such a role is associated with the performance of acts and activities of various kinds. More specifically, fulfilling a function depends on provision of the causal capacities that underpin particular acts and activities in a given context. Sometimes it is possible to identify the occurrence of particular acts or the performance of particular activities with goal attainment. However, we tend to attribute goal possession and attainment to only certain sorts of causal system, and to identify what it is that makes for such systems is not my immediate concern.

If we think back to the example of the clock then it can be argued that there is a certain perspective we adopt when we are confronted by such a mechanism and wish to understand how it works, relative to our knowledge that the clock is intended to display the time by appropriate variation of the positions of its hands. If asked what we think the function of a certain part of the clock is, we examine the physical structure of the clock and observe it in operation. Through mechanical reasoning we are generally able to hazard a good guess as to what it is that a part accomplishes in material cause/effect terms. Thus we might say that the function of the chime wheel is to raise the clapper arm so that the bell is periodically struck. This mechanical form of reasoning comes very naturally to us, and is interesting in its own right, but I shall not reflect further on it just yet. For now I shall be content to note merely that where material mechanisms – artefacts, machines – are concerned it is possible to think of the functions of parts in physical cause/effect terms in the overall context of the mechanism's 'normal' operation. Almost invariably, however, even mechanical reasoning takes place in the context of orienting knowledge (or guesswork) about overall artefact function, i.e. about the activity or overall ends to which the artefact's structures contribute.

A causal conception of mechanism

The machine conception of mechanism can be contrasted with a looser and more abstract sense that I shall call the causal conception. This pertains to the second dictionary

[30] Structural diversity in biology can be related to the idea that natural selection acts at the functional level and is blind to structure (Rosenberg 2001).

definition – which stated that a mechanism is 'a process by which something takes place or is brought about'. The range of phenomena this notion comprehends is vast; potentially it might be thought to include all the processes and events in the world to which the notion of causation might be applied. In principle it would therefore seem to include singular causal events and chains of events. Shortly I shall show how in practice a range of factors serve to narrow down substantially the domain of applicability. Consider the kind of singular causal process that we often recount in narrative terms and of which the following is an example:

> 'The man got on the bus at High Street at 5pm, but the traffic going out of town was slow owing to an accident that had occurred earlier in the day. By the time the bus reached the ring-road two miles away the supermarket had shut. As the bus the man had caught was the last bus of the day he had to walk home and, because he had been unable to buy food, he had to have beans on toast for supper.'

Such a narrative answers numerous potential questions. When did the man catch the bus? Why was the traffic slow? Why did he have beans on toast? In interpreting the narrative we make lots of assumptions, on the basis of common sense as well as through the application of more specific cultural knowledge. We know that most shops close in the late afternoon (in the United Kingdom at least – cultural context presumably matters), so it seems unsurprising that he was unable to buy food once the supermarket had shut – presumably other shops would have been shut at the time in question. Traffic accidents often do disrupt the flow of traffic. And so on. A picture of a series of entailments is created that explicates how events unfold in space and time, and this is what mechanisms in the loose sense of causal process are about.

We would probably not describe the pattern of causal connections sketched in the narrative as constituting a mechanism, but it is useful to reflect on why this should be. The most significant factor, I suggest, is the role we assume to have been played by chance in structuring the events the narrative describes. It was presumably just a coincidence that an accident occurred earlier on the day the man decided to catch the bus to the supermarket, and the fact that the traffic out of town was slow, whilst connected with that in a causally significant way, might also have involved a degree of chance. Had the traffic been lighter perhaps there would have been less of the congestion that we presume resulted from the bottleneck created by events related to the accident and that retarded the progress of the bus. Another factor besides chance that is of importance in guiding our mechanism-attributing tendencies may be the length of the causal event chains that connect events. If

they are so long that we cannot imagine – without making assumptions about large numbers of intervening events that we feel unjustified in making – how they could be connected, then (I conjecture) we tend to regard them as coincidental or down to chance.

How would we feel about applying the term 'mechanism' if the pattern of events described in the example happened every week, perhaps to multiple people in different locations? Perhaps then we could speak, without raising too many eyebrows, of a mechanism for public transport-related baked bean consumption (or the failure of a supermarket-reaching mechanism). But in that case we would presumably infer the existence of additional causal factors to account for the repetition of the same event patterns at different locations. In other words we would infer the presence of an underlying causal process, and its discovery would go some way towards legitimizing mechanism talk. If in even those circumstances we were to baulk at use of the term then this would point to the involvement of issues of social convention in its deployment (above and beyond the basic causal facts). The status of singular causal processes in the causal conception of mechanism is rather uncertain, then. But I suspect that where a singular causal process appears not to involve a strong element of chance it is for pragmatic reasons rather than substantive ones that we would not apply the term mechanism to it. When we talk about singular causal events and processes it is usually precisely their singularity, the specifics of the case, to which our attention is being drawn.

More abstract examples of mechanism in the sense of the establishment of entailment can be given. For example, the setting of interest rates constitutes the mechanism by which the Bank of England attempts to control inflation. Software applications consist of lines of code that can be thought of as logical mechanisms that act on data to transform or manipulate it in various ways. More biologically, but perhaps less mechanistically (I consider the point in the next chapter), the 'mechanism' by which protein molecules fold involves the aggregation of hydrophobic amino acid residues and the satisfaction of the polypeptide chain's hydrogen-bonding potentialities through the formation of hydrogen bonds either internally or with surrounding water molecules. These diverse uses of the term 'mechanism' all point to the existence of, or possibility for the existence of, particular patterns of implication, entailment or causation that serve to define – or could be used to instantiate – some kind of process, event or operation. In Chapter 6 I develop, in the context of a discussion of emergence, the idea that we recognise these

patterns by paralleling them in our minds.[31] We simulate bits of the world, and see entailment or causal structure, a causal 'mechanism', when our thoughts are able to stay in step with, retrodict or anticipate events in the world. Roughly speaking, then, the causal conception of mechanism relates to the accessibility of a pattern of cognitive entailments that connects the real or imaginary event or phenomenon that the mechanism is 'for' with real or imaginary antecedent events, conditions or states. Such a model adverts to the existence of a parallel system of entailment in the world, or the possibility of one, and where the scientific study of nature is concerned a physical system exhibiting a particular causal structure is the ontologically robust 'mechanism' indicated by use of the term.[32]

The neo-mechanistic perspective

Having briefly considered the material and causal conceptions of mechanism I now turn my attention to what it is that biologists mean when they speak about mechanisms, as in the examples I gave in the introduction to this chapter. On the face of it the machine conception would seem to have little to offer biologists, since they are generally not dealing with largely solid-state devices. Does that mean that mechanism talk in biology appeals to something like the second sense? In their 2000 paper ('Thinking About Mechanisms', or MDC as I shall refer to it), Peter Machamer, Lindley Darden and Carl Craver attempt to explicate what is going on when biologists talk about mechanisms. That paper has shaped much of the subsequent philosophical debate about mechanisms and explanation in biology. They begin with the assertion that in 'many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism. So it is not surprising that much of the practice of science can be understood in terms of the discovery and description of mechanisms' (MDC pp.1-2). They state their goal as being 'to sketch a mechanistic approach for analyzing neurobiology and molecular biology that is grounded in the details of scientific practice, an approach that may well apply to other scientific fields' (MDC p.2). That the aim is described as being 'to sketch a mechanistic approach' raises the issue of the distinction between mechanistic ontology – mechanisms as things in the world – and mechanistic epistemology. It sounds here as though mechanism is regarded by MDC

[31] An early statement of this sort of idea is found in Craik (1943/1967).

[32] The identification of a mechanistic pattern of entailments in the world can be thought of as an example of inference to the best explanation (Lipton 2004) – provided that what is considered to be 'best' is understood to be context-relative and contingent on what we already believe. 'Most adequate' might be a preferable phrase, but that is not to say that scientific explanations don't typically connect deeply with the phenomena they account for. (They don't just model surface phenomenology, in other words; they are dependent on postulated counterfactual-supporting structures and properties that underlie what occurs – as I discuss in Chapter 3 in relation to the simulation of protein dynamics.)

primarily as part of the explanatory toolkit we can bring to the world in order to make sense of it, although talking of 'the discovery and description of mechanisms' makes mechanism sound more ontologically grounded.[33] In fact their position has both ontic and epistemic aspects.

To motivate the case for their account MDC provide quantitative evidence regarding use of the term 'mechanism':

> Mechanisms have been invoked many times and places in philosophy and science. A key word search on 'mechanism' for 1992-1997 in titles and abstracts of *Nature* (including its subsidiary journals, such as *Nature Genetics*) found 597 hits. A search in the *Philosophers' Index* for the same period found 205 hits. Yet, in our view, there is no adequate analysis of what mechanisms are and how they work in science.
>
> (MDC p.2)

The aim in quoting these statistics is presumably to show that scientists and philosophers alike make extensive use of the term. The numbers given fail to make this point, however, for no information is given regarding how the numbers of hits for 'mechanism' in the corpora mentioned compares with the numbers of hits for other terms in the same corpora. Neither is it possible to judge whether a certain number of hits for a term in a specific corpus should be judged high or low unless one knows how frequently the same term occurs in a larger, less discipline-specific corpus. But this is a methodological gripe. I think MDC are right to claim that scientists make extensive use of the term 'mechanism', and I agree with them that the task of making sense of scientific explanation will be easier if we can come to some understanding of what is meant by it.

MDC encapsulate their view of mechanism in the following definition:

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.
>
> (MDC p.3)

Amongst the examples they discuss from molecular biology and neurobiology is the case of DNA replication:

[33] The phrase 'a mechanistic approach for analyzing neurobiology and molecular biology' is unfortunate, since it is ambiguous about whether the analysis in question is of the subject matter of those disciplines or of the conceptual content of the disciplines themselves. I suspect that MDC would defend both interpretations.

> In the mechanism of DNA replication, the DNA double helix unwinds, exposing slightly charged bases to which complementary bases bond, producing, after several more stages, two duplicate helices. Descriptions of mechanisms show how the termination conditions are produced by the set-up conditions and intermediate stages. To give a description of a mechanism for a phenomenon is to explain that phenomenon, i.e., to explain how it was produced.
>
> (MDC p.3)

This kind of example provides some sense of what MDC mean by set-up and termination conditions, but still it would be as well to try to spell it out. I take them to mean that such conditions are specifiable or visualizable material states of affairs that bracket or delimit, spatially and temporally, the phenomena of interest – a point to which I return later. The use of the concept of productivity is a distinctive feature of the account, but I think Torres is right to argue that it is best regarded as an idiosyncratic way of importing causal notions into the account by the back door (Torres 2009, p.242).

A central feature of the account is its commitment to the ontic dualism of entities and activities, and this is potentially both appealing and problematic. Some of the appeal may stem from a tendency to think of entities as corresponding to structures and activities to processes, and the idea that the account succeeds in combining structural and processual thinking. (Torres thinks that activities do correspond to processes (Torres 2009, p.241).) But in that case why do MDC not use the terms structures and processes? Why is novel terminology needed? Worry about this point makes it important to be able to state clearly what entities are and what activities are. Entities are described as 'things that engage in activities' (MDC p.3) and it seems reasonable to suppose that entities do correspond to structures or parts of structures.

At the molecular level this interpretation is straightforward enough, but it must be borne in mind that at the cellular level many 'structures' are actually highly dynamic. For example, the relatively stable appearance of the Golgi bodies and other forms of endoplasmic reticulum seen in electron micrographs obscures the fact that there is a constant cycling of membrane structures through the cell, inwards from the cell membrane and outwards to it. An activity associated with a structure in which the relationships amongst parts are constantly changing, but through which matter does not flow, is significantly different from one associated with an apparent structure maintained despite, or indeed by, an underlying turnover of matter. It is important not to lose sight of the distinctive thermodynamic character of cellular processes and one of the principal ways in which man-made machines differ from biological systems. If we do, the chance to gain

insight into particular methodological issues may slip away. The equal weighting MDC place on entities and activities should I think be understood as reflecting recognition of this very point – assuming that we can equate activities with processes. Entity/activity dualism represents an acknowledgement of the difficulty we sometimes have in assigning causal priority, and hence making an ontic commitment, to just structures or just processes.

Activities are described as 'the producers of change; they are constitutive of the transformations that yield new states of affairs or new products' (MDC p.4). This sounds like a causal matter, but the avoidance of overt causal talk problematizes the nature of the relationship between activities and causes. MDC themselves say that activities are 'types of causes', but then add a little confusingly that '[a]n entity acts as a cause when it engages in a productive activity' (MDC p.6). But as I discussed earlier, in the context of a discussion of the machine conception of mechanism, we can think of functions as just the kinds of acts of doing that MDC appear to associate with activities. This is Wouters' (2003) function$_1$ sense. Appealing to Wouters' function$_2$ sense of function as causal role they also say that functions are 'the roles played by entities and activities in a mechanism. To see an activity as a function is to see it as a component in some mechanism, that is, to see it in a context that is taken to be important, vital, or otherwise significant' (MDC p.6). Overall, the way in which entities, activities, functions, and causes are interrelated in the account is far from clear. Of functions MDC go on to say that:

> It is common to speak of functions as properties 'had by' entities, as when one says that the heart 'has' the function of pumping blood or the channel 'has' the function of gating the flow of sodium. This way of speaking reinforces the substantivalist tendency against which we have been arguing. Functions, rather, should be understood in terms of the activities by virtue of which entities contribute to the workings of a mechanism. It is more appropriate to say that the function of the heart is to pump blood and thereby deliver (with the aid of the rest of the circulatory system) oxygen and nutrients to the rest of the body. Likewise, a function of sodium channels is to gate sodium current in the production of action potentials. To the extent that the activity of a mechanism as a whole contributes to something in a context that is taken to be antecedently important, vital, or otherwise significant, that activity too can be thought of as the (or a) function of the mechanism as a whole.
>
> (MDC p.6)

The point about avoiding speaking of functions as properties 'had' by entities, related to the 'substantivalist tendency', is well-taken. Moreover it is consistent with hints in the paper that to think of biological mechanisms in structures-with-functions terms is simple-minded. But other authors with whom MDC appear to be broadly sympathetic,

such as Bechtel, sometimes construe biological mechanisms in terms of a position on structure—function relationships that sounds distinctly machine-like:

> The part—whole relationship between a mechanism's component parts and its structure can be understood as falling within the type of hierarchical, mereological framework that systematic biologists and others have long used to bring orderliness to types of entities at different levels. The relationship between a mechanism's component operations and its overall function have roughly the same character ... . What is important here is that both kinds of components (the parts and their operations) can be regarded as occupying a lower level than the mechanism itself (a structure with a function).
>
> (Bechtel 2006, p.40)

In MDC it is asserted that

> Mechanisms occur in nested hierarchies and the description of mechanisms in neurobiology and molecular biology are frequently multi-level. The levels in these hierarchies should be thought of as part-whole hierarchies with the additional restriction that lower level entities, properties, and activities are components in mechanisms that produce higher level phenomena'.
>
> (MDC, p.13)

This kind of mereological view, in which mechanisms are seen as nested part-whole hierarchies, sounds like an unpromising way of addressing cellular phenomena that are often characterized by fluidity and the dynamic turnover of material noted earlier. But further reflection raises an opposing worry, that mechanisms seen in this way might encompass too much. The problem I think is that it is hard, and perhaps ultimately unrewarding, to think of nested hierarchies of entities-and-activities, which are the terms in which MDC encourage us to think about mechanisms. A nested hierarchy of structures has at least the merit of ready conceivability, but what is a hierarchy of entity/activity complexes *like*? Perhaps part of the difficulty lies in the capacity of the entities and activities making up complex biological systems to pay little heed to neat mereological distinctions. The binding of a single signalling molecule to a specific receptor can trigger a ramifying set of processes that may be widely distributed throughout the cell and beyond, for example. The fact that the cell incorporates structures and aggregates of structures, contained within bounded compartments, encourages levels talk, but it is prudent to regard the idea of levels as a schematic idealization or as a fiction that is only sometimes explanatorily helpful.

In terms of scope and looseness the MDC conception of mechanism sometimes looks little different from the causal conception I outlined earlier. And indeed, when scientists talk about cell signalling mechanisms, say, I think a simple causal interpretation is a major part of what underlies the ascription of mechanism. (There is a way in which signalling events are brought about, but no parallel with machines is intended.) But the emphasis in MDC on process regularity demonstrates that something narrower is meant. Again, however, the language is unclear. Recall that the main definition states that '[m]echanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (p.3). Thus it seems that it is the changes that mechanisms bring about that are regular. (Although the question arises: what sorts of change?) But then MDC also say that mechanisms are regular 'in that they work always or for the most part in the same way under the same conditions. The regularity is exhibited in the typical way that the mechanism runs from beginning to end' (p.3). Later, in comparing activities with laws, MDC describe a mechanism as 'the series of activities of entities that bring about the finish or termination conditions in a regular way' (p.7). The sense that regularity is an important constraint on the projectibility of mechanism concepts is thus conveyed, but quite what the relevant regularities are is far less clear.

Despite these sorts of interpretative difficulty, I think that it is generally clear that what MDC are trying to do is address the relationship between the causality of molecular and cell biological phenomena and our methods for explaining them. They seek to do so in a way that reflects their distinctive spatiotemporal characteristics and their materiality, and which makes no appeal to strict nomic regularities. The problem facing anyone attempting to describe and explain cellular processes is how to talk about dynamic systems rather than mere static structures, and moreover systems that turn over their parts whilst retaining particular forms of process architecture. MDC pin their hopes on the flexibility conferred by thinking in terms of complexes of entities and activities. A danger, however, is that it is too easy to switch in an unprincipled way between entities and activities in order to reveal the mechanistic nature of a phenomenon. (If the status of entities looks problematic then trace out the mechanism through the activities, and vice versa: these do not appear to be obviously illegitimate strategies for mechanistic explication under an MDC-style approach.)

Torres has proposed a heavily modified version of the MDC account which, he claims, circumvents the difficulties associated with ontic dualism about entities and activities. These, he argues, attach especially to the latter: activities for MDC are reified causes, and the problem is that some molecular/cellular processes involve activities that

amount to *negative* causes.[34] He gives the example of neuronal long term potentiation (LTP), in which it is removal of blocking magnesium ions that enables (causes) calcium ions to diffuse throught the NMDA channel (Torres 2009, pp.242-247). MDC, Torres argues, conflate (through the notion of activities) *how* property changes are brought about with *what* it is that brings them about. The answer, he suggests, is a 'descriptivist' account in which activity verbs (including 'enabling' or 'allowing') explain by specifying how a property change is brought about without necessarily saying what does the bringing about. He defines mechanisms to be

> complex systems composed of entities organized in space and time such that (i) through engaging in activities they produce a phenomenon, and (ii) the activities in which the mechanism's entities engage are characterizable in interventionist terms of direct, invariant, change-relating generalizations.
>
> (2009, p.247)

This definition has much going for it, even if the phrase 'complex systems', imported from Glennan's account, is not especially helpful.[35] The idea that mechanisms bring phenomena about by engaging in activities that involve entities is intuitively satisfying. And the appeal to Woodward's interventionist account of causation (Woodward 2003a) circumvents the problems that would attend the invocation of universal laws. (In Woodward's account the causal relation is associated with the possibility of manipulating one variable in a system by adjusting the values of other variables. The relationships between the variables need not be lawful in a universal sense; rather they need hold only within bounds (Woodward 2003a, p.240).) However, it is not obvious that Torres' account out-performs the MDC account when it comes to explaining how mechanisms serve an explanatory purpose. For MDC, recall that mechanism description is constitutive of explanation – '[t]o give a description of a mechanism for a phenomenon is to explain that phenomenon, i.e., to explain how it was produced' (MDC, p.3).

Sometimes what are described are mechanism schemas, where a mechanism schema is 'a truncated abstract description of a mechanism that can be filled with descriptions of known component parts and activities'. MDC suggest that mechanism schemas play some of the roles traditionally ascribed to theories – they are 'discovered,

[34] This is also his reason for rejecting Glennan's interactionist account of mechanism (Glennan 1996, 2002). As Torres argues, in some mechanisms a causal role is played by absence of interaction.

[35] The 'complex systems' investigated by the 'sciences of complexity' are often exactly those that are regarded as least susceptible to mechanistic forms of understanding. (I discuss the complexity of biological systems in Chapters 4 and 5.) However, the phrase is arguably consistent with Torres' view of an activity as 'an occasion on which one or more entities brings about one or more property changes'. The danger with this view is that it threatens to result in most phenomena being construed as mechanisms.

evaluated, and revised in cycles as science proceeds. They are used to describe, predict, and explain phenomena, to design experiments, and to interpret experimental results' (MDC p.17). By way of example they discuss the discovery of the mechanism of protein synthesis, which they present as an illustration of the piecemeal discovery of a mechanism schema (MDC, pp.18-21).

The further discussion of some of the epistemic aspects of their account that follows their exposition of mechanism schemas is the most satisfactory part of MDC's paper. Especially suggestive is the discussion of intelligibility and explanation, which raises some intriguing ideas about explanation and understanding. What they propose is that mechanistic explanations render phenomena intelligible, and this is connected with 'showing how the phenomena might be produced' (p.21). This encapsulates their view that

> The understanding provided by a mechanistic explanation may be correct or incorrect. Either way, the explanation renders a phenomenon intelligible. Mechanism descriptions show *how possibly*, *how plausibly*, or *how actually* things work. Intelligibility arises not from an explanation's correctness, but rather from an elucidative relation between the explanans (the set-up conditions and intermediate entities and activities) and the explanandum'.
>
> (MDC, p.21; their italics)

Explanation is not, they argue, fundamentally a matter of regularity: rather, 'explanation involves revealing the *productive* relation. It is the unwinding, bonding, and breaking that explain protein synthesis ... . It is not the regularities that explain but the activities that sustain the regularities' (p.22; their italics). I take these passages to indicate that mechanistic explanation is a matter of making phenomena and their production conceivable or imaginable, even though they do not speak in those terms. This points to a connection with cognitive psychology, in which regard MDC consider briefly the sensory basis of intelligibility. This is not just a visual matter:

> But seeing is not our only means of access to activities. Importantly, our kinaesthetic and proprioceptive senses also provide us with experience of activities, e.g. pushing, pulling, and rotating. Emotional experiences also are likely experiential grounds of intelligibility for activities of attraction, repulsion, hydrophobicity, and hydrophilicity. These activities give meanings that are then extended to areas beyond primitive sense perception. The use of basic perceptual verbs, such as "see" or "show", are extended to wider forms of intelligibility, such as proof or demonstration.
>
> (MDC, p.22)

The link with emotional experience perhaps requires further explication (and clarification of what falls within scope) in order to amount to a convincing claim, but (as I argue in more depth in Chapter 6) the involvement of kinaesthetic and proprioceptive senses in the comprehension of phenomena rings true with psychological findings. And the implied centrality of the visual sense accords with the importance biologists place on visualizing phenomena and with their reliance on visual descriptions such as diagrams, images and visual metaphors – as seen in the next chapter in relation to protein folding.

Torres' modifications to the MDC account relate primarily to the ontological status of activities: he reinterprets the MDC notion in such a way 'that activity verbs are explanatory by virtue of their descriptive content, rather than by reifying activities as the causal components of mechanisms' (Torres 2009, p.249). This de-reification of MDC's activities is a useful change, and moreover it is not necessarily inconsistent with MDC's interesting epistemic ideas concerning intelligibility and understanding, if for example it is possible think of descriptions involving activity verbs (such as pushing or pulling) as internal cognitive stimuli of some sort for the senses that MDC posit as underpinning mechanistic interpretation. The change enables Torres to deal with negative (interaction-free) causation, as occurs in his example of the removal of magnesum ions from a channel to allow the flow of calcium ions. This is not, I take it, the sort of scenario we have any great difficulty imagining, and the chief philosophical difficulty (if I allow myself to speculate rather freely here) is to find a meta-language which appeals to robust concepts that enable us to relate the language we typically use to talk about these kinds of imaginable phenomena to their spatiotemporal and other physical characteristics.

If this thought has any basis then it seems to heighten the importance of remaining vigilant as regards the distinction between explanation in a broad sense that pertains to understanding (within a mind) and a narrower sense of explanations as vehicles for conveying and instilling such understanding (between minds). The interesting epistemic parts of the MDC account concern both senses, whereas Torres' philosophical disagreements arguably attach principally to the latter. But if verbal descriptions of causal structures, which might include the descriptive content of activity verbs canvassed by Torres, succeed in evoking thoughts that connect with sensory experience (and especially with visuospatial senses) then MDC claim that intelligibility is the likely result. Thus MDC and Torres do not disagree at the fundamental level of the epistemology of explanation.

MDC say very little about constraints on intelligibility, although they go so far as to state that what is intelligible is likely to be 'a product of the ontogenic and phylogenetic development of human beings in a world such as ours' (p.22). From this I take them to be sympathetic to the view that the mechanism ascriptions we make are relative to our psychological capacities and dispositions, and in this case it becomes interesting to consider the relationship between mechanism and complexity, say. This is not to claim that mechanism ascriptions have no basis in the world: as a fundamental metaphysical assumption I take it that phenomena involve entities engaged in processes that unfold in diverse ways over a wide range of spatial and temporal scales. But what looks complex, what looks like a process, and indeed what looks like a mechanism, depends on *our* natures as much as it does on the objective nature of the phenomena in question. I investigate biological complexity in Chapters 4 and 5, and for now merely note the absence from the MDC account of overt reflection on some of the key physical properties of the sorts of system that interest its authors. The fluidity of cellular phenomena is a striking characteristic, but they do not discuss it explicitly except when they note that mechanism schemata can be 'instantiated in biological wet-ware' (2000, p.17). A thought at this point is that by paying attention to these biophysical characteristics and to the complexity of cellular phenomena we might gain insight into constraints on the propensity to discern and delineate mechanisms. Then it might be possible to obtain a more nuanced and comprehensive perspective on biological understanding.

Varieties of molecular mechanism

I have argued that the MDC account of mechanism suggests an orientation towards the epistemology of mechanistic explanation and understanding in biology. At the same time, however, the ambiguities to which its distinctive ontological features give rise means that its scope – in terms of the fraction of biological phenomena it might cover – is somewhat unclear. Is explanation in molecular cell biology just mechanistic explanation? Is there is an explicable fraction of phenomena which corresponds to those that are mechanistic, and how large is any non-mechanistic fraction?

I have claimed that ideas about structure—function relationships are closely associated with traditional views about material mechanisms, yet they are largely absent from the neo-mechanist perspective. In this last section of the chapter I want to return to issues of structure and function, because they promise to provide an additional way of

thinking about mechanism, and moreover one which admits of a matter of degree of 'mechanism-likeness' or 'mechanisticity' inasmuch as functional decompositions can be more or less aligned with structural ones. By such means it may be possible to address the causal complexity and striking diversity of cellular phenomena, and the apparent success of science in providing rich conceptual frameworks that help to characterize the nature of that complexity and diversity. To that end I shall attempt to develop a perspective according to which a process may be described as being either more or less mechanistic (from an ontological point of view), and either more or less amenable to mechanistic explanation (epistemically speaking).

To begin it is worth acknowledging that what we are interested in where cellular properties, capacities and behaviours are concerned *is* causation. Understanding the cell is to a great extent about being able to answer questions such as 'What causes A?', 'What effect does B have?', 'How does C bring about D?', etc. The kinds of answer we give to such questions are shaped in important ways by the nature of the relationships that hold between structure and function. Sometimes, in relation to particular phenomena, it seems that we can identify particular structures with specific functions, while where other phenomena are concerned it is much harder to make straightforward associations. A structure may be associated with several functions, or a function may be associated with several structures – or it may be difficult to associate functions and structures at all. I suspect that it is on this kind of basis that some of the fundamental properties and behaviours of cells should not be thought of as mechanistic in any strong sense. And properties and behaviours that involve ambiguous or otherwise problematic structure—function relations are probably good candidates for being the ones we are most inclined to describe as non-mechanistic.

One approach to mechanism in cell biology, then, is to say that a phenomenon is mechanistic when it is possible to associate particular functions with the specific molecular and cellular structures it involves. This idea sets up a parallel with machine-type artefacts, in which as we saw earlier there exists just this kind of alignment of structure with function. However, it stops short of saying that cell mechanisms *are* machines. I have already stressed one significant respect in which cell processes are to be distinguished from machines: the fact that cell processes often take place in a more or less fluid phase of matter. This has important consequences. The interactions that take place between molecules are often quite specific, and interactional specificity provides an ontological basis for the establishment of specific causal networks that are to some extent functionally isolated and

independent of other networks. Specificity means that everything need not interact with everything else just as chance encounters dictate. In functional terms interaction networks may be thought roughly analogous to the parts of a machine-type artefact, but in structural terms they are quite different. The entities involved in a network need not all exist simultaneously, so a network can be partly virtual, an idealized conception derived by imaginatively integrating over time multiple causal steps; a network need have no definite or persistent morphology; networks may interpenetrate; networks may share entities, so that complex functional inter-relationships are formed; network entities can be replaced without functional disruption; and so on. These kinds of property make cell processes rooted in molecular interaction networks *machine-like* in just the rather abstract sense that particular structures can be identified to some degree with specific activities, capacities and functions, but it would be hard to mistake them for machines.

Many cell functions, on the other hand, depend not on these evanescent interaction networks defined over more or less transient populations of mobile molecules but rather on what biologists are increasingly referring to as 'molecular machines' (Morange 2006a). An influential 1998 special issue of the journal *Cell* brought together a number of papers emphasizing the machine-like nature of a variety of functionally specific protein complexes. Bruce Alberts, in his editorial overview of the issue, suggests that 'the entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines' (Alberts 1998, p.291; see also Reynolds 2007 regarding the factory analogy). He goes on to explain that it makes sense to view the large protein assemblies underlying certain cell functions as machines because 'like the machines invented by humans to deal efficiently with the macroscopic world, these protein assemblies contain highly coordinated moving parts. Within each protein assembly, intermolecular collisions are not only restricted to a small set of possibilities, but reaction C depends on reaction B, which in turn depends on reaction A – just as it would in a machine of our common experience'. (Alberts cites his own (1984) in relation to this idea.) These highly constrained causal relationships amongst localized parts do indeed sound highly machine-like. What is more, the parts appear to be identifiable with particular functions, supporting the idea that the use of strongly mechanistic terminology is underpinned or motivated by the possibility of identifying neat structure—function relationships.

A recent example from the literature illustrates some of these points. Michal Zolkiewski (2006) reviews what is known about the Clp ATPases, a class of 'protein machines' involved in protein degradation and disaggregation, all of which are members of

the so-called AAA+ superfamily of ATPases.[36] Different subsets of the Clp ATPases perform different functions. ClpA, ClpX and HsIU are involved in protein degradation, while ClpB acts as a disaggregator of aggregated polypeptides. The functional difference of ClpB correlates with a structural difference, while the functional similarities of the others is reflected in their possession of conserved AAA+ sequence modules and ATP-binding motifs. The ATPases form ring-shaped or cylindrical oligomers (usually hexamers) through the centre of which runs a hollow channel. Different ATPases contain different additional domains that confer particular properties, for example specifying cellular localization or substrate specificity. Proteins destined for degradation have a peptide tag added to their C-terminus, and this is recognized by the ATPases. ATP hydrolysis provides the energy needed to unwind and thread the polypeptide chain of the protein to be degraded through the Clp channel to associated peptidases that cleave the unfolded chain. Unstructured loop regions in the AAA+ modules appear to be involved in substrate binding – another example of 'parts' having particular functions. Mechanistic terminology appears to come readily to researchers working on macromolecular complexes like this: ClpX is described as working 'like a machine with a two-speed transmission', and a hexameric Clp ATPase as 'a 'six-cylinder' ATP-driven polypeptide-threading engine' (Zolkiewski 2006, p.1096).

This relatively recent talk about protein machines is perhaps the most overtly mechanist language seen in molecular and cell biology, and it appears to pick out properties of functional macromolecular complexes that bear comparison with the man-made artefacts discussed earlier. Common properties include comparative rigidity of structure, sufficient to confer specific molecular recognition capacities, and relative to which defined motions occur, e.g. via flexible hinge regions, and the association of distinct functions with different parts (often specific protein domains). The result is the reliable generation of highly constrained causal event sequences. Given the relative fluidity of the cellular milieu, the tight association of the functional sub-elements of a process that protein machines represent is, Alberts argues, biologically advantageous. Invoking what amounts to an entropic argument, he asks us to compare 'the speed and elegance of the machine that simultaneously replicates both strands of the DNA double helix ... with what could be achieved if each of the individual components (DNA polymerase, DNA helicase, DNA primas, sliding clamp) acted instead in an uncoordinated manner' (Alberts 1998, p.292).

[36] AAA+ comes from '<u>A</u>TPases <u>a</u>ssociated with various cellular <u>a</u>ctivities'.

Conclusions

In this chapter I have presented an initial analysis of the concept of mechanism, which begins by resolving it into several senses. First there is the machine conception exemplified by man-made artefacts and characterized by certain physical properties as well as by the more abstract idea of structure—function alignment. Then there is the looser causal conception, which is the sense that is often invoked when we say that A is a mechanism for bringing about or accomplishing B. Employment of mechanistic terminology in this way is connected with a capacity to describe how a pattern of entailment comes about. The machine and causal conceptions are related inasmuch as machines harness matter to reliably instantiate particular causal structures or patterns of entailment. An examination of the most influential recent neo-mechanist articulation of mechanism, the MDC account, leads me to doubt that it provides a sufficient basis for addressing satisfactorily the diversity of molecular and cell biological processes. However, it does appear to be promising in relation to the development of a cognitively informed perspective on explanation and understanding. Structure—function concepts appear likely to provide a useful adjunct or extension to neo-mechanist accounts, however, and may have a valuable part to play in shedding light on particular aspects of contemporary scientific practice.

Talk of protein machines shows how many cellular biological processes depend on functional structures that can be viewed as machine-like. These materially instantiate patterns of causality that are relatively independent of their typical contexts, and in doing so they exemplify the connection between the two dictionary definitions of mechanism with which the chapter began. In Chapters 4 and 5 I consider the diversity and complexity of cellular processes and discuss some of the perspectives that have been developed in response to the causal and other issues they raise. Further consideration in Chapter 5 of the structure—process distinctions leads me to propose a conceptualization of mechanism that addresses some of the most important issues head on. From the standpoint of this new perspective the MDC account is seen to be too simple and inexplicit to deal satisfactorily with those issues, even though it is shaped by recognition of their significance. First, however, I shall discuss protein folding and what it means to talk about the mechanism of protein folding.

3. Protein folding and mechanism schemas

Introduction

In the previous chapter I investigated concepts of mechanism. As well as reviewing a perspective that is particularly influential in contemporary philosophy of biology I outlined two broader senses. The first pertains to relatively stable material systems involving hierarchical structure—function relationships and constrained patterns of configurational change. The sorts of functional artefact we describe as machines, in which these properties result for the most part from the nature of the solid state, best exemplify this sense. This is the conception of mechanism to which contemporary talk of 'protein machines' was seen to appeal. The second broad sense relates more generally to causal processes and patterns. I suggested that in principle it takes in singular causal processes, but argued that our linguistic practices reflect a pragmatic concern to exclude certain sorts of event patterns – those that involve a significant element of what we take to be contingent interplay amongst events – from being treated as mechanisms. Thinking of mechanisms in general as causal structures serving particular ends brings out the overlap between the two senses: machine-type mechanisms materially implement specific patterns of entailment.

Now I explore the region between mechanistic and non-mechanistic phenomena by investigating the biophysically fundamental topic of protein folding. In particular I attempt to pin down what scientists mean when they talk about the 'mechanism' of protein folding. Reviewing the relevant science reveals how (consistent with what MDC say about intelligibility) increased scientific understanding can come about as much through progressive improvements in our abilities to visualize complex molecular processes as through more abstractly theoretical developments. These enhanced abilities are the result of the interplay of a number of parallel scientific trends and research directions. I discuss the growth and use of computational techniques for simulating and visualizing the molecular dynamics of peptides and proteins to gain insights into the folding process. These developments have been accompanied by evolving visual metaphors for thinking about the phenomena they address.

My investigation reveals that protein folding is not a causally straightforward process: neat structure—function relationships and determinate sequences of events are the exception rather than the rule. Each folding 'run' for a given polypeptide sequence is likely to follow a more or less unique trajectory. Despite this stochastic element, evidence suggests that in the context of a specific amino acid sequence particular residues sometimes play key roles in establishing local structure. One might say of such residues that they are structures with particular functions in relation to the folding of a specific protein, but with the proviso that in a different sequence context the same residues may have different folding functions. In general there is regularity of process only at a rather abstract level, then, and to suggest that protein folding is mechanistic in any strong sense can be considered tendentious. I argue that it should instead be viewed if not as non-mechanistic then as only schematically mechanistic (semi-mechanistic, perhaps). I unpack the notion of mechanism schematicity by equating it with the possibility of viewing a phenomenon as mechanistic at a particular level of abstraction. A mechanism schema is a pattern to which a set of phenomena conform, but which can be 'filled out' in different ways by different members of the set. Sometimes, however, schema filling amounts to little more than a pragmatic willingness to take for granted the existence of some process connecting antecedent and resultant system states.

Protein folding

Many of the activities and structures on which life depends are grounded in the properties of protein molecules. The chemical reactions of metabolism are catalyzed by proteinous enzymes, and proteins play a variety of structural roles within cells (where they form the cytoskeleton) and externally, where they make up the extra-cellular matrix. Proteins such as collagen and keratin fulfill important structural functions in a variety of body tissues. The antibodies of the immune systems are large proteins specialized for molecular recognition, and other proteins perform transport functions or carry out other specialist tasks. A protein molecule is a chain of amino acids, and to a first approximation one can say that its biological properties are determined by its shape and other chemical properties (all within a given context). Each kind of protein – corresponding to a particular amino acid sequence – has (again, to a first approximation) a particular structure, which is the result of a process in which following synthesis, by translation of mRNA resulting from the prior transcription of DNA, the polypeptide chain produced adopts a compact shape, the protein's so-called native conformation.

Perhaps surprisingly, work on protein folding has been little studied by philosophers of biology, whose attention has tended to be captured instead by the Central Dogma and the issues about information and causation that it raises (and which I discuss at length in Chapter 7).[37] Yet assumptions about protein folding played a key role in shaping molecular biology's early disciplinary identity, and the relative experimental intractability of the phenomenon has stimulated the development of a diverse range of investigative techniques. Study of these promises to provide insight into major aspects of scientific explanation and understanding. Here I aim to clarify the mechanistic status of protein folding, and reciprocally use the phenomenon to help refine the ideas about mechanism already discussed.[38]

In the earliest days of molecular biology it was not known what factors were responsible for determining a protein's structure (although the first X-ray diffraction patterns obtained from protein crystals implied at least that there was such a thing as a regular solid-state conformation (Ferry 1998, pp.91-94)). The sequencing of insulin by Fred Sanger and co-workers between 1949 and 1955 established that a protein is a chain of amino acids arranged in a specific sequence (Judson 1996, p.89)[39], which led to the proposal that

> no general conclusions can be drawn from these results concerning the general principles which govern the arrangement of the amino-acid residues in protein chains. In fact, it would seem more probably that there are no such principles, but that each protein has its own unique arrangement; an arrangement which endows it with its particular properties and specificities and fits it for the function that it performs in nature.
>
> (Sanger and Thompson 1953, p.371)

This conclusion has been described as a sine qua non of the rise of molecular biology (Stretton 2002, p.530), as it led the early molecular biologists to speculate that sequence alone is sufficient to specify a protein's structure. Enshrined in the so-called Sequence

[37] However, Michel Morange has discussed the parallel neglect and underestimation of structural biology in the historiography of molecular biology (Morange 2006a). He attributes it to three factors: (1) the steady pace of progress in structural work, (2) its technical complexity, and (3) the broader neglect of chemistry, with which it is continuous.

[38] The picture of protein structure and folding presented here is based on a wide range of sources, allied to earlier laboratory experience. Whitford (2005) provides a comprehensive modern introduction, while Dickerson and Geis (1969) and Schulz and Schirmer (1979) are dated but still valuable. Useful reviews and commentaries include Karplus (1997), Dobson (2003), Clark (2007) and Chen et al. (2008).

[39] Organic chemist Sir Robert Robinson thought it 'astounding' that proteins have such structure, because it was widely believed prior to Sanger's sequencing of insulin that proteins are statistical polymers (Brenner, in Wolpert and Richards (1988), p.100).

Hypothesis this idea, together with the Central Dogma, formed a distinctive quasi-theoretical foundation for the new discipline (Crick 1958). Brenner recalls the novelty of the idea:

> I can remember going to meetings where people said 'Well, there'll be genes for folding up proteins' and so on, and we had just this remarkably simple hypothesis, the sequence hypothesis, which said that all you had to do was to specify the amino acid sequence and the folding would look after itself, and the energy would look after itself, and everything would be all right.
>
> (Brenner, in Wolpert and Richards (1988), pp.100-101)

Evidence to support the Sequence Hypothesis had come in the 1950s when Christian Anfinsen found that ribonuclease, a small enzyme, can be made to unfold and refold in vitro. He found that refolding proceeded spontaneously once the conditions that induce denaturation (loss of native conformation through unfolding) – such as variation in pH or temperature, or the presence of specific bond-disrupting agents – were lifted, in a manner compatible with the idea that the 'native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment' (Anfinsen 1972, p.56).

The first protein structures to be elucidated, by X-ray crystallographic methods, provided important clues about the principles on which protein structure might be said to depend. A common feature of the structures of myoglobin and haemoglobin, studied by Kendrew and Perutz respectively (at Cambridge), and to a slighter lesser extent lysozyme, studied by David Phillips at the Royal Institution, is the presence of extensive alpha helical structure. Alpha helices are formed when a stretch of polypeptide adopts a spiral conformation through the formation of hydrogen bonds between the peptidyl amide group of residue *i* and the peptidyl hydroxyl group of residue *i*-4.[40] The sequence of amino acid residues is regarded as the primary structure of a protein, and the alpha helix represents one kind of secondary structural element.[41] The other main classes of secondary structure in proteins are beta sheets and loops or turns, and a protein's overall conformation is known as its tertiary structure.

Amongst the factors responsible for the structural complexity of proteins are the properties of the twenty commonly occurring amino acids that make them up. An amino

[40] Peptidyl means pertaining to the peptide bond.

[41] The structure of the alpha helix was predicted by Linus Pauling on the basis of model-building experiments, a feat that inspired Watson and Crick to approach the structure of DNA in a similar way (Olby 1974/1994, chapter 17; Judson 1996, pp.62-69 and pp.134-135).

acid consists of a carbon atom (the so-called alpha carbon atom, or Cα) to which are attached an amino group, a hydrogen atom, a variable moiety called the sidechain or R-group, and a carboxyl group. When amino acids are joined together to form a polypeptide chain the amino and carboxyl groups of adjacent residues combine to form a peptide bond (with the release of a molecule of water)[42]. Electrons are delocalized over the peptide bond with the result that it has a somewhat rigid, approximately planar structure. This means that the conformation of the 'main chain' of the polypeptide – the series of linked atoms that runs from one end to the other – is to a large extent defined by rotations around just two bonds per amino acid residue: the bond between the amide group and the alpha carbon (the angle of rotation of which is termed phi) and that between the alpha carbon and the carboxyl carbon (rotation angle psi) (but see Berkholz et al. 2009 for a qualification to this).

G.N. Ramachandran had the idea of representing the structure of a protein structure as a two-dimensional plot of phi against psi (Ramakrishnan and Ramachandran 1965; see also Kleywegt and Jones 1996, Ho et al. 2003). Using physical models in which atoms were represented as hard spheres he showed that steric (three-dimensional shape-related) constraints limit amino acid conformations to just particular values of phi and psi, or regions of the phi—psi (or Ramachandran) plot. Different amino acids are limited to different regions, according to the bulk of their side-chains. The amino acid residue with the smallest side-chain, glycine, can adopt almost any conformation, whereas a bulky residue such as tryptophan is much more constrained. Side-chains differ markedly in terms of properties other than mere bulk. Some are charged, positively or negatively, or are capable of forming hydrogen bonds. These residues are termed hydrophilic (water loving). Others lack the ability to interact with water molecules (which are polarized and can form hydrogen bonds), and are described as hydrophobic. The cysteine residue's side-chain, meanwhile, contains an –S-H group which is capable of forming a strong covalent bond with the –S-H group of another cysteine residue – a so-called disulphide bridge. Most side-chains are capable of adopting a range of conformations, constrained again by the physically feasible angles of rotation around particular bonds.

Most proteins consist of several hundred amino acid residues, each associated with a range of phi and psi rotational possibilities that may be compatible with large numbers of alternative side-chain conformations. These facts give rise to the so-called 'protein folding problem', which is how the protein reaches the native state given the astronomic number

[42] In the context of a polypeptide chain, each amino acid becomes an amino acid residue – the residue being what is left after formation of the peptide bond.

of conformational possibilities apparently open to it. Even allowing only 3 conformations per residue, a protein of 100 residues would be associated with 3^{100} possible conformations – a number greater than the number of protons in the universe. In 1969 Cyrus Levinthal noted the paradoxical contrast between this number and the fact that proteins are capable of folding spontaneously and rapidly (Levinthal 1969; see also Karplus 1997). He took this to be one argument for the widespread view that folding proceeds not through an exhaustive conformation-by-conformation sampling of conformations but via particular pathways.

Establishing the routes followed by polypeptide chains as they head towards the native conformation has involved the development and application of a range of often complex practical techniques and theoretical perspectives. These have been used to address an evolving set of questions. What is the nature of the unfolded state? What are the relative contributions made to folding by hydrogen bonding (internally, and with water molecules) and by hydrophobic effects? Does folding converge directly on the native conformation, or are intermediate conformations involved? Answering such questions was hampered for many years by the fact that structures were obtained in the solid state, by crystallography, whereas folding occurs in an aqueous environment. Over a period of several decades, however, nuclear magnetic resonance spectroscopy (NMR) and computer modelling evolved to a point where they could be used to address precisely these kinds of question (Wüthrich 1995). Knowledge of Anfinsen's solution studies of denaturation and renaturation, and the fact that thinking about folding and unfolding involves connecting an image of the folded structure of a protein with some conception of the unfolded polypeptide, means that protein scientists had known for many years that proteins have to be conceived as more than static structures. But X-ray crystallography depends on the presence in a protein crystal of numerous copies of the molecule, all of which are constrained in roughly the same ways, and it focuses attention on the averaged image of the solid-state structure that results. This structure is generally corroborated by, and is consistent with, numerous and diverse other empirical findings, and indeed frequently makes sense of them. Hence it would be foolhardy to call into question crystallographically derived protein structures simply on the grounds that the crystalline environment is unlike the fluid milieux in which proteins typically function. On the other hand it is probably true to say that the thinking of protein scientists was significantly biased in particular directions by the attainability only of a solid-state structure that could apparently be equated with the

native conformation.[43] No doubt thinking did to a large extent neglect the dynamic properties of proteins until practical and computational developments brought them within range of investigation (McCammon and Harvey 1989).

Views about protein folding have slowly evolved as additional evidence has been obtained, in the form of new structures (from crystallographic, NMR-derived or other experimental data) and results from computer modelling. Work carried out in the 1950s and 1960s suggested the importance of a 'hydrophobic effect' (Kauzmann 1959; Tanford 1978), and this led the interior of a protein to be seen as being akin to an oil drop. The idea is that hydrophobic residues become buried as a by-product of the optimization of the structure of the surrounding cage of water molecules, which as well as hydrogen bonding with each other form hydrogen bonds with the hydrophilic residues at the protein's surface. An aspect of this picture is that – to the extent that a protein is partly hydrophobic – a protein's structure is the result in part of the thermodynamically driven minimization of its surface-to-volume ratio. This can be explicated in terms of the entropic cost associated with the creation of a cage of relatively immobilized water molecules around the protein. (The smaller this cage, the lower the cost.) Debate continues today about the nature of the hydrophobic effect and its importance for folding. Early views also tended to interpret folding kinetics in terms of a simple two-state model of structure, featuring only unfolded and folded states. (Baldwin has described this as the 'classical view' (Baldwin 1995).) With more sophisticated kinetic measurements it became clear that folding kinetics are often more complex than can be accounted for by such a model.

The 'new view'

The contemporary view of folding begins with the idea that the unfolded state relates to an ensemble of conformations, populated according to relative thermodynamic stability.[44] Constant jostling of a polypeptide chain by molecules of the surrounding solvent oscillates its atoms and causes the reorientation of parts of the chain relative to other parts, through rotation about those covalent bonds that are free to rotate. The resultant writhing of the polypeptide represents a sampling of accessible conformational space (CS), which is

[43] David Phillips, who determined the structure of lysozyme, wrote that '[t]he period 1965-1975 may be described as the decade of the rigid macromolecule. Brass models of DNA and a variety of proteins dominated the scene and much of the thinking' (Phillips 1981, quoted in Karplus 1997).

[44] The relative occupancy of two states differing in potential energy by ΔE is given by the Boltzmann formula $N_U/N_L = \exp(-\Delta E/kT)$, where N_U/N_L is the ratio of the number of molecules in the higher-energy state to the number in the lower-energy state, T is temperature and k is the Boltzmann constant.

the ultra-high-dimensional abstract space defined by extrapolating the concept of the Ramachandran plot such that each rotatable bond corresponds to a distinct dimension. Every possible conformation of a protein then corresponds to a point in CS, and protein dynamics deals with motions through that space. The unfolded state can be thought of as a set of points in CS, and folding can be identified with the set of pathways that begin at those points and that converge on the small region corresponding to the native state (Powell 1989). These ideas gave rise in the 1990s to the 'new view', in which folding is understood in terms of a funnel-shaped potential energy 'landscape' that directs unfolded conformations to the native conformation (Baldwin 1995; Dill and Chan 1997; Karplus 1997; Matagne and Dobson 1998). What is still unclear is exactly how rough the landscape is (Chavez, Onuchic and Clementi 2004; Krivov and Karplus 2004). The kinetics of folding of some proteins is consistent with the existence of relatively stable folding intermediates (i.e. they exist for long enough to show up in spectroscopic experiments, for example), and some proteins fold via several major alternative pathways. Some pathways are fast folding routes, whilst others are slower, depending on the stability of the intermediate conformations along a pathway. In energy landscape terms, fast pathways are routes that avoid both high-energy hills (traversal of which would require the supply of sufficient kinetic energy to push the molecule over the hill, via relatively improbable high-energy collision events) and low-energy wells in which a conformation might become stuck (until kicked out of the well by similar chance high-energy events). Some fast-folding proteins are known that approximate the theoretical possibility in which folding is always 'downhill' – there being no stable intermediate conformations along the folding pathway to the native conformation (Dyer 2007; Li et al. 2009).

Increasingly there is evidence to suggest that many unfolded polypeptides retain a considerable amount of secondary structure, and in particular a significant amount of alpha helical structure. And some proteins appear quickly to assume a non-native conformation or set of conformations that nonetheless is almost as compact as the native conformation. It is thought that this represents the fast burial of hydrophobic residues, and the corresponding realization of energetically favourable interactions between hydrophilic residues and solvent molecules. Rapid 'hydrophobic collapse' is thought to yield what is termed a 'molten globule' state, which in a slower stage then rearranges to attain the native conformation (Ohgushi and Wada 1983; Dobson 1992).

Thus the current orthodoxy is that folding generally involves multiple pathways, in the context of an overall picture in which non-native states are funnelled towards the native

conformation largely through a progressive built up of structure based on the transient formation of elements of local structure (Ho and Dill, 2006; Jayachandran 2007; Mok et al. 2007). Particular residues may be key to the formation of these stabilizing 'flickering intermediates', for example through the formation of specific hydrogen bonds or clustering of hydrophobic sidechains (Ybe and Hecht 1996; Neuweiler, Doose and Sauer 2005), but there are few hard and fast rules. A residue that plays a critical role when it occurs at a certain sequence position in one polypeptide may play no such role in the context of another sequence. And the order in which particular elements of structure form will vary from one instance of a particular polypeptide sequence to another, depending on the way in which contingent thermal events steer the trajectory through CS. Those are the rough outlines of an account of how proteins fold, but there is not yet complete consensus. The status of hydrophobic effects I mentioned earlier, and there is debate too around the significance of mainchain interactions versus sidechains as folding determinants (Rose et al. 2006). Such debates are likely to be resolved – perhaps more through a change in the emphasis placed on different terms in the account than through the development of an altogether new perspective – as more data are gathered from structure determinations, 'wet' experiments (especially those based on spectroscopic methods) and modelling.

Modelling and simulation

That there has been real progress over recent years in the computer modelling of protein structure is shown by the increasing success modellers have had in accurately predicting structures (Lovell and Papp 2005). The bi-annual CASP (Critical Assessment of Techniques for Protein Structure Prediction) initiative provides an opportunity for modellers to pit their methods against each other by using them to predict the structure of a protein whose structure has been determined crystallographically but not yet published (Tramontano 2006, pp.51-54). Strategies and techniques for structure prediction are many and varied, but a basic distinction is between ab initio methods, which attempt to compute structure on the basis of just the amino acid sequence and general physical principles, and knowledge-based methods that utilize knowledge of the structures of other proteins (Helles 2008, p.387). In recent years reasonably good results have been obtained using 'threading' methods, in which an attempt is made to fit the sequence of a protein of unknown structure to a known protein fold (Higgins and Taylor 2000, chapter 1). This kind of approach has become possible only with the availability of a large database of experimentally derived protein structures, and their success is contingent on the extent to which the set of known structures represents a comprehensive sampling of the folds that

occur in nature (Zhang 2009). Sometimes a combination of structure prediction methods is used, with an early example of such a hybrid approach being that of Martin et al. (1989). In recent years the 'Rosetta' methodology of the Baker group in particular has achieved some notable results (Das and Baker 2008; Raman et al. 2009).

Computer modelling of protein structure has its origins in multiple lines of research, but one important branch goes back to the development of computational methods for visualizing protein structures. X-ray crystallographers at first constructed physical models from their data, but this was a laborious process on account of the size of macromolecules and the unsuitability of the available model-building kits for macromolecular work (Platt 1960). As computer power and capability increased, and with the development of generic computer graphics techniques (Sutherland 1963; Foley and Van Dam 1982), it became feasible to generate and display images of protein structures from the computerized atomic coordinate data yielded by crystallography (Francoeur and Segal 2004). The Brookhaven Protein Databank (PDB) was established in the early 1970s as a central repository for storing and sharing protein structural data (which essentially consists of lists of atoms and their Cartesian coordinates) (Berman 2008).

The first interactive molecular graphics system was developed in the mid-1960s by Levinthal and co-workers at MIT (Levinthal 1966; Francoeur and Segal 2004), and within five years or so a variety of similar systems had been developed. Early systems were expensive, however, as the computations needed to view, rotate and manipulate these structures on-screen taxed even advanced computing resources. One especially widespread protein display system consisted of a software package called FRODO running on a high-end computing platform and usually outputting images via specialized graphics display hardware from Evans and Sutherland (Jones 1978). Technology developments led to the availability of increasingly affordable high-performance workstations, such as those from Sun Microsystems running the Unix operating system. By the 1990s it became possible to use powerful molecular display and modelling software, such as Biosym's Insight/Discover applications or similar packages from, for example, Polygen or Oxford Molecular, on standard hardware. These software applications often incorporated a variety of novel molecular modelling algorithms exported from the academic research environment.[45]

[45] For example, Oxford Molecular was formed in 1989 as a spin-off from the chemistry and biophysics laboratories at Oxford University, in part to commercialize software that had formed the basis of recent doctoral work (http://www.isis-innovation.com/spinout/index.html#pre1998 – last accessed 20 October 2009).

Interactive molecular graphics techniques can be thought of as an epistemic prosthesis that compensates for our limited ability to view a 2-D image of a molecular structure and infer from it, and keep in mind, a 3-D structure. By rotating a molecular model on screen in real time it is possible readily to form an impression of molecular shape on the basis of the relative motions of the atoms making up the molecule. However, if the molecule is represented simply by connecting the x, y coordinates of the atoms with lines to represent covalent bonds, with no cues to indicate relative 'depth' (i.e. position along the z axis), then the stereo geometry of the molecule and its direction of rotation are ambiguous. This is because the shape of the molecule is inferred from the relative motions of the atoms across the screen. If an atom moves to the left then this could mean either that the atom is at the front of the molecule and that the latter is rotating clockwise about the vertical (y) axis (if one were to look down that axis from above), or that the atom is at the rear of the molecule and the direction of rotation is anticlockwise about the vertical axis. Graphical indications of relative depth (depth cues), such as drawing closer bonds with thicker or brighter lines, overcome this ambiguity. If molecular motion is controlled via hardware, e.g. by rotating knobs, then the problem may not arise however – and the sense of interacting with a real object may be greater. Efforts to heighten this impression have involved utilizing immersive virtual reality techniques and 'haptic feedback' to provide researchers with simulated physical experience of interacting with molecular shapes and forces, although such techniques have not entered the molecular modelling mainstream (e.g. Cruz-Neira, Langley and Bash 1996; Francoeur and Segal 2004, p.422).

Computer modelling of protein structure involves more than just interactive visualization, however. Techniques for modifying and analysing structures, and for computing physico-chemical properties such as surface area, electrostatic potential and potential energy (e.g. Connolly 1983; Warwicker et al. 1985; Bash et al. 1987), provide insights into the properties of known structures and the likely properties of counterfactual structures. Such theoretical techniques also provide the foundation for a computational technique that now plays an important part in protein structural work and increasingly contributes to our understanding of folding: molecular dynamics (MD) simulation (Karplus and McCammon 2002). This is used to model, in what is intended to be a physically realistic manner, the behaviour of a molecule over time by iteratively solving Newton's equations of motion for each of its atoms. Sometimes the solvent is also explicitly represented in atomic terms, or its bulk properties may be represented by modifying various parameters. To compute the atomic displacements at each time step it is necessary to compute the overall force acting on each atom (Wang et al. 2001). This is achieved by

calculating the value of a potential energy function that includes a number of terms to represent different kinds of interaction and force. For example there are terms in the function to represent the electrostatic interactions between the atom in question and each other atom in the system, or the (much weaker, and shorter range) Van der Waals interactions between an atom and its near neighbours. Reduced to its essentials the MD algorithm is rather simple:

Bare-bones molecular dynamics algorithm

(Time t at start of simulation = 0 s)
Step 1: Calculate the overall force acting on each atom at time t as a function of the locations and identities of the other atoms in the system, in accordance with the specified potential energy function;
Step 2: Displace each atom according to the force acting on it;
Step 3: Increment t by the simulation time interval Δt;
Step 4: If $t < t_{end}$ then go to Step 1.

(t_{end} is the simulation duration – i.e. the amount of time to be simulated, not the actual time such a simulation might take to run.)

When building a model from crystallographic data it can usually be automatically determined whether two atoms are covalently bonded or not: if they are separated by less than a certain distance then covalent electron sharing can be presumed. Different approaches exist for representing hydrogen bonds, but often they are treated as a special kind of electrostatic interaction. Because MD techniques are so computationally demanding (especially as regards processor time) simulations of even short peptides – of 20 amino acid residues or so – can generally be run only over simulated timescales of the order of microseconds (Karplus and McCammon 2002, p.650). This compares with typical real folding times for proteins on the order of milliseconds to several minutes. Nonetheless it has been possible to shed light on important aspects of folding, for example concerning the formation of local structure, by use of MD simulations. Studies have necessarily tended to focus on small proteins and peptides, such as bovine pancreatic trypsin inhibitor (BPTI) (Brooks and Karplus 1983) and more recently the villin headpiece (Jayachandran et al. 2007), and often simulation work is paralleled by NMR and other spectroscopic investigations (e.g. Brewer et al. 2005).[46] The findings of in silico and in vitro work are complementary, and when the results from the one are consistent with those from the

[46] It might be interesting to compare the role of these 'model molecules' with that of model organisms in biological research more generally.

other confidence is gained that the picture we have of how proteins fold is on roughly the right tracks (Spörlein et al. 2002).

Computational trends mean that ever-greater timescales are coming within reach, and the development of techniques for splitting problems into chunks that can be parcelled out to multiple networked processors has had a further amplifying effect. (MD simulations performed in 1998 of peptides of around 40 amino acid residues in length covered simulated timescales of 800 ps (Smith et al. 1998), whereas in 2007 researchers were able to follow the dynamics of the villin headpiece, a 35-residue peptide, over a simulated timescale of 20 μs (Lei and Duan 2007). This represents a 25,000-fold increase in simulation timescale.[47]) The results that have been obtained using MD simulations, their approximate congruence with findings from 'wet' experimental techniques such as NMR, and increasing success in predicting protein folds on the basis of both sequence similarity with known structures and ab initio methods mean that for the past decade the protein structure community has generally been optimistic about prospects for the dissolution of the protein folding problem (Baker 2000; Wolynes 2004; Dodson 2007; Service 2008). The principal obstacle to the accurate prediction of structure from sequence appears to the computational one of sampling enough of conformational space at sufficiently high resolution (Raman et al. 2009, p.99).

Simulation, models and laws

It has been argued that simulation techniques raise few new issues for the philosopher of science (Frigg and Reiss 2009). Perhaps that is so, but whether novel issues are raised is a different matter from whether old issues are touched on in interesting and perspicuous ways. How might MD simulations influence how we think about physical laws? The nature and status of laws has attracted considerable philosophical attention over a number of decades, but the relevant debates have often had a somewhat abstracted character. Seeing laws in a particular scientific application setting raises the possibility of unearthing hitherto neglected issues and aspects – or may simply help to appreciate the salience of issues exposed by more traditional decontextualized reflection.

[47] These figures relate to different molecules and different computing setups, and are merely indicative of the sorts of performance improvements that have been made.

Two broad and perhaps contrasting points about laws are prominent in the context of MD simulations and the physical models that they are based on. The first relates to the fact that the laws seem to do real explanatory work: in large part it is the nature of the laws (in terms of their mathematical form) employed and coded into MD software that determines whether simulations generate virtual molecular behaviours that have any consonance with, and hence capacity to account for, ('wet') experimental data. I say 'in large part' since simulation accuracy is dependent too on the forms of molecular representation employed – for example as regards the treatment of atoms as point objects identifiable with specific locations in Cartesian space, or the way in which amino acid sidechains are represented. So perhaps it would be better to say that the laws used by current simulation techniques, such as the inverse square law governing electrostatic interactions between atoms, collectively appear – in tandem with the ways in which molecules are represented in virtual models – to succeed in allowing many of the qualitative aspects of the phenomena they are intended to model to be indeed modelled, along with some of their quantitative aspects. Replace an inverse square law with an inverse cube law and the simulation yields virtual behaviours that bear no relationship to the available empirical evidence, for example. On the face of things this seems grounds for presuming that the physical models that are 'run' in MD simulations of peptides and proteins in solution quite accurately mimic – at least within limits – the properties and behaviours of real physical systems involving those molecules.

The second point is that notwithstanding the apparent adequacy of the laws used, they are readily acknowledged by scientists not to be fundamental but rather to be approximations. Current scientific orthodoxy tells us that a physically thorough approach to molecular simulation would involve solving extremely complex quantum mechanical equations for the relevant molecular systems, which would be even more computationally demanding than present, approximate MD methods. (So demanding as to be, today and for the foreseeable future, unfeasible for all but the simplest systems.[48]) Thus the fact that MD simulations appear to 'work' can be seen as confirming a working hypothesis to the effect that it is sometimes possible to model macromolecular processes realistically despite having to make substantial approximations to the way in which physical phenomena are represented. This hypothesis appears to be justified in the case of many such processes, but not all. Important exceptions include the electron transport processes involved in

[48] New techniques for solving systems of QM equations for molecular systems have been developed that extend the range of simulation methods, however (see e.g. Leach (2001) and the special section 'Challenges of Theoretical Chemistry' which appeared in *Science*, 8 August 2008 issue (volume 321)).

respiration and photosynthesis. Additional support for this pragmatic confidence in the scientific validity of molecular simulation methods comes from the fact that as representations are made more detailed – for example by including hydrogen atoms in molecular models, or by reducing the size of simulation time steps – results model empirical findings with greater fidelity. This imparts a progressive character to simulation work which suggests that as computational resources become more powerful it will be possible not only to simulate molecular systems over longer timescales, but also to run more realistic simulations.

Recent philosophical work on physical laws has tended to take place in relation to debates about realism, and hence has involved a focus on theoretical truth. Nancy Cartwright has argued that there are no exceptionless laws, even in physics – all are subject to a ceteris paribus qualification (Cartwright 1983, pp.46-47). However, MD work provides a powerful reminder that the idea of exceptionless physical laws is not necessarily incompatible with the existence of a 'dappled world' (Cartwright 1999). In other words it shows how a complex observable phenomenal 'surface', which is often resistant to encapsulation in neat mathematical generalizations, can come about as a result of the interplay of the underlying laws we infer. In an MD simulation laws function as components in a mechanism for generating virtual phenomena. It is hard not to believe that whatever the 'real' properties of atoms and molecules are, they result in behaviours that are well-modelled (that is to say adequately, if not perfectly, modelled) by the laws we employ.

An interesting feature of Cartwright's work on laws and explanation is the way her realism about theoretical entities is combined with an anti-realist stance about theoretical laws. Her reasoning is that often we explain something by inferring the most probable cause, and that causal explanations typically involve existential claims about entities rather than just the invocation of theoretical laws (Cartwright 1983, 'When Explanation Leads to Inference', p.92). What is striking about MD simulations, however, is just how spare the theoretical entities can be, and how intimately they are tied to theoretical laws. The atoms of an MD simulation of a peptide or protein are just points in Cartesian space, each one nothing more than the origin for a particular potential field, e.g. of electrostatic potential. The simulation scientist heavily approximates both entities and laws, does so knowingly, and obtains results that are nonetheless explanatory.

Caution is still called for, however, and we should remember that MD simulations only work within the ultimate bounds set by their approximations. To model the electronic phenomena involved in chemical reactions we cannot avoid using quantum mechanical approaches. And which of the laws we infer and use in our simulations are not to be regarded as themselves phenomena of a sort, part of a pseudo-nomological 'surface' generated by events and processes that conform to more fundamental laws? When can we be sure we have struck solid nomological bedrock? These are the kinds of worry I take Cartwright to articulate, and they are well-motivated. But I would contest the claim of Frigg and Reiss that simulation science is philosophically unfertile territory. Understanding simulation techniques is beneficial because it helps us to appreciate the nature of such worries and gauge the severity of the potential epistemological constraints they point to.

Entropy and the stochastic exploration of state space

Protein folding can be viewed as a battle against entropy: out of all the possible structures that a polypeptide can adopt, one or two must be favoured over others if the functional capacity of the native conformation is to be fulfilled. Perhaps the most significant anti-entropic factor is the formation of the polypeptide chain in the first place. If a protein's constituent amino acids were not covalently linked then it would be overwhelmingly probable that they would diffuse apart, never to come together simultaneously in the right configuration to form a stable functional entity. Once the polypeptide chain exists, folding becomes a matter of ensuring that sufficient stabilizing interactions can occur to make remaining in a folded or partially folded state, once such a state is attained, favourable relative to returning to unfolded states. Proteins are presumably the size they are on account of the trade-off that must be achieved between there being enough stabilizing interactions in respect of one region of CS for there to be a functionally robust native state on the one hand (i.e. short peptides are too unstable), and ensuring that such a region is accessible from unfolded conformations in reasonable time (to avoid aggregation and misfolding, and to deliver the functionality afforded by the native conformation swiftly) on the other. The set of naturally occurring amino acids provides consistent backbone properties – enabling for example the formation of the hydrogen bonds that stabilize the alpha helix and beta sheet secondary structures – while providing a diverse set of side-chain properties, in terms of size and shape, hydrogen-bonding capabilities, electrostatic properties, and other chemical properties.

Complexity is a term used increasingly in a variety of scientific fields, although it can be quite difficult to explicate exactly what it means. In the context of protein folding complexity is associated with the vastness of a polypeptide's conformational space (CS). Each rotatable bond represents a degree of freedom, a set of alternative configurations specifiable by variation of a single structural parameter. But the fact that sampling of CS at physiological temperatures is adequate to the relatively rapid 'discovery' of the native conformation implies that the topography of CS is such that unfolded conformations are separated from the native conformation by (perhaps a succession of) only relatively small potential energy barriers. The native conformation is stable enough that once attained it is not readily and irreversibly left. So the complexity of protein folding involves the stochastic sampling of a giant state space, through the random Brownian motions that are inevitable in molecular systems at non-zero absolute temperatures. The dependence of protein folding on the existence of what might sometimes be regarded merely as thermodynamic background noise makes the process rather different from the kind of machine-type mechanisms discussed in the previous chapter. (Blomberg (2006) uses the suggestive phrase 'Brownian ratchet' to capture the way in which noise plays a part in driving biological processes in a particular direction.) Machines were characterized in terms of highly constrained patterns of configurational variation, consistent with the degrees of freedom associated with particular points, lines and planes or articulation defined by the shapes of stable parts. To operate effectively, these man-made solid-state machines are typically engineered to operate consistently irrespective of environmental conditions. Of course, some artefactual mechanisms sense aspects of their operating environment and adjust to it, but in such cases the aim is often to compensate for such sensed properties (through feedback mechanisms) so that the usual mode of operation remains effective. Protein folding, on the other hand, *depends* on thermal noise: it is incorporated as an essential element of the process.[49]

Mechanism schematicity and abstraction

How does what I have just said about protein folding relate to the earlier discussions of mechanism and causal processes? On the one hand there seem to be broad physical generalizations that capture much of the nature of protein folding. There is an overall process trajectory in the progression from unfolded to folded states, and a certain amount of regularity as regards the causal roles played by classes of secondary structure and

[49] Except, presumably, for fast-folding proteins for which folding is always energetically 'downhill'.

key residue. It is not, therefore, that it is completely impossible to assign folding function to particular kinds of structure. There is substantial reliability of process too (i.e. most tokens of a particular sequence type fold to yield the same structure in a given environment). These attributes are, I suggest, enough to justify saying that folding has at least a partially mechanistic character. On the other hand it is a process in which chance, in the guise of the 'random' thermal motion of atoms and molecules, plays a major part. This element of contingency, acting in concert with the astronomic number of degrees of freedom a polypeptide's structure represents, and the relatively limited variation in energetic stability across alternative configurations relative to those degrees of freedom (i.e. few sterically feasible conformations are not accessible at least occasionally at the energies available at physiological temperatures), gives rise to a high degree of process complexity. Any instance of a given polypeptide can reach the native conformation by way of a large number of routes, and it is difficult, if not frequently impossible, to assign specific functional roles in respect of folding to specific amino acid residues that hold good across an entire population of instances of a particular polypeptide. The relative lack of causal constraint (when contrasted with the very limited number of pathways through configurational space associated with the operation of a machine), and the difficulty we have in identifying simple structure—function relationships, makes folding look somewhat non-mechanistic. Hence I suggest we should think of protein folding as a *semi-mechanistic* process; it is a process poised somewhere between the machine and causal process senses of mechanism.

So the short answer to the question of what the phrase 'the mechanism of protein folding' means is that it is the process by which proteins fold. But at a detailed level, in terms of what a particular amino acid does, there may be no general facts of the matter that hold true across either (a) multiple folding runs of the same polypeptide sequence or (b) the folding processes of different polypeptide sequences. Nonetheless I have argued that there is sufficient determinacy of outcome in the case of any given polypeptide, and sufficient regularity amongst folding processes in general, to warrant placing a slightly stronger interpretation on mechanism. One way of further explicating this sense of mechanism involved in protein folding – the process by which they fold – is through a concept of mechanism schemas and mechanistic schematicity. I shall try to make clear what I mean.

Generally the phenomena studied by science represent instances of particular types or related sets of causal processes or events. As they are viewed at higher resolutions

and/or as the contextual scope is broadened the sets of which they are members become smaller. (All events presumbly have to be regarded as unique if the contextual bounds are set large enough, and if viewing resolution is sufficiently high.) But conversely when events and processes are viewed at a higher level of abstraction, by limiting the contextual bounds and / or by ignoring details visible only when the resolution exceeds a certain level, many phenomena appear to be repeated at different times or places or both. In addition, phenomena may be similar if not in their details then in terms of their effects in a particular context. We see patterns in phenomena and amongst events. So far as overall process causality goes, details may not matter. For example, a causally important event in a molecular setting might be the formation of a hydrogen bond between a residue at the surface of a protein and a water molecule. But the orientation of the hydrogen bond, and by implication perhaps the orientation of the water molecule, might not matter if the other water molecules that interact with it can adjust their orientations accordingly. Similarly, the exact identity of the amino acid may not matter. In such cases we might say that there is schematic similarity between two distinct cases, or that there is a shared mechanism schema. Here the schema is not very abstract, however, by which I mean something quite specific: there are not many ways of satisfying, or 'filling out', the schema. (That there are not many ways of filling the schema means that it is readily filled out. This might initially sound paradoxical – what are the chances of finding just those few ways of filling out such a schema? But the point is that the schema is highly constrained, and the constraints act as a stringent filter on our thoughts.)

A mechanism schema, then, is a counterfactual space standing for a class of causal processes that must be filled by specific events and processes for the process to be instantiated. The more abstract a schema is then the greater the number, and perhaps diversity, of possible counterfactual events and processes that could fill it. A schema that imposes a large number of constraints on filling events and processes (as conditions to be satisfied) may be judged to be less abstract, or it may be that the filling processes must be more complex. For example, we can readily conceive of a process in which entity A must cause entity B to move in a certain way via a mediating event or process of some kind, such that B moves as close to simultaneously with A as possible. If B must be displaced in the same direction as A, and by the same amount, then the least complex counterfactual process needed to complete the overall process by which A displaces B is clearly quite simple. If, however, B must amplify the displacement of A, and reverse its direction, the minimally complex coupling process must be more complex. If it does not matter when B moves, so long as it is only after A moves, then A and B are more strongly decoupled and

the counterfactual space of mediating processes is larger. Presumably there is an upper limit to the complexity of a mediating process, beyond which – as a contingent, epistemic matter – the features of the mediating process in some sense come to dwarf the events and processes it couples.

That is one relevant sense of abstraction relevant to talk of mechanism schemas. It focuses on mediation between two stages of a process. Another sense addresses the process stages that are coupled in a schema. It depends on our capacity to group kinds of antecedent event or process, and resultant ones. This may be on functional grounds, or causal grounds, or perhaps according to some other basis. But however it is done, the larger and more diverse the sets of antecedent and resultant events and processes, the more abstract the mechanism schemas they represent. (This 'process terminus'-oriented sense of mechanistic schema abstraction can be combined with the first, 'mediating process'-oriented sense, of course.)

A minimal schematic sense of mechanism is invoked when the term is used to describe an unknown causal process presumed to connect one state or set of conditions (the origin) with another (the destination). This use of the term amounts to little more than an affirmation of faith in materialism and determinism (i.e. belief that one state of affairs, materially speaking, will lead to another). Only where repetition of a phenomenon can be discerned without the need for much abstraction can we say that the phenomenon is mechanistic in a non-schematic sense. The greater the degree of abstraction required the more strongly we feel that a phenomenon is mechanistic only schematically.

Conclusions

What does the phrase 'mechanism of protein folding' mean? It cannot be taken to refer to anything that can be likened to a machine, if we take those to be material structures that implement particular patterns of physical entailment in order to fulfil some role within a broader context. In the protein folding process a polypeptide chain is the target, so to speak, of a variety of physicochemical principles (laws of electrostatic force, the tendency to form chemical bonds by sharing electrons, etc.), but these principles are not organized into any overall structure – qua diachronically stable spatial configuration – that stands comparison with the structure of a machine. (It is a mistake to attempt to associate them with any such structure at all.)

General statements can, however, be made about how folding occurs, in terms of hydrophobic effects, hydrogen bonding propensities and so on, and associated with these general statements are a variety of striking visual metaphors, such as 'energy landscape', 'potential energy well', and 'funnel'. Such topographical metaphors help us imaginatively fill the gap between unfolded and folded states despite the complex stochastic nature of the processes that we believe to occupy that gap. This I think can be seen as bearing out MDC's proposal that intelligibility is connected with 'showing how' some phenomenon or effect comes about. Molecular dynamics (MD) simulations appear capable of filling the gap in physically explicit detail from moment to moment, by paralleling algorithmically what we take physical laws to do in reality. The laws used in MD simulations appear to approximate closely the effects of the causal capacities we attribute to atomic and molecular structures on the basis of their behaviour in a wide range of contexts.

Minimally and schematically, then, the phrase 'the mechanism of protein folding' adverts to the existence of sets of causal processes that connect unfolded polypeptides (in general) with their folded variants, about which it is possible to make a variety of counterfactual-supporting generalizations.

4. Complexity and cellular causality - I

Introduction

This chapter and the next concern the nature of cellular causation in relation to the processes it involves. These are distinctive in a variety of ways, and the limited understanding we have of many of the fundamental phenomena of cell biology implies that this distinctiveness constitutes a significant barrier to understanding. I have already touched on some of the characteristics of cell processes in relation to mechanistic concepts (for example when I mentioned interaction networks in a fluid environment) and I now examine them more closely, with a view to providing some account of how our understanding is shaped and constrained by specific issues.

The epistemic deficit we encounter when faced with the cell seems in large part to be connected to the *complexity* of cell processes, but how is that characteristic to be understood? Perhaps it is unsurprising that attempts to capture the nature of complexity often appear partial or unsatisfactory in some respect, for example by being so abstruse that the connections with specific scientific, and especially biological, problems are unclear (e.g. Rosen 1991; Darley 1994; Jones 2008). A corollary of one view of the complexity of cell processes is that no simple unifying account of them can be given. However, by looking at a range of cell properties and processes and by considering several perspectives on them I hope to shed some light on what it means to say that cells are complex and also clarify how the issue of complexity relates to mechanism. (To what extent, if any, does it make sense to think of cells as mechanisms?)

My approach consists in part of demonstrating how, depending on where our interests lie, a number of rather abstract ideas are potentially relevant to our understanding of cellular causality. The other part of my approach is to reflect on the dependence of cell function on a variety of different kinds of causal factor. These I describe in terms of the structures they involve, and the way in which those structures participate in the realization of particular functional outcomes. They range from the highly localized and machine-like at one extreme to, at the other, highly delocalized processes in which the functional capacities that contribute towards particular phenomenal outcomes or patterns are distributed across

multiple structures (if specific capacities or sets of structures can be identified at all). Some functions are implemented by relatively stable structures while others are instantiated fleetingly by highly dynamic or short-lifespan structures. Processes that are relatively insensitive to their environment can be contrasted with others that are more heavily modulated by contextual factors.

Protein machines, which can be thought of as a kind of molecular gadget, I discussed in Chapter 2. As a causal factor in the processes in which they participate – which typically are chemical reactions – they act rather differently from the highly dynamic and rather labile structures of the cytoskeleton that are important causal factors in a broad range of cell functions and processes. These two cases are different again from solvent effects or macromolecular crowding, say, which are non-specific and collective in character (i.e. involve properties of large molecular ensembles that are to some extent independent of the exact structures of the molecules), and from the way in which DNA and RNA function as passive molecular templates. I suggest that the cell is epistemically intractable in part because of this diversity of causal factors, which often act simultaneously to bring about particular events, processes or outcomes. But causal diversity does not necessarily beget epistemic intractability, even if it can be an important contributory factor. Intractability may also stem from the nature of any logical or functional architecture exhibited by cell processes, supervening on events involving molecular structures and their transformations. (Or perhaps, to keep an open mind, the problem relates rather to the lack of such architecture.)

It could be argued that an implication of the picture I present of the complexity of cellular causality, expressed in terms of the cognitively taxing interplay of diverse causal factors organized in functionally opaque ways, is that there are many routes into the topic. There would be a certain historical logic to beginning with the Central Dogma and the genocentric perspective on cellular causation to which it gave rise, which is a perspective now frequently considered misleadingly simplistic. In fact, however, I will only touch on the topic of gene action in the next chapter, and address it in depth only in Chapter 7. Instead my starting point is to review several related and influential accounts of complexity that touch on issues pertinent to biological systems. From the platform provided by this contemporary understanding of complexity I then look at recent work on the modelling of metabolism. This connects complexity with the topic of emergence (which I discuss at length in Chapter 6), and leads into a broader discussion of fundamental causal issues surrounding metabolism. The next chapter continues the discussion by describing some of

the biophysical and biomolecular specifics underpinning a variety of cellular processes. I argue that problems centring on the structure—process distinction, which the MDC account of mechanism tries to address through the ontic duality of entities and activities, remain pressing. In the final part of the next chapter I sketch a perspective on mechanism that seeks to avoid privileging either structure or process, whilst at the same time gratifying intuitions about the status of interactions and functions, for example. In addition I attempt to be explicit about the distinction between ontic and epistemic issues, and comparison with the MDC account suggests that some of that account's problems stem from its conflation of the two categories. To begin, however, I want to motivate the case for a broader perspective on cellular processes than appears often to be countenanced by neo-mechanists.

Structures, processes and material flux

The two broad senses of mechanism I outlined in Chapter 2 can be combined and situated within a rough and ready framework for thinking about material causal processes in general. At one extreme are systems involving relatively stable, persistent structures associated with discrete functions. Structural and functional organization tends to be modular and hierarchical, and the structure—function relationships are straightforward and robust. Causal influence in such systems typically acts in a localized and concentrated way, operates relatively independently of context, and conforms to simple schemas that are usually linear or cyclical. Such patterns of causation are in general readily described, analysed and comprehended. The operation of such systems may be so simple – if not globally then certainly within functionally significant local pockets – as to be expressible by way of simple mathematical functions involving few variables. These are the sorts of system we tend to think of as being mechanisms in a strong and relatively non-figurative sense. So far as an overall framework of causal processes goes we can think of these systems as implementing strongly mechanistic causal processes.

At another extreme lie quite different kinds of causal process, in which it is hard to associate structures reliably with particular functions, or functions with specific, stable structures. Structure—function relationships appear neither simple nor robust, and causal

influences are complex, often being distributed, collective or delocalized in nature.[50] It is generally much harder to describe, comprehend or analyse what is going on in these kinds of non-mechanistic systems, in which phenomena and behaviours sometimes occur unexpectedly, as if 'out of nowhere'. These kinds of phenomena, and the systems that give rise to them, are often described as emergent or as exhibiting emergent properties. I want to stress that this framework for thinking about material causal processes is highly approximate and is not to be thought of as a simple one-dimensional spectrum of possibilities. I have just described two extreme kinds of causal process, but they are not simple polar opposites, for they are functions of a number of variables (i.e. those relating to the characteristics listed).

Another way of thinking about material systems, some of which we tend to see in mechanistic terms, is to focus on (non-causal) spatiotemporal correlations between the particles that make them up. (For purposes of argument it could be assumed that the particles are the nuclei of the atoms making up the system, or the small groups of nuclei that co-occur in particular molecules.) Imagine that a virtual box defines the region we take to include the system of interest. If this virtual box contains a homogeneous fluid then the particles will be mobile and will move relatively independently. (Suppose that the fluid is a gas, so that the particles are not unduly constrained by their neighbours.) (See Figure 2A.) As a result, over a particular time interval Δt each particle P will come within a certain distance r of $N(r)$ other particles. For large values of Δt each particle will come within r of the same number $N(r)$ of particles. And suppose we number each particle P, and associate with each particle an 'encounter list' consisting of the numbers of all the particles that come within range r of P during Δt. As Δt increases the particles will tend to become associated with increasingly similar encounter lists.

The encounter list properties that hold for a system consisting of a homogeneous gas are very different from those generated when the virtual box is filled by a solid (Figure 2B[51]). Then, each particle is associated with a unique and fixed encounter list that is independent of Δt. If instead of a fluid or a solid the system consists of a solid wheel – of radius $2r$, say – immersed in fluid (see Figure 2C) then the encounter list characteristics are

[50] My former Egenis colleague Jonathan Davies has explored the idea of distributed causal explanation in his PhD thesis (Davies 2008). In Chapter 8 I reflect further on the differentiation of systems according to how functions are distributed.

[51] In Figure 2B the particles have been drawn larger relative to r as compared with Figures 2A and 2C, for ease of preparation.

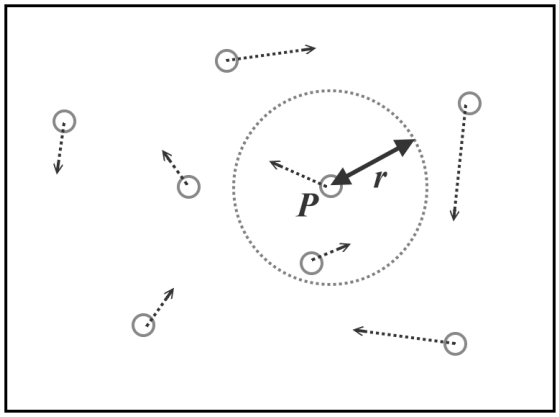


Figure 2A – Tracking particles: homogeneous fluid

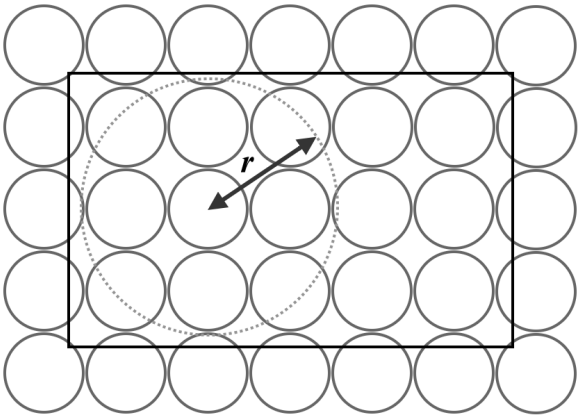


Figure 2B – Tracking particles: solid-state system

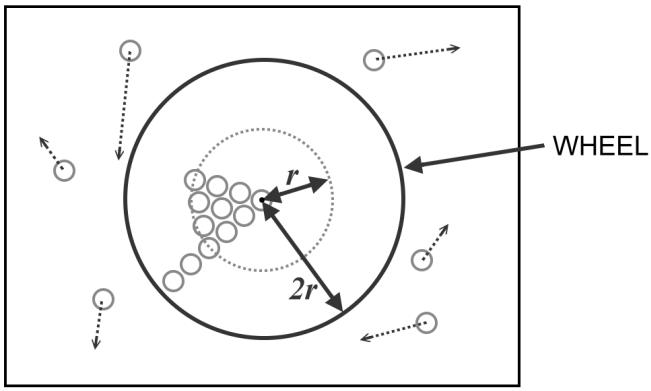


Figure 2C – Tracking particles: solid-state wheel in fluid environment

modified again. Assuming that the composition of the wheel is stable then irrespective of the value of Δt the encounter lists of particles lying within r of the centre of the wheel will have a fixed composition, representing just that fraction of the system's particles that lies within a distance of r. Particles in the wheel lying further than r from the centre of the wheel will have encounter lists made up of two parts: a fixed set of particles (their neighbours in the wheel), and a changing set that grows with time (their fluid neighbours).

If now the system is a cell in a fluid medium containing nutrient molecules then thinking in terms of particle correlations helps to make clear the distinctiveness of living systems. For although over short timescales (certainly shorter than the cell division interval) the cell looks to be structurally relatively stable, over longer timescales the compositional dynamics of the particle encounter lists will be unlike those seen in the cases just described. In particular, there will be few particles for which the corresponding encounter lists remain fixed for very long. This is because atoms are continually being taken up from the environment in particular molecular contexts, which are then broken down and their constituent atoms are recombined. Moreover, catalytic reactions shuffle atoms amongst molecules that differ widely in terms of mobility and lifetime. Thus the encounter lists will be both highly diverse (at a population level), and highly dynamic (individually).[52] Over timescales longer than several cell division intervals, say, the system's particle encounter lists may exhibit patterns of various kinds such that in statistical or formal terms they are similar to those that occurred earlier. In this respect the lists are like those of a machine-type artefact. The key difference is that whereas in the artefact the particle numbers in the lists stay the same, in the living cell the particle encounter lists at one time will be associated with a different set of particles than at substantially earlier or later times.

This 'particle tracking perspective' may be thought to offer no insights that more literal descriptions of material and chemical systems in terms of physical state and so on cannot yield. But it does help to draw attention to the different sorts of material dynamics that can occur in different kinds of system. It shows how different living systems and machine-type artefacts are – and yet how they are not unrelated. We can imagine intermediate systems of various kinds, and can see how a system can be more or less like a solid-state machine-type mechanism in terms of the dynamics of its underlying material composition. This is consistent with an aspect of the view I want to promote: where material systems are concerned, mechanistic status is a matter of degree, not an all-or-

[52] It may be that the statistical or other quantitative properties of particle encounter lists could form the basis for some measure of system complexity.

nothing affair. It also provides a salutary reminder that structures can be thought of – when considered at a particular level of detail – as processes.[53] That living systems are different in kind from non-living systems would be reflected, I predict (detailed analysis and perhaps computer simulation would be required to prove the point), in differences between their respective particle encounter lists. The greater predicted diversity of those for living systems compared with those of simple gases or solid-state devices seems related to the increased complexity of biological systems, and indeed the idea that the complexity of biological systems is one of their most significant attributes is an influential one. It constitutes an important element underpinning the rapid growth in recent years of systems biology (O'Malley and Dupré 2005; Noble 2006; Alon 2007a). That said, what is complexity, and what form does biological complexity take?

Capturing complexity

The 1980s and 1990s witnessed the re-emergence of the so-called sciences of complexity or complexity science, which has been defined succinctly as 'the study of systems with many interdependent components' (MacKay 2008). Perhaps, however, this definition is *too* succinct: machines fit the definition while not being the focus of complexity research. An alternative approach is to see complexity science as a trans-disciplinary set of conceptual orientations and approaches to understanding systems made up of numerous interacting entities and / or characterized by non-linear, and hence analytically intractable, mathematical relationships. Their origins lie in the growth of cybernetics in the 1950s and 1960s (Wiener 1948; Pask 1961), and while interest in complexity never disappeared completely it does for a time at least appear to have been displaced to the scientific margins as cybernetics gradually mutated into more focused lines of research in artificial intelligence, systems theory and computer science.[54] Renewed interest in complex systems is closely associated with the explosion of interest in fractal mathematics and chaos that occurred in the 1980s, and it can be related too to the rise of personal computing and the consequent ready availability of the computing resources required to investigate them. Complexity research is concerned especially with developing an understanding of a variety of ubiquitous patterns that occur across a wide range of natural phenomena, such as chaotic or emergent properties and behaviour:

[53] Even atoms can be thought of as processes, constituted by the interactions of sub-atomic particles such as protons and electrons, which in turn are understood in terms of the dynamics of quarks.
[54] An extensive list of complexity-related references is to be found at http://bruce.edmonds.name/thesis/ (last accessed 26 October 2009).

> Complex emergent systems, in which interactions among numerous components of "agents" produce patterns or behaviours not obtainable by individual components, are ubiquitous at every scale of the physical universe, for example in neural networks, turbulent fluids, insect colonies, and spiral galaxies. Complex systems also appear in a range of artificial symbolic contexts, including genetic algorithms, cellular automata, artificial life, and models of market economies.
>
> Life, with its novel collective behaviours at the scale of the molecules, genes, cells, and organisms, is the quintessential emergent complex system. Furthermore, the ancient transition from a geochemical world to a living planet may be modeled as a sequence of emergent events, each of which increased the chemical complexity of the prebiotic world.
>
> (Hazen et al. 2007, p.8574; reference citations omitted)

As well as reports of specialist work on topics such as fractal mathematics and chaos (e.g. Peitgen et al. 1992), emergence (e.g. Johnson 2002), self-organization (e.g. Kauffman 1993, 1996) and network dynamics (e.g. Watts 2003), a number of more general discussions of complexity have been published. In practice, however, these frequently amount to compendia of concepts and findings from those other specialist areas (e.g. Mainzer 2007). Sometimes discussions relate the concept of complexity to reduction, or rather to irreducibility (Casti 1994, chapter 5). Indirectly this reinforces the intuitively appealing idea mentioned in Chapter 1 of a possible link between reduction and simplicity, since the latter must presumably be considered the complement of some everyday conception of complexity.

I shall not spend much time discussing the canonical technical ideas of the complexity sciences, or at any rate I shall not address them via the simple physics-based models characteristic of much of the field[55]. This reflects my belief that the biocomplexity exhibited by the cell is of a different kind, and perhaps degree, to that of simple model systems. In the introduction I hinted at the reasons for such an assertion, and one of the purposes of this chapter is to amplify those hints into something like a defensible position.

Before going further, however, it will be useful to summarize briefly several of the more influential abstract accounts of complexity that have been advanced in recent decades. One especially prominent perspective is that of Herbert Simon, who associated complexity with the idea of 'nearly decomposable systems' (Simon 1996). Decomposable biological and physical systems are – to a first approximation – those in which the strength and frequency of causal interactions between system parts correlates in an approximately

[55] For example lattice gas and cellular automaton-based simulations of physical systems (see e.g. Manneville et al. (eds.) 1989).

inverse manner with their separation ('... in most biological and physical systems relatively intense interaction implies relative spatial propinquity' (Simon 1996, p.187)). Thus the closest parts interact most strongly, while the weakest causal relationships are between the most distantly separated parts. Simon generalizes the notion of decomposability, however, by stressing the primacy of interaction intensity over spatial proximity, as this gives him a concept that can be applied to social as well as material systems.

Decomposability tends to be associated with modularity, and in material contexts this in turn gives rise to the idea of mereological levels – or structural hierarchy, as Simon describes it more generally.[56] A system is made up of parts or modules, and parts in turn are made up of sub-parts. A system can be decomposed into parts that can be individuated – theoretically and practically – because causal interactions amongst the sub-parts within a part are stronger than those between the sub-parts and other parts. Hence sub-parts have a coherence and identity that appears to be prior to, or at least largely independent of, that of the parts they make up. Where attention focuses on interaction rather than structure the concepts of modularity and levels remain useful, since they can be used to discuss parts of systems – sub-systems – which are causally relatively self-contained, irrespective of issues of spatial distribution.

From this initial articulation of the concept of decomposability Simon identifies a class of system in which overall decomposability is accompanied by causal inter-relationships that cut across the clustering of interactions that makes for hierarchical decomposability. These kinds of systems he terms nearly decomposable[57], and he argues that many of the systems we think of as complex are nearly decomposable in this sense. A pertinent question from the point of view of my concerns in this chapter then is whether such systems are complex in a sense capable of subsuming biological systems such as the cell. Hierarchical and modular organization seem often to make for systems that are analytically tractable and thus rather simple (and in an organizational setting hierarchical structure provides a way of simplifying the management of large numbers of people). If this is so then we should associate complexity especially with the module- or level-crossing causal interactions that mandate the 'nearly' in 'nearly decomposable', rather than with the modularity that is the typical accompaniment of decomposability.

---

[56] Modularity does not obviously follow from the idea that 'intense interaction implies relative spatial propinquity'. It seems necessary, to account for modularity, to additionally postulate that there is not a continuous spectrum of interaction intensity but rather a clustering of intensities around particular values. In physical systems it is the different forces, with their different ranges and strengths, that give rise to structures at different scales and thus establish the 'levels' that naively we see in natural phenomena.

[57] Systems that are nearly decomposable are often referred to as 'near-decomposable'.

Another perspective on complexity, related to Simon's but less well known, is that of William Wimsatt. He distinguishes between descriptive complexity and interactional complexity. Descriptive complexity is a measure of the extent to which different theoretical concepts and frameworks, when applied to the same system, recognise the same sets of structural and process boundaries. (He terms the different spatial decompositions of a system S into sub-systems under different theoretical perspectives the different K-decompositions of S.) If a system is amenable to analysis or treatment according to two distinct theoretical frameworks then the system is more descriptively complex when the terms and operations of each framework serve to pick out or define spatially distinct objects, regions or processes than when they establish or respect similar boundaries and objects (Wimsatt 2007, p.184; see also Peacocke 1989, pp.245-249). The concept of interactional complexity is intended as a tool for distinguishing between systems on the basis of the complexity of their causal interactions 'with special attention paid to those interactions that cross boundaries between one theoretical perspective and another' (Wimsatt 2007, p.184). A system is interactionally simple if none of the sub-systems in a decomposition made according to causal interaction criteria cross boundaries between its different K-decompositions, and interactionally complex to the extent that they do. These ideas are somewhat abstract, but an important result is that if S is descriptively complex – i.e. if the different K-decompositions respect different spatial boundaries – then the researcher has to take into account the relations of parts for all parts of the different decompositions, and S is correspondingly epistemically intractable.

Wimsatt claims to part company with Simon over the consequences of evolution for system organization. Simon argues that evolution will favour modular system organization, on the basis that the attainment of a functional configuration will be more probable given hierarchical and progressive self-assembly. The argument he gives is essentially entropic: for all the parts of a complex system to come together simultaneously in the 'right' configuration (i.e. that selected on the basis of functional capability) is far less likely than for some subset of the parts to come together in the right way. Hence evolution will favour possibilities for the 'successive aggregation of individually stable sub-assemblies into larger sub-assemblies', as Wimsatt puts it (Wimsatt 2007, p.188). In biological contexts Simon's idea, if it supposed to provide some kind of account of system development and/or functioning, sounds somewhat simplistic. As I discuss in the next chapter, the functional sub-systems of biological systems are typically not mereologically simple

structures assembled like building blocks, which is what seems to be implied by the mereological tenor of Simon's treatment.

Wimsatt takes issue with Simon's thinking on the consequences for complexity of evolution, although his reasoning is hard to follow as it builds on what he says about interactional complexity. What it boils down to, however, is that 'considerations of efficiency in evolution would lead to the co-adaptation and increased interdependence of parts of a functional system', and this results in an increase in a system's descriptive and interactional complexity (p.190). (He notes that modern man depends for survival on particular social environments, and similarly many bacteria prosper only in the context of a bacterial culture.) One result is a drift away from neat structure-function mappings – he speaks of 'recognizable physical objects' rather than structures – of the kind discussed earlier in relation to mechanism. This idea is an especially attractive one: a significant aspect of complexity of the kind seen in biological systems such as the cell is to do with the way in which functions are distributed over structures. (About which I say more later.) Moreover it speaks to the connection between complexity and mechanism in an intuitively plausible manner. But it is unclear to what extent Wimsatt's conclusion is actually incompatible with Simon's position. Simon's entropic self-assembly argument concerns structure-building interactions, whereas functional relationships of the kind relevant to Wimsatt's thinking about complexity are more to do with chemical interaction and reaction.[58]

Emergence and metabolic reaction networks

Having reviewed some fairly a prioristic, albeit apparently suggestive, thinking about complexity I now look at a specific strand of research in systems biology. This investigates the kinetic properties of the biochemical reaction networks that constitute an important element of cells as systems, and I shall look in particular at a paper by Westerhoff and co-workers (Boogerd at al. 2005). Reflecting on this work and the appeals it makes to philosophical ideas provides the opportunity both to introduce the idea of emergence and to identify a number of issues pertinent to causation in the cell that are connected to but distinct from the issue of complexity. (Making sense of the connection is

[58] Wimsatt appears to be aware of this issue when he attributes to Simon a position on the implications of evolution for complexity that combines the ideas of near-decomposability and mereological hierarchy (p.188 – 'But Simon's use of the concept of near-decomposability in the same article sometimes appears to suggest that he believes such hierarchical systems to be nearly decomposable in a nestable manner ...').

one of the things I will attempt to do over the course of the remaining chapters of the thesis.)

Boogerd et al. discuss the kinetic properties of coupled biochemical reactions in terms of the values of parameters such as enzyme and metabolite concentrations, the reaction-specific *K* values that express the affinities of metabolites for enzymes, and the equilibrium constant K_{eq} for each reaction. (The lower the *K* value, the greater the avidity with which an enzyme binds the corresponding metabolite. K_{eq} is the ratio of the product and substrate concentrations at equilibrium.) The dynamics of a system of reactions is, Boogerd et al. explain, a function of the properties of system constituents, system configuration and internal and external conditions (p.149). Modelling such a system requires, they claim, knowledge of two kinds of properties of the system parts: intrinsic properties, such as the mass or amino acid sequence of a protein, and relational properties such as the dissociation constant that characterizes the dissociation of a complex into its constituent parts. These relational properties, it is suggested, are 'sufficient to determine which parts of the system interact with each other and in what manner'. The relational properties, plus what they term a composition relation or a description of the spatial organization of the cell, constitutes a model of the 'static system' which, under the application of physical laws, yields the 'state-independent properties' of the system.

To obtain the dynamic properties of a system it is necessary to combine the state-independent properties with boundary and initial conditions, such as (in the case of a bacterial cell) nutrient, enzyme and metabolite concentrations, temperature and pressure. Simulation given these quantities then generates a set of state-dependent properties of the system as a function of time; these include rate of free energy release, nutrient uptake fluxes and growth rate. Boogerd et al. then argue that the parts making up the system now, when the system consumes energy to perform work, display 'component properties', which are functions of the relational (i.e. non-intrinsic) properties of the parts and the entire state of the system. They include the rates of metabolite conversion and the sensitivities of these rates to change in metabolite concentrations.

So far, Boogerd et al. have been treating the cell as a single-compartment system. Now they introduce the idea of modularity. Suppose that the total system A comprises two sub-systems, A_1 and A_2. (One could imagine that A_1 is an organelle in a cell that constitutes A_2.) Suppose in addition that A_1 contains enzymes 1 and 2, and that A_2 contains enzymes 3, 4, and 5. Sub-systems (modules) A_1 and A_2 interact, and each contributes to the boundary

conditions that determine the dynamic component properties of the other. The key idea is that if A_1 is studied in isolation then the interactions between A_1 and A_2 must be mimicked if its component properties are to be understood (p.156). But this may be easier said than done when the component properties of A_1 and A_2 co-determine (or co-constrain) each other. As the authors note, the system behaviour given compartmentation

> could then include oscillatory, or chaotic states that are not present in simpler systems. The behavior of A_1, in isolation is sometimes qualitatively different from the behavior of A_1 in A, and therefore, since the behavior of A is a function of A_1, understood as a component, the behavior of A cannot generally be derived from studies on simpler subsystems of A. In general, the (dynamic) behavior of A is not simply the superposition of the (dynamic) behaviors of its subsystems studied in isolation. Dynamic interactions can bring about qualitatively new behavior in complex systems. This is precisely where prediction of system behavior on the basis of simpler subsystems fails. We cannot predict the behavior of the components within the entire system and so cannot predict systemic behavior. This is emergence, with novel system behavior that cannot be predicted on the basis of the behavior of simpler subsystems.
>
> (Boogerd et al. 2005, p.156)

Such invocations of the concept of emergence are increasingly common, as possibilities for computer simulation bring the properties and behaviour of increasingly complex systems within range of scientific investigation. But what is emergence? I shall defer attempting to answer that question fully until Chapter 6, instead confining myself to considering briefly several relevant ideas. Popularly, the term is associated with notion of a whole being 'greater than the sum of its parts'. This phrase points to the way in which the properties and behaviour of certain systems prove difficult or impossible to explain or predict on the basis of the properties of their parts, as Boogerd et al. discuss in relation to metabolic reaction systems. One aspect of the emergent properties and behaviours such systems exhibit, related to unpredictability and apparent inexplicability, is unexpectedness or surprise (Casti 1994, 1997). The oscillatory or chaotic system behaviour that Boogerd et al. note, if unseen in non-modular system configurations in which enzymes 1-5 are present in a single compartment, would presumably be unexpected in just this way. But do Boogerd et al. succeed in demonstrating, as they claim to do, the existence of what they term a 'strong' form of emergence in cell biology, whilst avoiding what they regard as unpalatable metaphysical issues? ('We seek an account of emergence that is not merely epistemological and yet does not suffer from the problems of *a priori* metaphysics' (p.133).) I regard the claim that the emergence that is observed is of a different, 'stronger', kind to that typically studied in the complexity sciences, as being more problematic than the basic claim that the phenomena they see are emergent.

At the outset of their paper Boogerd et al. suggest that if 'some phenomenon is emergent, in the metaphysical sense, then it is somehow fundamental and irreducible. It is fundamentally different from the physical basis on which it nonetheless depends' (p.131). The concern is that emergent properties, construed in the way in which Kim construes mental qualia as metaphysically emergent (on account of the way in which they involve intrinsic properties that elude functionalization[59]), 'defy any naturalistic explanation' (p.132). They suggest that

> For emergence to have any positive role to play in a scientific setting, it must be understood differently. It must be compatible with the thought that scientific explanations are mechanistic explanations. ... Ideally, emergence would be a natural consequence of physical processes. Finally, emergence should not be merely an epistemic notion, as Ernest Nagel thought it would be.
>
> (Boogerd et al. 2005, p.132)

These passages raise a number of issues, with two in particular being especially important. First, there is the meaning of naturalism assumed by the phrase 'naturalistic explanation'. I consider the account of emergence I offer in Chapter 6 to be thoroughly naturalistic, but it is not naturalistic in the sense of naturalism implicit in Boogerd et al.'s stated worry. Naturalism in my account is just the stance that sees the psychological capacities that arise from the operation of our mind/brains as causally integral parts of the world, and hence a naturalistic explanation may legitimately – and may sometimes – depending on what we are interested in – have to reference knowledge of our mind/brains as well as knowledge of the world external to the mind/brain. Explanations of the world beyond our mind/brains, on this account of naturalism, are explanatory relative to the structures, processes and properties of our mind/brains. For Boogerd et al., however, naturalistic explanation apparently means explanation in terms of functions of properties and objects that do not reference our psychological constitution, as a substrate, or the capacities that arise from it. Emergence, as I shall explain, I see as lying in a particular region of explanatory problem space, as a phenomenon that lies as close to problems of perception and cognition as it does to those to do with the trajectories of thrown objects, for instance. Boogerd et al. presumably would disagree with such an assessment.

[59] That is, phenomenal experience has properties that cannot be defined in terms of causal or nomic relations to other properties in what Kim terms the reduction base, which is 'the domain of properties (also phenomena, facts, etc., if you wish)' which contains the 'basal conditions for our emergent properties' (Kim 1999, p.10). When a property can be functionalized, on the other hand, it is possible to establish such relations. Kim suggests, perhaps rather tendentiously, that [a region of] DNA is functionalizable as a gene in virtue of its fulfilling the causal role of transmitting phenotypic characteristics from parents to offspring, which property means that it satisfies the property of being a gene.

The second issue raised by these brief passages concerns mechanistic explanation. Specifically, they say that 'scientific explanations are mechanistic explanations', and cite Machamer, Darden and Craver (2000) and related sources as providing support. This assertion is contentious, for it seems to (a) imply that *all* scientific explanations are mechanistic explanations, whilst (b) leaving open the possibility that there are mechanistic explanations that are not scientific (if the assertion is not to be construed as a statement of the identity of scientific and mechanistic explanations). An MDC-style account of mechanism is invoked, but there is a suggestion that they are simply paying lip-service to such an account, and that what they actually mean is just something like mechanical causation, in which what happens is the result of physical interactions amongst material entities. This is suggested when they say that

> Microorganisms essentially are large biochemical networks. They exhibit a range of system properties, such as homeostasis, regulation, plasticity, and adaptation, that appear to transcend the physical properties of their constituent parts, including enzymes, individual pathways, organelles, and other systems smaller than the cell. It seems that it is here where life emerges from its inanimate constituent matter. Nonetheless, every phenomenon of the cell is mechanistically explainable. Emergent phenomena are mechanical effects.
>
> (Boogerd et al. 2005, p.133)

The account of emergence to which Boogerd et al. cleave is that of C.D. Broad, and this is an account that now looks rather dated. Broad discusses emergence in *The Mind and Its Place in Nature* (1925) and in a 1919 article entitled 'Mechanical explanation and its alternatives'. The two accounts are similar but not identical. In the former, better-known account, he defines (synchronic) emergence thus:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A, B, and C in a relation R to each other; that all wholes composed of constituents of the same kind as A, B, and C in relations of the same kind as R have certain characteristic properties; that A, B, and C are capable of occurring in other kinds of complex where the relation is not the same kind as R; and that the characteristic properties of the whole R(A, B, C) cannot, even in theory, be deduced from the most complete knowledge of the properties of A, B, an C in isolation or in other wholes which are not of the form R(A, B, C). The mechanistic theory rejects the last clause of this assertion.
>
> (Broad 1925, p.61)

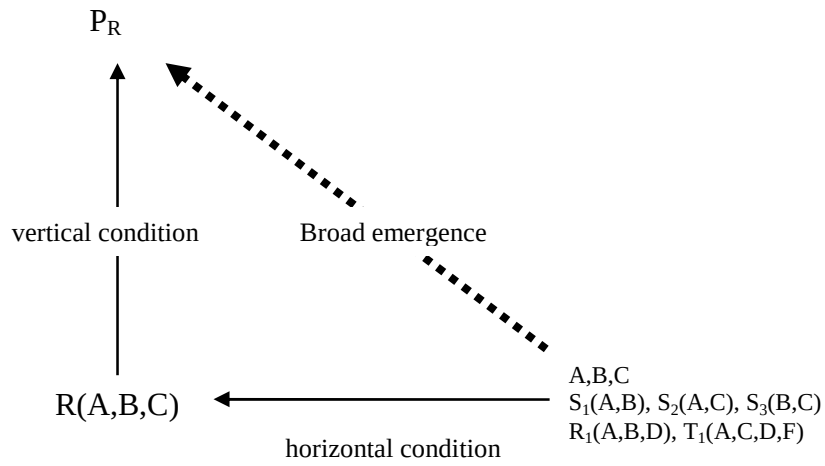
The seeming epistemic relativism of the assertion about property deducibility is problematic. Deep theoretical knowledge allied to powerful simulation techniques is

perhaps a combination Broad could not have anticipated. (He was writing before the development of the first computers, and in the infancy of quantum mechanics.) In his earlier account Broad discusses the nature of 'new and unexpected' chemical properties. In a rather elliptical way he expresses the idea that what we know about the properties of particular chemical groups in one chemical context provides an inadequate basis for predicting the properties of the same groups in different contexts. But inasmuch as chemical theory, allied to a body of empirically derived chemical knowledge that has expanded greatly since the time when Broad was writing, does provide a basis for rationalizing and in many cases predicting the properties of chemical compounds, the force of the passage is now heavily attenuated. We tend not to think that the properties of novel chemical compounds are emergent in any interesting sense, even if we cannot in practice predict them. Rather we suppose that if we had the computing resources to carry out more detailed calculations we would be able to predict the properties of the compounds. (See, for example, Bukowski et al. 2007, reporting the prediction of the properties of water from 'first principles'.)

The account of emergence to which Boogerd et al. appeal, then, is one that seems to set the bar sufficiently low that they can make their claim to see emergent properties. But the weaknesses of Broad's account are to some extent ameliorated by the way in which Boogerd et al. extend it by developing the idea that the fulfilment of either of two independent conditions is sufficient for emergence. They term these the horizontal and vertical conditions. The vertical condition asserts that

> A systemic property is emergent if it is not mechanistically explainable, even in principle, from the properties of the parts, their relationships within the entire system, the relevant laws of nature and composition principles.
>
> (Boogerd et al. 2005, p.135)

The horizontal condition on the other hand relates to the possibility of deducing the properties of the parts of a system from their 'properties in isolation or in other wholes, even in principle'. The vertical condition is about the explicability of *system* properties on the basis of properties of the parts and their intra-system relations, then, while the horizontal condition is about the *part* properties within the system, and emergence arises from the fulfilment of either condition:



In this figure (based on Boogerd et al. 2005, p.136, Figure 1), A, B and C are the parts of a system R; S_1, S_2 and S_3 are simpler wholes including these parts in different combinations. R_1 is a system with the same number of parts as R, T_1 is a system with more parts than R, and P_R is a systemic property.

Prima facie the horizontal/vertical distinction looks to be a meaningful one, and notwithstanding my reservations about Broad's discussion of emergence it appears at first glance to represent a framework within which the phenomena studied by Boogerd et al. fit quite comfortably. But further reflection leads me to wonder how necessary it is in any case to appeal to Broad's account of emergence to make the points Boogerd et al. wish to make. Key to Broad's account is that it deals with synchronic, not diachronic phenomena, and that it makes heavy use of the concept of a property, of a part or of a system of parts. This introduces a tension that Boogerd et al. address by implying, but not stating, that component properties (which are determined in part by the state of the entire system) are emergent, while actually pinning the label of emergence to phenomena such as oscillations that are clearly diachronic in character, even if that fact goes unacknowledged. The examples of emergent phenomena I outline in Chapter 6 tend to be readily visible patterns or behaviours, and it is these kinds of phenomena that in general attract the designation of emergence, rather than any underlying changes to system parameters that result in such visible effects. On the other hand if we really are to take quantities such as metabolite flux rates and so on as emergent phenomena, then Boogerd et al. seem in danger of falling into the very trap they warn against (pp.134-5), of seeing everything as weakly emergent.

What is clear is that when Boogerd et al. describe as strong the emergence they identify in the properties of reaction networks they mean something rather different from the sense of strong emergence indicated by Bedau (1997), or by Silberstein and McGeever (1999). Those other thinkers reserve the term for just two categories of phenomena: quantum mechanical EPR-style 'spookiness', and consciousness. Neither of those categories seems amenable to mechanistic explanation, in the MDC sense, in more straightforward 'mechanical' terms (referencing physical interaction between entities), or along lines compatible with what I have said about mechanism.

Overall, I am willing to concede that some of the properties of biochemical networks they describe can be considered emergent, albeit with reservations along the lines stated above. However, Boogerd et al.'s heterodox 'strong' classification looks highly unconvincing when set alongside the special ontological difficulties posed by quantum mechanical non-localism and by consciousness. Hence I contend that the emergence they see is best bracketed with the other cases of weak emergence identified in the complexity sciences and elsewhere. If their emergent properties do seem different in kind from some of those other instances then I suggest that this argues for a model of emergence founded not on ontological claims of the kind advanced by Boogerd et al. but rather on psychologically grounded epistemological ideas (which I describe in Chapter 6).

How does the work that Boogerd et al. report relate to the earlier discussion of complexity? One point of contact is in the area of functional distribution and implications for epistemic tractability. Irrespective of whether we see them as emergent phenomena, the oscillations in metabolite conversion rates, metabolite levels and so on that can result from the interaction of co-constraining metabolic modules or sub-systems seem likely to have significant causal consequences for the biology of the cell. Indeed, fundamental phenomena such as the cell cycle are known to depend on the levels and variations in levels of particular molecular species, and on the relative timing of processes that affect such quantities (Morgan 2007). (Relatedly, reaction-diffusion processes are thought to play a part in establishing chemical gradients and patterns that may have major developmental consequences (Turing 1952; Crampin, Hackborn and Maini 2002).) To the extent that the dynamics of the relevant processes are established through the kinetic specifics of reactions of the kind Boogerd et al. discuss, it is clear that key cell functions are sometimes to be associated not with specific structures but rather with interaction processes involving multiple structures.

This functional coupling of processes seems to illustrate rather well what Wimsatt says about interactional complexity and the co-adaptation of the parts of a system. And again, an important aspect of this kind of causal complexity is the difficulty we have in assigning functions to isolated entities or localized processes. If we interpret mechanism in terms of the alignment of structure and function then this, clearly, is non-mechanistic behaviour – yet such non-mechanistic processes may serve functionally critical ends within the larger context of the entire cell. Thus it is possible to see in this kind of example a reason why MDC frame their account of mechanism in terms of activities as well as entities: they want to speak of a set of coupled processes as constituting the mechanism for a certain phenomenon. I shall be looking again at the idea of mechanism at the end of the chapter, but for now it is worth noting how the MDC account is interposed between the machine (material) sense of mechanism and a looser causal process sense. They would say, I take it, that there is a way in which certain phenomena come about, and it involves particular entities, but they come about because of the processes (activities) in which those entities engage.

This discussion of the complex dynamics of cell processes touches on biologically profound phenomena, as mention of the cell cycle made plain. The big trends, constancies and circularities of cell life – metabolism and homeostasis, growth, genome replication, and cell division – can perhaps be thought of as the overall ends to which such life is directed, and even as largely constitutive of it. Some theoreticians have argued that biological understanding must begin by addressing these 'ultimate' ends, to develop a theoretical framework that captures what it is that makes living systems more than mere ensembles of interacting molecules.

Metabolic circularity and system autonomy

M,R systems

Theoretical biologist Robert Rosen argued that organisms are irreducibly complex, with knowledge of the parts of a living system providing an inadequate basis for comprehending the properties of the whole (Rosen 1991).[60] By complexity, however,

[60] One area in which Rosen found support for his view that machines and organisms are profoundly different was the many years of lack of success experienced by protein structurists in predicting structure from sequence. This led him to conclude that 'when contemporary physics claims to speak about matters

Rosen appears to have meant something rather different from the sense implicit in contemporary work in the complexity sciences.[61] There the sense prevails that complexity can be made a computable quantity, or is a system attribute that can be given an account in terms of the properties of a system's entities and the relationships their properties establish. Rosen, on the other hand, connects complexity with non-computability. I will focus on a separate claim he makes, however: that what distinguishes living from non-living systems is what he refers to as the closure of the former to efficient causation.[62] In other words, only living systems have the means to maintain and renew themselves from within – even though they are open systems in thermodynamic and material senses. Over a number of years he developed a theoretical framework concerning the causal circularity of metabolism that aimed at capturing this insight, which he termed the theory of M,R systems. (Here M stands for metabolism, R for repair – although some later interpreters have preferred to think of R as standing for replacement or resynthesis (Cornish-Bowden and Cárdenas 2007, n.1).) Traditionally biochemists focused on the catalysis by enzymes of specific chemical reactions, which might for example take the form:

$$A \xrightarrow{E1} B + C$$

This focus on the kinetics of individual reactions, and on the chains of reactions that constitute the major metabolic pathways, led to the neglect of what for Rosen was the biologically more fundamental issue of the collective properties of the totality of reactions that occur in an organism. These become apparent when we enlarge our picture of metabolism to encompass the systemic origin and fate of enzymes such as E1 in the reaction scheme above. Turnover of E1 is associated with the creation of new enzyme molecules to replace those lost through damage and subsequent targeted destruction (by routing to proteasome complexes). Synthesis of E1 means that it must be seen as another

---

biological, it either has nothing to say, or it gives wrong answers' (Rosen 1991, p.270). Perhaps, however, he was partly right when he said that the '[molecular] species involved are not *synthesized* from elements in any ordinary sense; rather, they *emerge* through a process of morphogenesis. Thus, when we ask 'why?' about them, when we treat them as effects and inquire into their causal correlates, we cannot make do with syntactic answers framed in terms of sequence. As always, there is not enough entailment in syntax alone to permit it' (p.275). I take these passages to draw attention to the difference between computation construed as deduction according to a simple logical schema, involving the performance of linear operations on tokenized representations, and computation of the kind involved in the molecular simulations discussed in the previous chapter. Protein structure prediction is clearly incompatible with the former but may, it seems, yield to the second.

[61] Although perhaps what he meant can be regarded as something akin to a strong manifestation of Wimsatt's interactional complexity.

[62] '[*A*] *material system is an organism if, and only if, it is closed to efficient causation.* That is, if *f* is any component of such a system, the question "why *f*?" has an answer within the system, which corresponds to the category of efficient cause of *f*' (Rosen 1991, p.244).

product of a metabolic reaction, not fundamentally unlike the products of the reaction it catalyzes (B and C above). The reaction by which E1 is synthesized is presumably catalyzed by another enzyme – E2, say – and that enzyme will itself be turned over in the same way as E1. The theoretical problem that Rosen sought to address was the potentially infinite regress implicit in this line of thinking, in which we need always to posit an additional enzyme to replace that catalyzing what we hoped would be the last reaction in a series.

Rosen's articulation of these ideas was highly abstract, and moreover was expressed in the obscure mathematical language of category theory. Recent interpreters, however, have taken up his ideas and rendered them more readily intelligible by explicating them in terms of biochemically plausible reaction schemas (Cornish-Bowden and Cárdenas 2005, 2007; Letelier et al. 2006). Their more concrete interpretations go some way towards demonstrating how the metabolic circle can be closed so that metabolism is closed to efficient causation. Such closure represents a considerable constraint on the permissible sets of metabolic reactions, for it requires that a metabolic reaction set generates its own catalysts. But, importantly, once a reaction set with this capacity arises it is self-sustaining to the extent that it requires only to be fuelled with the appropriate nutrient molecules.

By explicating metabolic circularity and M,R systems principles in terms of the closure properties of sets of reaction pathways, Cornish-Bowden and co-workers have succeeded in narrowing the gap between Rosen's abstract vision of what it is that makes living systems distinctive and the understanding we have of biomolecular processes. They offer a schematic picture that is simple enough that we can comprehend it almost without analysis, in the sense of not having to work through a lengthy chain of reasoning, and I suggest that it is this property of schematic conceivability that underpins any sense of understanding we are able to derive from it. However, just as it is troubling that the overall perspective articulated by the school of systems biology to which Boogerd et al. belong involves a heavy emphasis on reaction kinetics (Boogerd et al. 2007 – see especially Chapters 1, 2 and 4), so the focus on metabolic inter-relationships implicit in Rosen's thinking appears to represent an insubstantial basis for gaining a comprehensive perspective on living systems. The basis for the unease in the two cases is the same, and relates to the importance of structure and spatiality.

Suppose that a reaction generates a product molecule that in a particular milieu spontaneously associates with itself, or perhaps with another molecule. Maybe self-assembly proceeds to create structures that go on to structure cellular space, or maintain a

pre-existing structuring of cellular space, in a manner that sets constraints on possibilities for molecular diffusion and interaction. Then reactions that would otherwise occur may, thanks to structural consequences of the occurrence of a particular reaction or associative process, be prevented from occurring at all (and reactions that would otherwise be incompatible may co-occur, say if they are confined to separate compartments). The structural properties of particular molecules, including their symmetry attributes, thus may be important factors in determining the specific morphological characteristics of cell structures of various kinds. The specifics of such structures may determine what chemical entities and transformations a metabolic reaction network might include, and will have to be specified in any description of the total system network as constraints on network structure. Moreover, it seems almost inevitable that the possibility of specific kinds of system intervention, and the explanations we give for why and how certain phenomena occur, will depend on structural factors as much as on relational, kinetic or stoichiometric details of the reactions implemented within a system. In short, it seems likely that our representations and explanations of cellular phenomena must often reflect or incorporate spatial properties – in spatial terms – as well as kinetic properties. Growing recognition of this can be seen in the online abstract for one of the plenary symposia at FEBS-SysBio2009:

> Until now most of Systems Biology has focused on describing biochemical, gene expression or signal transduction networks in terms of changes in cell-wide concentrations with time. Less attention has been paid to the fact that these processes are also occurring in space. This symposium will address how movement of molecules through the cell is important for biological function and how macroscopic structures influence non-spatial processes.[63]

When we think about the connection between reaction and structure, any approach that focuses exclusively on reaction network connectivity and kinetics looks somewhat simplistic, even if the proponents of such approaches express a strong commitment to systems thinking. Worry about the neglect of structure (for example as articulated by Harold 2005) is arguably addressed in part, in a somewhat abstract manner, by another framework that emphasizes the circularity and closure of living systems, namely the theory of autopoiesis[64].

---

[63] Online scientific programme, FEBS-SysBio2009: 3rd FEBS Advanced Lecture Course on Systems Biology, 'From Molecular Biology to Biological Function', 7-13 March 2009, Congress Centrum Alpbach, Austria. *Symposium S: Systems Biology in Space* (Chair – Hans Westerhoff; lecturers – P Bastiens, M White, B Kholodenko). (Available at http://www.febs-sysbio2009.org/ - last accessed 30 October 2009.)

[64] The name autopoiesis derives from 'auto' meaning 'self' and 'poiesis' meaning 'producing'.

Autopoiesis

In the theory of autopoiesis, Maturana and Varela ('M&V') (1980) treat a living system as

> a machine organized (defined as a unity) as a network of processes (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.
>
> (M&V 1980, p.79)

They go on to explain that an autopoietic system

[A] continuously generates and specifies its own organization through its operation as a system of production of its own components, and does this in an endless turnover of components under conditions of continuous perturbations and compensation of perturbations. Therefore, an autopoietic machine is an homeostatic (or rather a relations-static) system which has its own organization (defining network of relations) as the fundamental variable which it maintains constant.

Clearly there are significant similarities between autopoietic machines and Rosen's M,R systems. In particular, there is the idea of a system that replaces its own resources. The major difference relates, as already hinted at, to the importance placed on spatial and topological considerations. In the context of the cell, particular significance is attached to the cell membrane and other structures in determining the 'relations that determine the topology of the autopoietic organization' (pp.90-91). It is important to note that it is not a specific structural organization that is kept constant (as one might infer from quotation **[A]** above), but rather an organization that preserves the capacity for autopoiesis – and it is towards the maintenance of [an organization that maintains] such a capacity that an autopoietic system is organized (as 'a unity'):

> What makes this system a unity with identity and individuality is that all the relations of production are coordinated in a system describable as an homeostatic system that has its own unitary character as the variable that it maintains constant through the production of its own components. In such a system any deformation at any place is not compensated by bringing the system back to an identical state of its components as it would be described by projecting it upon a three-dimensional Cartesian space; rather it is compensated by keeping its organization constant as defined by the relation of the relations of production of relations of constitution, specification and order which constitutes autopoiesis. In other words,

> compensation of deformation keeps the autopoietic system in the autopoietic space [i.e. of self-production].
>
> (M&V 1980, pp.92-93)

For Maturana and Varela, autopoiesis is the defining property of living systems – a living system is such 'because it is an autopoietic system in the physical space, and it is a unity in the physical space because it is defined as a unity in that space by and through its autopoiesis' (M&V 1980, p.112). This unity is the result of 'operational closure' in that 'every process in an Autopoietic network is the direct consequence of the interplay of components produced in other parts of the network' (Letelier et al. 2003, p.266). Letelier et al. reason that while all M,R systems exhibit the operational closure characteristic of autopoietic systems, by being closed to efficient causation, there is nothing in Rosen's work to show how an M,R system represents a unity in the topological sense of M&V. Hence it makes sense to think of autopoietic systems as a subset of M, R systems that fulfill additional topological constraints.

In the context of the present chapter key issues concerning the circular causality characteristic of the perspectives of autopoiesis and M,R systems theory are how it relates to the ideas of mechanism and complexity, and what the possible implications are for the analysis and understanding of biological systems. These questions can be addressed in part by focusing on the obvious point that these perspectives are pitched at the level of systems as wholes. This immediately invites comparison with conventional biological understanding grounded in the details of specific molecular and cellular processes, and draws attention to the possibility of an epistemic gap of some sort between 'top-down' and 'bottom-up' perspectives. An intuitive worry is that autopoiesis threatens to demonstrate that analytically grounded understanding, especially that based on detailed structural knowledge of biomolecules and cellular structures, will simply prove inadequate as a basis for recognising the salient causal stories that such structures play out. To elaborate the precise nature of this threat is non-trivial – and perhaps that is part of the picture of causal complexity with which it could be associated. It is helpful, however, to think about how different kinds of system respond to perturbation. A mechanistic system – in something approximating the machine sense of mechanism discussed earlier – is especially sensitive to perturbations that disrupt the systemic organization of its parts. The overall operation of such a mechanism depends on the individual functional contributions of its parts, and disruption of a part usually leads to the disruption of some sub-function necessary for overall system function. Sometimes, especially in 'mission critical' systems, functional sub-

systems are replicated so that if a particular functional element is damaged its function can be fulfilled by a back-up element. But even in this case of in-built functional redundancy we would be entitled to say that the system was now compromised relative to its condition prior to damage being inflicted on the original functional element, for if nothing else it now has a diminished capacity to tolerate further damage.

Inherent in the concept of an autopoietic system is the idea that it is far less brittle in the face of perturbations than the kind of non-autopoietic machine just discussed, as quotation **[A]** above makes clear. Within certain limits perturbations can be accommodated by compensating adjustments amongst the parts of an autopoietic system so as to maintain the capacity to absorb further perturbations. This makes for the possibility of continuity of identity not in terms of the persistence of a specific structural configuration but rather in terms of the ongoing operation of a system as a delimited physical object or region that continues to have the capacity to absorb and adapt to further perturbations. The ability to reconfigure in response to perturbations is a very striking difference between the two kinds of system. A man-made machine is standardly capable of existing in only a very limited number of configurations of its parts, and lacks the resources to renew its parts from within. An autopoietic system on the other hand can – if my interpretation of Maturana and Varela is correct – exist in a multitude of different configurations, and the different possible system configurations can be chained together, and in fact do entail each other, in ways that are capable of tracking and responding to environmental changes to preserve autopoietic capability.

The compensatory capacity of the cell to which the theory of autopoiesis draws attention appears potentially to relate to several distinct characteristics or sets of properties. First there is the physical amorphousness or plasticity of many cells[65], and their consequent ability to conform to physical aspects of their environment. This is related to the flexibility of the plasma membrane and the ability of fluids to occupy volumes of arbitrary morphology. In a fluid medium particular structures can often be arranged in space in numerous different ways, and possibilities for altering their relative configurations are less constrained than in systems assembled from solid state structures. This kind of physical deformability cannot be considered a fundamental component of the autopoietic condition, however, for its utilization to absorb the effects of a perturbation sometimes represents an overall reduction in perturbation-absorbing capability – just as when a spring is compressed

[65] I am thinking of the cells of animals, of course, not (for example) plant cells or bacterial cells possessing a hard cell wall.

the possibility for further compression is reduced. Simple physical changes that result in this way are thus non-adaptive, in that they do not promise to maintain a system's capacity to absorb further perturbations. Perhaps we should think of the factors underlying the physical deformability of the cell just as buffering mechanisms that serve to reduce structural brittleness – or perhaps sometimes merely as by-products of other physical properties necessary for the maintenance of life.

Quite distinct from these properties pertaining to the nature of different physical states and configurations of matter are chemically grounded possibilities for alternative patterns of gene expression. These provide a basis for the generation of different sets of metabolic reactions, structural characteristics and cell identities or phenotypes, and this looks like a more promising basis by which to understand the autopoietic capacity of the cell. The idea here is that it depends on the abilities of cells (and organisms) to sense their environment and for this to trigger specific patterns of gene expression, in order to generate or stabilize particular phenotypic characteristics appropriate to those conditions.[66] 'Appropriate' in this context presumably has to be cashed out in terms of adaptiveness, by saying something to the effect that the triggered phenotypic features render the cell (organism) better able to withstand or exploit new or changed environmental conditions than would be possible under the prior expression pattern.

The phrase 'in order to' several sentences ago points to the fact that such an account has a teleological component. In other words it sometimes seems natural to explain the behaviour of a system in terms of particular goals or purposes. Di Paolo (2005) argues for a position that is very congenial to my own, but he suggests that the notion of adaptivity is merely implicit in various interpretations of autopoiesis. It needs to be added as additional ingredient if the theory is to provide a basis for biologically grounding what he refers to as 'intrinsic teleology' and 'sense-making', as indeed I have done in my interpretation. Intrinsic teleology refers to the ideas Kant sets out in the *Critique of Judgement* concerning the 'mutual generative relations between organismic components and between them and the whole, making the living system a natural purpose' (Di Paolo 2005, p.433), while sense-making Di Paolo explicates in terms of an individuality that through self-

[66] A variety of epigenetic mechanisms also have the capacity to shape cell phenotype, and again these are 'chemically grounded' in the sense that they depend on the specific properties of particular molecules. In the next chapter, however, I touch on how physical force can be sensed and generated within cells via interactions between the cytoskeleton and a range of metabolic and regulatory processes. The cell is thus a causal nexus in which behaviours arise as a result of the interaction and integration of aspects of external and internal contexts.

production has sensation and agency, 'an entity for whom its own continuation is an issue' (p.433).

The link between autopoiesis, adaptivity and teleology is intriguing because it suggests a connection between the causal distinctiveness and complexity of living systems, in their metabolic aspects, and the roots of living systems as cognitive systems. To explore this topic fully would take me some distance from my central overall theme of biological explanation, so I merely advert to it. Before moving on, however, it is worth noting an additional link: that between biological adaptiveness and Rosen's idea of anticipatory systems (Rosen 1985; Rosen and Kineman 2005; Pezzulo 2008). An anticipatory system is one 'containing a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to the latter instant' (Rosen 1985, chapter 6). Perhaps biological complexity can be considered in part the price that must be paid if a system is to be anticipatory. A system that is anticipatory is capable of exhibiting many of the behavioural characteristics we associate with life – such as the appearance of agency or a capacity for goal-directed action. And it is perhaps by being autopoietic that a system is able to maintain the organization that underlies its anticipatory capacities.

5. Complexity and cellular causality - II

The diversity of causal factors

Whether or not autopoiesis and anticipation are indeed the ends which biological complexity is best seen as serving, it remains for me to provide some sense of the material nature of that complexity. To do so now I shall describe some of the different kinds of physicochemical factor that potentially enter causally into cell processes. These factors are many and varied, and rather than attempt to catalogue them exhaustively I shall contrast a number of strikingly different causal factors that figure prominently in our understanding of the cell, in order to appreciate the nature and degree of the diversity involved. The five kinds of causal factor I shall examine are: (1) synthesized 'molecular gadgets' such as the protein machines discussed in Chapter 2, which often implement chemical reactions or are instrumental in enabling processes such as the transport of molecules through membranes; (2) self-assembly, which underpins the formation of cellular membrane boundaries and compartments; (3) competitive assembly/disassembly, as involved in the processes that construct, rebuild and dismantle cytoskeletal structures; (4) the persistent molecular templating system represented by the genome and associated gene expression componentry; and (5) collective properties relating to the fluidity of the cytoplasm, such as diffusion coefficients and the phenomenon of macromolecular crowding.

Molecular machines

The protein machines discussed in the previous chapter are characterized by relative compactness and stability, and by the generally clear alignment of structure and function. Like machines they consist of parts that stand in particular relationships and that perform specific functions which contribute to the performance of the gadget's overall function. These properties are exhibited by enzymes, which are protein molecules that catalyze specific chemical reactions, i.e. that speed up processes in which atoms are rearranged into new molecular configurations. Substrate molecules bind, generally in a very specific orientation, at a particular region of the enzyme molecule termed the active site. (It is thought that substrate molecules are sometimes guided into the active site by the electrostatic field surrounding the enzyme.) Substrate binding brings specific chemical groups of the enzyme into precise spatial relationships with parts of the substrate(s), and

generally the effect is to distort the substrate(s) into a conformation that makes the structures of the reaction products thermodynamically accessible. In other words, the catalytic action of an enzyme depends on stabilization of the 'transition state' of the reaction, thereby reducing the activation energy of the reaction. In this way a reaction that would usually occur only very slowly – if at all – at physiological temperatures can be speeded up by many orders of magnitude. The key is the precision and specificity of the interactions between enzyme and substrate chemical groups. (Increasing the energy of random atomic and molecular motions by raising the temperature often will increase the speed of a reaction, but by nothing like the kind of factor routinely achieved by enzymes. This is because such Brownian motion is non-specific, and will distort the substrate in the required manner only infrequently.)

In general the reactions catalyzed by enzymes constitute steps within larger networks, of the sort studied by metabolic biochemists and systems biologists. Regulation and coordination of the rates of different reactions is sometimes achieved by modulating enzyme action in some way. A co-factor may bind non-covalently to an enzyme molecule at a site remote from the active site but in such a way that the structure of the active site is modified. Or another enzyme may catalyze a reaction that adds or removes a chemical group to or from the enzyme, with similar activity-modifying effects. Thus many enzymes exist in phosphorylated and dephosphorylated forms of differing chemical activity. The enzymes that carry out phosphorylation (usually termed kinases) and dephosphorylation (termed, more prosaicly, dephosphorylases) may act quite specifically on a particular enzyme or class of enzymes. It is easy to see how this mechanism of enzyme regulation provides a means for establishing a high degree of causal interconnectivity between cell reactions. Moreover it enables rapid adjustment to the rates of specific chemical reactions without modifying the number of molecules of a particular enzyme present in the cell – although mechanisms also exist for making just such modifications through changes to the rates of both gene expression and protein destruction.

In relation to earlier discussions about structure, function and the degree of 'mechanisticity', it is worth noting that even a single enzyme molecule is likely to contain regions that are more machine-like than others, in the sense that specific functions are more readily assigned to some parts than others. In particular, the chemical groups disposed around the active site will tend to be functionally highly specific: it is common to say that the function of group X is to bind group Y of the substrate, for example. Other parts of the enzyme's structure may be functionally much less individually distinctive. They

may serve simply to provide, collectively, a stable framework that establishes the structure of the active site and confers on it specific attributes relating, for example, to flexibility or (perhaps) to the establishment of an electrostatic field capable of selecting and steering substrate molecules towards the active site.

As I said, not all molecular gadgets catalyze chemical reactions. Some function more passively to enable or facilitate particular processes – with some of the pore complexes that support the transport of molecules across membranes perhaps providing good examples.[67] Functional protein complexes can be large structures, so we should probably resist the temptation to pick out what might count as a molecular gadget on the basis of size alone. Rather, it is functional specificity in relation to structure that is criterial.

Self-assembly

The second causal factor I mentioned was self-assembly, or the tendency for a structure to come together through the spontaneous association of its components. There is overlap and continuity here with the molecular gadgets discussed above, since functionally specific protein complexes may form through self-assembly processes (and protein folding can be regarded as a form of self-assembly). Different epistemic contexts emphasize different causal factors. If what we want to explain is the formation or stability of a functional complex then localized self-assembly may be one of the relevant causal factors, whereas if we seek to explain how a protein complex implements a particular function then the causal characteristics of molecular gadgets will have explanatory significance.

Some of the more functionally non-specific structures that make up the cell are formed through the assembly of multiple copies of the same molecular species, or large numbers of similar species. Examples are the plasma membrane and the proteinous filaments of the cytoskeleton, the former closing off volumes of space and the latter serving to structure membrane-delimited space and provide a dynamic scaffold capable of resisting or exerting force and of supporting other more specialist structures and processes. The plasma membrane of the cell consists – as predicted by Gorter and Grendel (1925) – of a bilayer of amphiphilic lipid molecules, each of which consists of a hydrophilic head region and a hydrophobic tail (Singer and Nicholson 1972). In an aqueous environment

[67] Although some such complexes function selectively and actively by orchestrating complex reaction schemes.

self-association of lipid molecules takes place spontaneously, with the bilayer structure in effect being the result of two sheets of lipid molecules coming together in tail-to-tail orientation, so that their hydrophilic heads face the surrounding water molecules. (The biophysics of the process has obvious parallels with aspects of protein folding.) Bilayer closure results in the formation of lipid vesicles, and the cell itself can be thought of as being based upon a giant lipid vesicle. Lipid bilayer membranes, all other things being equal, are highly flexible, but in cells their structural properties are modulated in a variety of ways. These include alteration to lipid composition – for example cholesterol serves to make the membrane less permeable to protons and sodium ions – and anchoring membrane via specific membrane proteins to cytoskeletal elements.

Vesicular compartments can serve to sequester particular molecular species so that their activities take place in specific conditions distinct from those prevailing in the bulk cytoplasm. Some activities are ones from which the rest of the cell must be protected, for example the proteolytic activities that occur in lysosomes, while other processes can be seen as requiring protection from other cell components. (Chaperone complexes serve such a purpose by protecting folding proteins from aggregation, as well as favouring compact conformations and thus promoting folding. However they are not lipid but proteinous structures.)

Competitive assembly/disassembly

The cytoskeleton itself consists of at least three main types of filament: actin filaments ('microfilaments'), intermediate filaments, and microtubules (see Lodish et al. 1999, chapters 18 and 19). Microfilaments and intermediate filaments maintain cell shape primarily by bearing tension. The former are found beneath the cell membrane, while the latter structure the internal space of the cell and provide support for organelles. Actin filaments form part of many signalling mechanisms, coupling the action of transmembrane receptor systems to internal signal processing reaction cascades. In conjunction with actoclampins (actin filament-associated molecular motors) actin is involved in a variety of cell motility functions, including the movement of podia of various kinds, dendritic spines, and intracellular vesicles, and it is implicated in membrane-trafficking processes such as endocytosis, exocytosis and phagocytosis. Many of these processes are powered by ATP hydrolysis.

Microtubules are hollow polymers of dimers of alpha and beta tubulin, and form a variety of intracellular structures including the mitotic spindle involved in segregation of the chromosomes during the division of eukaryotic cells. They also provide more general structural support within the cytoplasm, for example for organelles. The microtubule organizing centres (MTOCs), which play an important part in microtubule nucleation and organization, consist of alpha, beta and gamma tubulin subunits arranged in a ring complex that acts as a scaffold onto which alpha/beta dimers polymerize.

A key feature of the cytoskeletal filaments is their highly dynamic nature, attributable in part to the competition that exists between independent polymerization and depolymerization processes (i.e. between assembly and disassembly). In the case of microtubules rapid shrinkage is possible via a fast dissociative process. This depends on the fact that tubulin exists in GTP- and GDP-bound states. GTP-tubulin is what is added in polymerization, but the bound GTP may be hydrolyzed to GDP after addition, and GDP-tubulin has a greater propensity to depolymerize than GTP-tubulin. A GTP-tubulin cap has the effect of stabilizing a microtubule, whereas GDP-tubulin at the tip of a microtubule will dissociate. Hence if the protective cap is hydrolyzed to GDP-tubulin then rapid disassembly of the microtubule results. Polymerization and depolymerization rates are influenced by a variety of microtubule-associated proteins (MAPs), and the ability to regulate these provides the means to couple cytoskeletal properties and dynamics to cellular events and processes, and, more indirectly, to extracellular events.

Some argue that the cytoskeleton acts as an intelligent tensegrity structure, in which tensile forces can be distributed over the entire cell through the combination of tension-resisting microfilaments and intermediate filaments and compression-resistant microtubules (Ingber 2003a). Whether or not this picture is correct, evidence exists for the coupling of many cell processes and events, including metabolic reactions and processes connected with regulation of the cell cycle, with cytoskeletal dynamics (Norris et al. 2005; Michie and Löwe 2006). Although the coupling mechanisms are in general currently poorly understood, knowledge is rapidly being gained of how mechanical force can be sensed and generated by macromolecular fibrils and skeletal structures both inside the cell (Ingber 2003b; Albiges-Rizo et al. 2009; Christof et al. 2009; del Rio et al. 2009) and in the extracellular matrix (Smith et al. 2007; Hytönen and Vogel 2008). Such research suggests that cell processes are highly integrated despite their plasticity and adaptive capacity – or perhaps one should say that it is through being highly integrated that the cell is adaptively plastic.

Templating structures

Of course, any discussion of the causal factors that contribute to cellular life would be incomplete if no mention were made of the genome. I shall say much more about genome function in Chapter 7, so here will confine myself to the briefest of remarks. In contrast to the labile, evanescent nature of many cell structures, the DNA complement of the cell is notable for its stability and persistence. Ignoring mitochondrial DNA, the cell's DNA is packaged into chromosomes which for most of the cell cycle are present (in diploid cells) as pairs. Barring cell-specific mutation events each member of a chromosome pair has the same base sequence as every other equivalent instance in the organism (ignoring the gametes). Mechanisms exist to ensure that damaged DNA is repaired, and hence sequence is preserved across chromosome replication during mitosis. Despite the causal priority which has often been accorded to the genome (see Chapter 7), the DNA molecule itself can be regarded as playing a rather passive role in genetic processes.[68] These depend not just on DNA but also on the presence of a large number of protein-based and RNA-based molecular machines, and it is only in conjunction with them that the DNA sequence can have any cellular effects. Then, as is well known, it acts – sometimes only rather indirectly, on account of complex downstream RNA processing operations (Nilsen 2003; David and Manley 2008) – as a template for the production of specific proteins.

Fluidity

My final example of a causal factor that might necessarily figure in the explanation we give of some cellular phenomenon is the fluidity of the cellular milieu and properties related to it. It is an enabling condition for almost all the processes on which cellular life depends. Self-assembly and disassembly processes like those discussed above, for example, are possible in large part precisely because of the fluidity of the cytoplasm, for it is this that typically enables components to translocate, re-orient, come together and disperse again. Fluidity is a collective property, even if it depends on the properties of the molecules that form the fluid. If once it was supposed that the cell amounts to a bag of enzymes in aqueous solution then that image must now be discarded almost in its entirety. Just now I attempted to emphasize the contrast between the tight functional packaging characteristic of molecular gadgets on the one hand and the sometimes generic and often space-defining nature of self-assembly processes on the other. Now my inclination is to stress the

[68] 'DNA is a dead molecule ... it has no power to reproduce itself. Rather it is produced out of elementary material by a complex cellular machinery of proteins' (Lewontin 1992).

continuities that exist between the various molecular species that compose the cytoplasm, for it is the high 'bandwidth' of molecular species that makes it so distinctive.

In addition to water and other small solute molecules cell compartments are densely packed with heterogeneous populations of molecules that vary greatly in size, chemical properties, life-time and copy number (the number of tokens of a given molecular type that are present at a given time). Cytoplasm has been estimated to be a 70% solution of macromolecular species, and this phenomenon of 'macromolecular crowding' is increasingly thought to influence a range of processes (Ellis, 2001; Hall & Minton, 2003; Schnell & Turner, 2004; Golding & Cox, 2006). These include protein folding and stability, enzyme activity, DNA condensation, intracellular signalling, and pattern formation (Bray 1998; Weiss 2003; Weiss et al. 2004). Presumably this is principally a result of the importance of molecular motion in cell processes. Reactants must encounter enzymes if metabolic reactions are to occur; reaction products must be able to leave an enzyme's active site if the enzyme is to perform multiple catalytic cycles; synthesized proteins must reach particular sites, e.g. at the cell surface in the case of receptors. Some transport processes are active, i.e. involve the hydrolysis of ATP molecules (for instance those powered by motor proteins), but others occur passively, by diffusion. The rate at which molecules diffuse depends on a range of properties including size, shape, electrostatic charge, and hydrophobicity (and propensity to aggregate). This makes the cytoplasm unlike the simple fluids usually studied by physical chemists, the properties of which can often be expressed in terms of a small number of variables. To attribute to cytoplasm simple liquid properties looks naive or mistaken: its properties, because of the wide spectrum of molecular types that make it up, are complex and will affect different molecules differently, according to their size and other properties.

Fluidity is a major factor behind the cell's robustness and ability to replicate itself. This can be appreciated if we think about how a broken part in a man-made solid-state artefact is replaced. It is necessary first to dismantle the artefact by performing a series of specific manipulations, probably in a particular sequence, in order to access the damaged part. This can then be repaired in situ, repaired after its removal from the artefact, or replaced with a new part, and the artefact re-assembled. Re-assembly re-establishes the precise spatial relationships amongst the parts that define the causal action of the artefact's mechanism, and again this involves performing another highly constrained series of manipulations. The artefact may well be unusable until the damaged part is repaired or replaced, and the repair must be carried out 'from the outside' – artefacts generally lacking

the capacity or resources to repair themselves from within. If a protein in a cell is damaged, on the other hand, then chances are that the cell is not thereby completely shut down. Other copies of the same protein will continue to perform the protein's usual function, and the damaged copy can be tagged and routed through the cell to a proteasome complex for destruction. Additional copies of the protein may be synthesized using metabolites whose atoms are sourced ultimately from outside the cell, in accordance (to a first approximation) with a specification stored in the genome. The fluidity of the cellular environment means that multiple functions can occupy roughly the same volume, for functions are not defined by the maintenance of rigid relations amongst parts. Indeed, the avoidance of the formation of fixed relationships amongst parts, e.g. through macromolecular aggregation, is often a condition of cellular and organismic function. (Certain pathological conditions such as Alzheimer's disease or Creutzfeld-Jacob disease appear to be associated with the formation of 'plaques' brought about by the large-scale aggregation of misfolded or partially folded proteins (Chiti and Dobson 2006).)

Logic and structure

It seems clear that important aspects of the complexity of the cell are the diversity of the structures it contains, the variety of the processes in which these participate, and the range of causally relevant properties these exemplify. The five CFs just described illustrate this diversity. Sometimes the biological phenomena we seek to explain (and perhaps control) depend primarily on just one kind of CF, but more often than not biologically significant phenomena depend on multiple processes acting in concert (Toettcher et al. 2009), with each process in general depending on multiple CFs. For example, specific catalyzed reactions capable of modulating the dynamics of assembly and disassembly of structures have the potential to bring about large-scale structural changes in the cell. Mechanisms for sensing the conditions that result from these might conversely then have specific effects on particular reactions or indeed on the expression of particular genes. This latter possibility implies an additional level of complexity, inasmuch as the cell not only involves the interactions of a diverse set of molecules and ions but the interactions of a set that is constantly changing. And while some chemical species are present at very low copy numbers (such as the chromosomal DNA molecules), others are present in large numbers (for example ATP or calcium ions).

Notwithstanding the variety of cell structures and processes, the diversity of the molecular types that implement them, and the wide variation (high 'bandwidth') we observe with respect to molecular token number, lifetime and mobility, increasingly it is suggested that we should expect to discover an underlying functional simplicity to cellular life. Such claims come in particular from the developers and adherents of the network approaches which are common to many current interpretations of systems and synthetic biology (see e.g. Alon 2007a). These approaches focus on the dynamics and topology of genetic regulatory circuits and interaction networks rather than on structural details (Barabási & Oltvai 2004; Weiss 2005). It is argued that cellular networks draw on a finite set of functional motifs that in many cases appear to be connected together in ways that ensure they do not interfere with each other. The resulting decomposability means, it is argued, that the cell can be expected to be complicated but not insuperably complex (Alon 2007b).

Network approaches potentially raise once more the worry enunciated by Harold (2005), in that they threaten to discard too readily explanatorily important structural information. An interaction or regulatory network diagram that omits details about compartmentation and context, for example, or one based on interaction properties of proteins measured outside their normal cellular milieu, might imply a degree of causal simplicity (or for that matter interactional complexity) which is at variance with biological reality.

Contrasting with the optimistic vision of an abstract underlying simplicity to cellular processes are perspectives that draw attention to the problematic complexity of fundamental phenomena such as the cell cycle. This complexity makes it hard to describe the cell concisely or in terms of a simple analogy with some other class of objects, and Norris suggests that the cell is uniquely complex. As a result it 'cannot be reduced to a single simplified model system without losing its essence' (Norris et al. 2004, p.125):

> The cell is neither a neural net nor an oscillating system of diffusible enzymes nor a dissipative system nor a set of phase-separated membranes and cytoplasm etc. Rather, the cell is the creator and the creation of an extraordinarily high density of different organizing processes that have autocatalytic relationships with one another. It is a system that produces self-organization and assembly by recruiting and dismissing a multitude of processes and molecules. Whatever does this is a cell and, for the moment, we only know one example: the biological cell. In other words, the cell is its own metaphor.
>
> (Norris et al. 2004, p.125)

A possible parallel here is with the sense of complexity articulated by Chaitin and Kolmogorov (Mainzer 2007, p.193), which relates the complexity of a string to the minimum length of the algorithm required to describe it. A string consisting merely of repetitions of a simple substring can be expressed more readily and compactly than one in which substrings tend not to recur, for example. In the limiting case of extreme algorithmic complexity a string is its own shortest description, and analogously 'the cell cannot be understood adequately as anything other than itself' (Norris et al. 2004, p.125).

Norris et al. cast doubt on the idea that focusing on just biochemical and molecular biological details will yield satisfactory explanations of fundamental cellular phenomena. Certainly, detailed knowledge of many of the cell's individual mechanisms at the macromolecular level has not yet resulted in our having a good sense of how such mechanisms collectively give rise to such core cellular activities as chromosomal replication and cell division – despite the considerable progress that has been made (Hartwell and Weinert 1989; Nurse 2000). Norris argues for an approach reminiscent of that advocated by Marr in relation the study of vision (Marr 1982): we must pose functional questions in order to identify the 'design principles' a system embodies. ('What is the cell cycle for? ... Are 'key' regulator proteins simply the messenger boys instructed by the dynamics of structures? Are these proteins just part of a molecular overlay that behaves as a coupled oscillator?' (Norris et al. 2004, p.124)).

Hyperstructures

The aim in asking and attempting to answer such questions is perhaps not wholly dissimilar to that of the network researchers mentioned above: to identify and describe a functional architecture behind the material complexity of cell structures and processes (Nurse 2008). But the differences of attitude and approach are considerable: network researchers are willing to countenance the elimination of the material and spatio-structural from their descriptions and subscribe to the optimistic vision of underlying simplicity, whereas Norris and co-workers embrace spatial materiality and seek an account that confronts and accommodates the dynamic organizational complexity – indeed 'hypercomplexity' (Norris et al. 2005) – of the cell (with a focus on bacterial cells). They do this by positing a level of organization situated between the molecular and cellular scales involving what are referred to as hyperstructures (Amar et al. 2002; Norris et al. 2007). One definition of a hyperstructure is 'an extended assembly of diverse molecules and

macromolecules ... that is associated with at least one function' (Norris et al. 2005, p.317), but the notion is richer than that of the protein machines discussed in Chapter 2. For whereas those we think of as stable functional molecular assemblies, hyperstructures can exist as non-equilibrium as well as quasi-equilibrium structures:

> A non-equilibrium hyperstructure is assembled into a large, spatially distinct structure to perform a function and is disassembled, wholly or partially, when no longer required. Its continued existence depends on its consumption of energy.
> (Norris et al. 2005, p.317)

Moreover, a hyperstructure 'interacts with other hyperstructures at a level of organization between the macromolecule and the bacterial cell' to control the phenotype (Norris et al. 2007). But how are hyperstructures to be individuated and classified? Norris proposes an approach that combines both structural and functional considerations, based on the idea that '[a] hyperstructure that appears, matures, and disappears follows a trajectory in a space in which the axes are the processes responsible for its existence' (Norris et al. 2007, p.310). He then gives a number of examples of hyperstructures that might be classified according to form (topology) and process (function). These include coupled transcription-translation-insertion ('transertion') non-equilibrium hyperstructures that (it is hypothesized) perform a variety of cell functions, and 'metabolons', which are large putative assemblies of metabolic enzymes in which intermediates are channelled from one enzyme to the next (Norris et al. 2007, pp.311-312). It is also postulated that re-initiation of DNA replication in *E. coli* is prevented by a SeqA hyperstructure in which multiple copies of SeqA sequester the genes encoding replication-related proteins (Norris et al. 2007, pp.314-315).

The concept of hyperstructures represents an attempt to describe elements of functional organization in a way that respects the dynamic character of cell processes. It shows the explanatory value biologists attach to structure—function relations, but argues for the necessity for a concept that extends beyond the stable structure—function relationships seen in machines and which the protein machine concept invokes. There is surely overlap here with the task MDC set themselves: to elaborate a concept of mechanism appropriate to the dynamic nature of molecular and cell biological phenomena. But the resources the MDC account provides for addressing the distinction between equilibrium and non-equilibrium structures and processes are limited, and the fact that the account is largely devoid of functional language seems to bring few benefits. On the other hand the concept of hyperstructures has not so far been articulated in especially precise terms, and it is unclear whether it should be regarded as primarily an epistemic notion or an

ontic one. Now I want to investigate whether it is possible to devise a descriptive-cum-normative perspective on mechanism that retains the flexibility that the MDC account's ontic dualism confers as regards processes but which is more explicit about functional issues. At the same time I want the new account to be rich enough to subsume yet have the potential to distinguish between different kinds of structure, process and organizational basis. Here the hyperstructure concept provides a useful model, although from a philosophical point of view it will be a virtue if we can be clear about when we are talking about ontic matters and when epistemic. As a starting point it is helpful to reflect further on the difference between equilibrium and non-equilibrium structures and processes, which in a systemic context can be thought of as constituting different organizational bases.

Statically versus dynamically maintained organization

The stability and resilience of machines, qua the kinds of solid-state artefacts discussed in Chapter 2, makes structure the natural basis for their explication, and has resulted in mechanistic explanation acquiring a certain sense – that which we associate with the word 'mechanical' – that again seems to foreground structure over process. On this view mechanisms are materially persistent structures, and when a mechanism such as a clock stops we are strongly inclined to say that the mechanism itself continues to exist; it is merely 'not working'. This association of the term 'mechanism' with structure rather than operation is very striking when it is compared with the situation that obtains in the case of biological systems such as living cells. For when the molecular motions associated with a cell's processes cease, the cell's organization collapses and decays and the cell ceases to be alive. If the cell is a mechanism then it is one that ceases to exist when it stops 'operating', or when the processes in which its parts participate cease. This is what it means to say that a living system is far from equilibrium.

Thus an important respect in which systems may differ is the way in which their organization is maintained. In machine-type artefacts of the kind I have described, organization is maintained by the properties of the solid state. To reiterate something I said earlier, parts have structures that are stable under ambient conditions, and this is true whether or not the parts are involved in the collective motions their shapes enable and which are characteristic of the operation of the machine. The organization of the cell, on the other hand, is dynamically maintained. The characteristic structures we see in the living cell reflect and depend on the occurrence – perhaps one could say that they are just the

visible manifestation – of a large number of dynamic processes, involving the sorts of causal factor already discussed.

A crude but nevertheless helpful analogy is with a fast-flowing stretch of river. The overall form of the river, as expressed by its width, depth and the shape made by its boundary with the riverbank, may be relatively constant. But this constancy depends on the maintenance of water flow, i.e. on the continuity of operation of a particular process. The flow rate can be thought of as the amount of water passing within a certain time interval through a (notional) plane roughly normal to the direction of flow. Two such planes could be used to define a stretch of river, and the overall form of this stretch will be constant if the flow rate as measured at both planes is the same. Differences between the flow rates as measured at the two planes will be reflected in changes to the form of the intervening body of water: the river will be become deeper or more shallow, wider or narrower (depending on the structure of the river banks). If the two notional planes are now replaced by solid barriers (and the inflow of water appropriately re-routed), then the dynamically maintained stretch of river is transformed into an equilibrium system that has the same overall form as the non-equilibrium system but which is importantly different in terms of the basis for its form.

An expanded view of mechanism

The contrasts between structure and process and between static and dynamic maintenance of organization suggest a basis for distinguishing between different kinds of material systems which harks back to the particle correlation ideas outlined at the start of Chapter 4. This is to consider the relative timescales over which matter flows through, or is turned over by, a system and its parts. Such flux rates are relative to particular regions of space, defined by real or imaginary boundaries. The solid-state parts of machine-type artefacts are of roughly equal stability and persistence, and the changes relevant to accounting for the action of a machine relate principally to alterations to the relative configurations of the parts. In a complex system such as the cell, on the other hand, there is a wide variation in the stability and persistence of different parts of the system, and persistence where it does exist may be achieved in different ways. As I have noted previously, much of the apparent structural stability is somewhat illusory, in that it is the outcome of highly dynamic processes of parts generation, maintenance and replacement within an overall framework of causal circularity. The genome looks highly stable, and

indeed in sequence terms *is* exceptionally stable – for reasons that we believe we understand (if sometimes only in outline) in terms of the requirement to produce proteins that fold reliably and which are functionally specific relative to particular contexts, and the need for key regulatory events and processes to unfold in constrained (and, through their phenotypic effects, selected) ways. But the apparent stability of genome sequence depends on the action of an array of error detection and repair systems, and moreover with every cell division one strand of each chromosome is generated anew, in accordance with the template provided by the other strand. Thus the genome has a dynamically maintained stability.

Why might this matter, so far as an account of mechanism goes? Because whether a system is statically or dynamically maintained has important implications for its ontogeny and operation, as well as for how we come to understand the system. The static organization characteristic of machines is maintained by the strength of its parts and of the couplings between them, with the latter generally being established by a complex spatiotemporal pattern of assembly operations. Breakages of parts are typically calamitous for the machine's ability to perform its usual functions, and remedying them usually involves undertaking complex disassembly and re-assembly operations. Once established, the machine's organization is not readily altered, and it is in general physically – topologically – impossible for different machines to share the same parts. None of these characteristics necessarily holds for the mechanisms making up cellular biological systems, thanks to the dynamic basis on which their organization is maintained. Fluidity and weak couplings mean that parts can be replaced without complete functional disruption, parts can be shared amongst mechanisms, and different parts can fulfil the same functions within a mechanism to tune and tailor its cellular role.[69]

Presumably it is recognition of these points, or something like them, that underpins neo-mechanist denials that biological mechanisms are to be regarded as machines. However, it is hard to find confirmation of this suspicion. In seeking to go further than MDC, in terms of being explicit about the organizational issues just discussed, it is important not to throw the baby out with the bath water. This means retaining their even-handed approach to the relative causal status of structures and processes. At the same time the explanatory importance to biologists of the idea of function, which the MDC account downplays to no great effect, argues for its incorporation. I propose that this is best done

[69] In addition some proteins – so-called moonlighting proteins – are capable of fulfilling multiple functions according to context (Jeffrey 1999).

by structuring an account not around the ontic dualism of entities and activities but rather on the basis of epistemology versus ontology. The perspective I have in mind is illustrated in Figure 3.

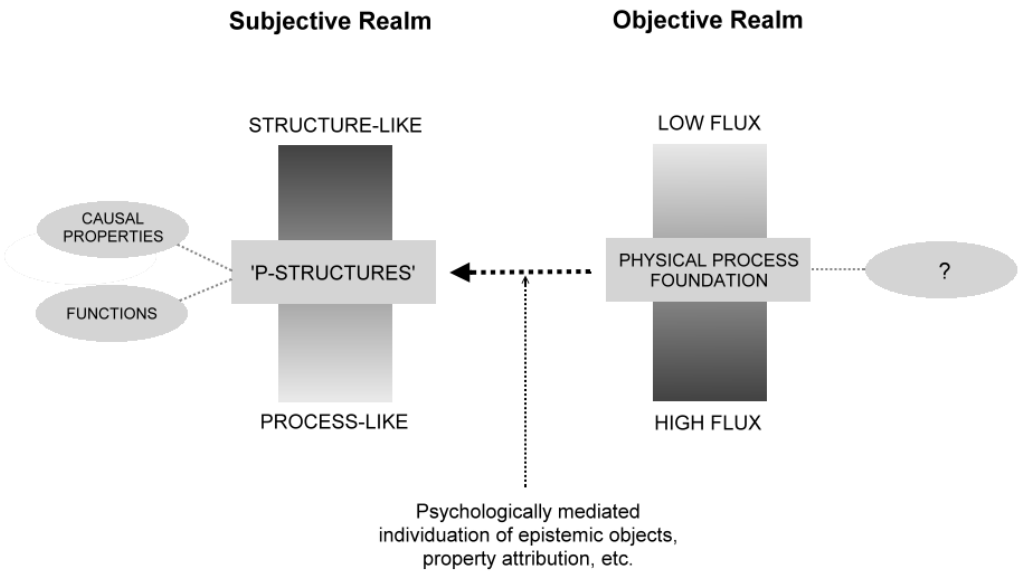


Figure 3 – Ontic/epistemic mechanistic framework

The right hand side of the figure provides for the idea that within a particular system or system part, or within a particular region of space, material is turned over and reconfigured at a particular rate. We can therefore think of systems, parts of systems, and regions as being 'high flux' or 'low flux'. What are in flux are entities, and it makes sense to say that these are the smallest structures that are stable over the time spans that we are interested in. These structures and their interactions, which I take to be real physical phenomena, establish what could be termed a physical process foundation. This is the set of physical processes that we observe, interact with, and intervene in, and on the basis of our scientific and perceptual engagement with them we individuate various structures and processes as having particular epistemic significance. (The question mark on the right hand side indicates the idea that the real physics making up the physical process foundation is presumably to a large extent unknown to us. The postulation of interacting entities flowing through regions of space – or of processes that support interpretation in these terms – represents the minimal metaphysics I think an account along these lines requires.)

Engagement with the physical process foundation, and the perceptually mediated and psychologically grounded object individuation and property attribution that this entails,

yields the epistemic objects and properties shown on the left hand side of the figure. My approach to maintaining neutrality about structures and processes, which appear in the MDC account in the guise of entities and activities, is to combine them into a class of hybrid objects I refer to as P-structures (the 'P' denoting both their potentially *p*rocessual character and the idea that they are *p*sychologically grounded rather than objectively real[70]). P-structures are the structures and processes in terms of which we frame our explanations, and to which we attribute functions and other properties. Some P-structures are principally structural (they do not change much over the timescales we are interested in) whilst others have a more processual character. The functions of parts and systems are not given to us and do not inhere in things in the objective sense in which the physical structures and interactions on the right hand side of Figure 3 might be said to exist. Rather we attribute them on the basis of a wide range of criteria (some of which I discuss in Chapter 7 in relation to the genome).

What can this more complex perspective on mechanism do that the MDC account cannot? First and foremost, it provides a way of distinguishing – on the basis of material flux – between machines, machine-like biological mechanisms, and more dynamically complex structures such as those indicated by the concept of hyperstructures. While there are strong overlaps on the left hand side of the diagram between these different kinds of system – in terms of the association of structures and processes with functions and causal roles – there are marked differences in terms of ontic underpinnings. Biological mechanisms are typically high flux processes (or sets of processes with different flux rates) whereas machines are almost invariably founded on low flux processes. This matters from a practical point of view, for high flux organization makes for physical intractability as well as epistemic opacity. The functionally significant dynamic material ensembles encountered in cell biology are hard to extract physically from their cellular contexts, and hard to study within them. To understand specific methodological issues in molecular and cell biology I think demands that we have an account along the lines just outlined. Moreover such an account provides the epistemological elbow room needed to address Torres' point (mentioned in Chapter 2 in the discussion there of the MDC account) about the possible involvement of negative causation in our mechanistic explanations.

The perspective just outlined remedies the inability of the MDC account to address squarely the issue of equilibrium versus non-equilibrium structures and processes, and I

[70] The term 'P-structures' also, not unhelpfully, calls to mind the 'P-' notation used by Alva Noë to distinguish between objects of various kinds and their perceptual equivalents (Noë 2004).

think represents an ontologically more straightforward way of tackling issues of structure, process and function. Arguably the picture it presents is still too broad, however, in that it lacks the capacity to distinguish between systems on the basis of the rate at which their organization changes, over and above the basis on which their organization at any particular moment depends. Some systems exhibit an organization which is stable over time but which is dynamically maintained, with the genome providing (to a first approximation) an example of this. Other systems display an organization that while also dynamically maintained is itself changing over time. A fast-developing embryo might be a good example. Perhaps systems of the first kind, systems that have a dynamically maintained but stable organization, pose a particular explanatory challenge. It may be the case, for example, that they wrong-foot any tendency we have to interpret causality within a system using cognitive strategies that are appropriate, and perhaps especially suited, to comprehending equilibrium phenomena but not necessarily non-equilibrium processes.

Another objection concerns the ability to distinguish between living and non-living systems. The earlier example of the river shows that it is not just living systems that exhibit the dynamic maintenance of stable form. In order to gratify our intuitions about what is deemed to be alive it is probably necessary to combine a hybrid ontic/epistemic account along the lines sketched above with ideas of the sort outlined in Chapter 4 about metabolic circularity, system autonomy, intrinsic teleology and the like. That ambitious project is alas not one I have the time or space to embark on here.

Conclusions

This chapter and the previous one have had to cover a lot of ground, but given the nature of their topic that is hardly surprising. In them I have attempted to show that the cell is complex in highly distinctive, indeed unique, ways. Much of its complexity can I think be summed up in the idea that it is constituted from a diversity of structures that exhibit extensive property variation across a number of dimensions, and these interact to participate in processes that unfold over a range of timescales. It is thus a 'high bandwidth' object, indescribable in terms of a few concepts or parameters. There is more to cellular complexity than this, however. In addition there is the way in which cellular processes are highly interconnected – as seen for example in the apparent coupling of metabolism, cytoskeletal dynamics and the cell cycle. This interconnectedness or high degree of integration tests, I think, some of our basic intuitions about the origins of robustness.

Where artefacts are concerned we are often at pains to decouple or causally insulate processes from each other, so that the effects of defects in or insults to one functional region of a system do not propagate to other regions. The cell's tolerance of substantial environmental variation, however, is positively correlated with its integration, since this is of a kind that confers a high capacity for adaptive compensation.

In the light of all these findings it is reasonable to ask whether the cell should be considered as some sort of mechanism, or even as a kind of machine. The obvious question then, however, is relative to what conception of mechanism or machine? Ganti has memorably described the cell as a fluid machine, an idea which he models via the concept of the chemoton (Ganti 2003). This is a minimal living system composed of a membrane-defined compartment, a heritable information store, and autocatalytic metabolic machinery. This notion, like Rosen's idea of M,R systems and the theory of autopoiesis described earlier, expresses much of the overall causal character of the cell. But it is hard to connect such perspectives with the extensive and detailed structural knowledge we have of molecular biological processes. The concept of hyperstructures helps to bridge the gap by providing a conceptual resource for thinking about dynamic processes defined in terms of specific groupings of molecular structures and events. I have tried to combine the spirit of hyperstructures with some of the insights provided by the MDC account to develop a new perspective on mechanism. This is flexible enough to subsume many (and maybe most) cellular phenomena, but that is not to say that the cell overall should be thought of as a mechanism by its lights.

I suggest that in the end it comes down to whether we reach a point at which we can specify a set of P-structures in terms of which we can both conceive of the cell and describe how it operates. This is largely an epistemic matter to do with our ability to imagine processes involving multiple entities, and the ability to attribute functions to structures and processes is probably key to the data reduction problem we face when our finite cognitive capacities are confronted with an object involving as many causally interwoven elements as the cell. The regularity of the cell cycle, as well the robustness of individual development and the stability of identity exhibited by the individuals making up a species population, all argue powerfully for the idea that there is an underlying order to cellular entailment structures. Perhaps we would prefer the idea of mechanism at root to be associated simply with the existence of a robust entailment structure within a system. But in that case the cell may be a mechanism we can never explain. Faced with that possibility we might decide that it is better to draw a strong epistemic boundary around the concept of

mechanism and say that something is a mechanism only to the extent that we can explain, in a finite (and humanly reasonable) amount of time, even if only schematically, how the behaviours it exhibits come about.

6. Emergence and cognition

Introduction

As Chapter 4 demonstrated, discussions of complexity often raise the topic of emergence (Darley 1994). Like complexity it is a concept that has proved remarkably resistant to analysis – as indicated by the growing literature surrounding attempts to go beyond the well-known notion of a whole being 'greater than the sum of its parts' (e.g. Holland 1998; Clayton and Davies 2006; Ryan 2007). A number of taxonomies of emergence, sometimes quite intricate, have been proposed (e.g. Stephan 1999; Fromm 2005; Deguet, Demazeau and Magnin 2006), with a common feature of several recent accounts being to distinguish between strong and weak variants. Just such a distinction was noted previously in the context of the work of Boogerd et al. (2005) on metabolic modelling. As I explained there, strong emergence is usually mentioned in connection with consciousness and some of the puzzling aspects of quantum mechanics, and is typically held to be ontologically problematic (Kim 1999; Silberstein and McGeever 1999). The weak emergence associated with complex systems, on the other hand, is generally considered to be more benign – often seemingly on the grounds that it is considered to be as much a problem of epistemology as ontology – and a number of more or less technical explications have been proposed (e.g. Bedau 1997; Holland 1998). Nonetheless some recent authors are at pains to stress the ontological reality of emergence (Bedau 2008; Huneman 2008), as if any epistemic aspect is irrelevant to understanding the phenomenon. In this chapter I investigate the concept of emergence in more detail and propose that it should be seen jointly in ontological and epistemological terms. Failure to do so renders the basis on which emergence attributions are made appear needlessly mysterious.

My starting point is to note that the word emergence, like reduction, can be used in a variety of senses which range from the trivial to the technically sophisticated. A rabbit appearing at the mouth of its burrow exemplifies emergence of the former, everyday, kind. But even this example is not quite as banal as at first it appears, and I will return to it later. John Holland begins his 1998 book with the example of the tale of Jack and the beanstalk. Holland suggests that the transformation of the bean into the beanstalk is a demonstration of emergence in a sense that relates to his subsequently elaborated, highly technical, account of the concept. It seems preferable to say, however, that the example combines

both everyday and more technical senses: there is the mere fact that something appears from the ground, and then there is the more mysterious transformative process by which bean becomes plant. Some of the intuitive force of terms like reduction and emergence, when used in technically specialist ways, no doubt trades on ambiguities arising from the possibility of conflation with everyday construals.

An interesting characteristic of Holland's account of emergence is that it takes for granted our ability to know it when we see it. Many accounts are like this: they attempt to address the question of what it is that characterizes the phenomena we recognise as emergent, but take recognition itself to be unproblematic and philosophically uninteresting – presumably because it is supposed that emergent phenomena are unified by their possession of an ontologically robust common property or set of properties. The task then becomes, if one inclines to this way of thinking, to identify the relevant properties. Thus although my particular interest is in the relevance of the concept of emergence to biological systems and consideration of our capacities to explain them, I begin my investigation by reviewing some of the ideas that recur in discussions of emergent phenomena across a range of areas. The diversity this reveals leads me to propose that there is not some one feature, or set of features, that when exhibited by phenomena leads us to describe them as emergent. Rather than asking what emergence is, in terms of objective features of phenomena in the world outside our minds and brains, we should adopt a more epistemically oriented stance. In other words we should ask what the psychological basis is for the ascription of emergence. I contrast our facility for explaining the behaviour of other people (and, to a lesser degree, animals) with the difficulties we sometimes encounter in relation to complex physical systems. This leads into a discussion of imagination, simulation and understanding, from which I derive a view of emergence that goes some way towards making sense of the application of a common term to highly disparate phenomena.

Facets of emergence

Reducibility, deducibility and predictability

Philosophical accounts of emergence, such as that of Broad (1925) outlined in Chapter 4, have often been framed in terms of irreducibility or non-deducibility. Broad was concerned especially with synchronic emergence, or with the dependencies that hold

simultaneously between the properties of wholes and parts of systems. This contrasts with diachronic emergence, which deals with how phenomena unfold over time. A property of a system is emergent, he says, if it cannot be deduced from the properties of its parts as they occur in it or simpler systems, and with this non-deducibility comes unpredictability.[71] Closely related to this, especially if reduction is seen in terms of formal derivability, is the idea that a system property is emergent if it is not reducible to the properties of its parts. For this to be illuminating requires that we already have a satisfactory account of reduction, but as Chapter 1 showed reducibility can mean many things. The formal derivability interpretation has the merit of clarity but in general lacks applicability to biological phenomena, even though we presumably would not say that those phenomena are in general emergent. On the other hand if we construe reducibility as explicability in terms of X then we obtain a concept which has greater potential salience to biology, but which now seems to be almost too weak to be worthy of anti-reductionist ire.[72] (It is possible to see a stronger motivation for that as arising from a *combination* of factors. One is the fact that an organism's phenotype is not fully accounted for by its genotype, even though genes are undoubted difference-makers (Waters 2007). Another factor, more spatial or mereological in character, is the recognition that causal influence does not simply radiate outwards from the micro to the macro, but rather acts bidirectionally.)

One problem with linking predictability and deducibility is that doing so seems to make predictability too dependent on the possibility of formal derivability. The possibilities presented by in silico simulation methods, such as those applied to the study of protein folding and dynamics discussed earlier, substantially decouple prediction from the constraint of mathematical analyticity. Some of Broad's examples, to do with the derivability of the physico-chemical properties of compounds from those of their constituents, are rendered suspect if not actually mistaken by scientific developments.[73] Nonetheless, the relationship between predictability and emergence is interesting, and not

[71] Many discussions of emergence highlight unpredictability as a major characteristic, e.g.: '[o]ne of the characteristics of many complex systems is the phenomenon of emergence in which properties of the system emerge that cannot readily be predicted from a knowledge of the constituents of the system' (Norris et al. 2005, p.313).

[72] Curiously (if we start off by assuming that reduction and emergence are straightforwardly complementary concepts), philosophers of science and philosophers of biology tended not to mention the concept of emergence in the context of the extensive debates about reduction that occurred from the 1950s onwards. If the concept was deemed to be philosophically uninteresting then this may have been because it was assumed to connect more closely with epistemological than ontological matters. If this is the correct explanation then it seems revealing of the philosophical values then prevailing, and suggests a bias which perhaps owed something to a positivist aversion to psychology and a preference for identifying what might be said about the objective nature of things in the world without reference to properties of the mind.

[73] For example, success has recently been reported in simulating the properties of liquid water from quantum mechanical principles (Bukowski et al. 2007).

straightforward. Many things happen in the world that we do not, and could not, foresee, just because of interactions between multiple contingencies in space and time – contingencies of a kind we often expect. In addition, many phenomena we call emergent occur reliably given certain conditions, and hence we expect them to occur.[74] Thus often the issue is not that phenomena are unforeseeable as such, for frequently we do expect them to occur, so much as that they threaten us with unintelligibility, in that we are unable to connect initial or underlying conditions with the resulting phenomena.

Downward causation and levels

Sometimes emergence is associated with downward causation, in which macro-level phenomena 'loop back' to exert a causal influence on, and so partially determine, micro-level phenomena. A classic articulation of the idea is that of Campbell (1974). He describes how the morphological specifics of an ant's jaws can be accounted for in terms of evolutionary selection forces acting on multiple generations of entire organisms. Ultimately these large-scale influences (large-scale in both spatial and temporal senses) serve to determine the DNA sequence of the genome that through developmental processes is associated with the generation of the particular jaw morphology of an individual ant.

Downward causation appears to run counter to the microphysicalist intuition that phenomena at a given level are fully accounted for by properties, processes and events at lower levels. In Campbell's example, population- and organism-level phenomena are seen to play a part in an account of how particular structures at the molecular level come about. Two salient points here concern first the nature and status of his example as a representative of downward causation in general, and second the idea of organizational levels in biology and the ontic and epistemic weight we should attach to them.

Regarding the first point it could be argued that selection is somewhat complex and causally rather distinctive when compared with more proximate biological processes (Mayr 1961). Many of the material phenomena we wish to explain can be accounted for in terms of the continuous or direct interactions and transformations of particles and larger aggregations of matter – it presumably being the frequent realization of this kind of possibility that establishes and reinforces the microphysicalist's intuitions. Evolution by natural selection operates by applying a filter to a population of genome-bearing organisms,

[74] Relevant examples here include the deterministic chaos seen in the dynamics of systems that are indefinitely sensitive to initial conditions.

against a backdrop of slowly accumulating genetic change. The generation of individual organisms does seem to be resolvable in principle, from one perspective, into continuously connected micro-level material events and processes. But the filter of natural selection typically cuts across, i.e. applies simultaneously and without regard to, the levels we standardly recognise, from the whole organism down to its constituent molecules. It effectively deletes subsets of a population so as to change the extent to which a particular class of genome sequences is represented in that population. This change in relative representation level biases the chances that a particular sequence will be propagated to subsequent generations.

The fate of a specific instance of a particular DNA sequence thus depends not just on interactions and transformations at its own spatiotemporal scale (i.e. the molecular) but also on the fate of the organismic container in which it resides – which is a matter of chance and the overall fitness of an organism in the context of a particular environment. Whether a particular sequence carries through to subsequent generations, however, depends on the fates of all the individual organisms in which it occurs. Therefore the presence of the particular DNA sequence that might account for the morphology of a specific ant's jaw may have to be explained by appealing to several rather distinct sets of factors. One is the continuous series of molecular-scale material events and processes (genome replication steps and so on), in principle traceable back to the ant's earliest ancestors, that resulted in a specific genome sequence being present now in the particular ant in question. Another, more negatively, is the history of random drift and selection events that these micro-level material genetic processes survived. Subtly different concepts of downward causation apply to these different sets of explanatory factor.

The sense of downward causation I take Campbell to have intended to emphasize relates primarily to the diachronic determination of the microphysical (DNA sequence) by population-level and macroscopic events and processes. This connects the term with the causally distinctive feature of evolutionary selection noted above: its capacity to 'cause by deletion' (of a subset of a population). Another, more exclusively spatial, sense of downward causation is pervasive, however. It enters into Campbell's account in terms, for example, of the way in which particular selection events may be realized by the consumption of one organism by another (think of an amoeba engulfing a bacterium). But it can also be illustrated rather well by thinking about a protein's conformation as represented by way of the Ramachandran plots described in Chapter 3. When the crystal structure of a protein is analysed to produce a Ramachandran plot, most of its residues are

found to fall within the allowed areas – but there is considerable variation amongst the precise phi/psi values, even amongst residues of the same type. Whilst some of the variation can be attributed to imperfections of crystallographic structure refinement (Kleywegt and Jones, 1996), for most residues local steric factors of course leave considerable latitude as to the exact conformation adopted. So what determines the actual phi/psi values that obtain in the context of a specific protein? One surely has to say: downward causation. To a first approximation, in the context of the folded structure any single residue is subject to the sum of the forces exerted on it by the other residues, and these will compel it to adopt that configuration which is energetically accessible and in which the overall energy of the molecule is minimized – even if as a result the individual residue is distorted into a conformation it would not, 'in isolation', frequently adopt. Proteins thus illustrate in microcosm how wholes can determine the properties of their parts.

Generalizing from this example it is possible to express one kind of downward causation – which could be referred to as *inward causation* – in terms of the way in which the influences, or causal potentials, that act at a particular location or region can come from multiple directions and distances. The way the influences sum and act, together with the nature of whatever exists at that location, determines what happens there. [75] (Influences here often means forces arising from fields, whether acting over long or short ranges. At different scales, different fields predominate.) Certain sorts of effect depend for their realisation on the attainment of a minimum local concentration of causal potential, and sometimes this can only be brought about by a summation of the individual causal potentials of multiple parts. (Think of light being focused onto a small region so as to heat it to its ignition point.) Conversely, parts are compatible – in the sense that they can retain their integrity or identity – only with particular contexts, within bounds beyond which they are changed. The conditions under which, and the manner in which, they are changed depend in complex ways on energetic and other factors. Highly non-linear behaviours can thus be generated: if a causal threshold is met, X happens; if it is not met, X does not happen.

I mentioned above that the topic of downward causation raises the issue of organizational levels and how we should regard them. In this connection it is worth bearing in mind that biological causes may be highly dispersed. Extracellularly, fluid media such as

[75] This sense is the same as that implied by the title of Abraham Pais' book on the history of atomic and high-energy physics, *Inward Bound* (Pais 1986).

the bloodstream can be used to broadcast a wide variety of signals to different cellular targets with varying degrees of specificity. Intracellularly, causes may similarly be distributed, taking advantage of the properties of the intracellular environment in order to exert their effects. Should these kinds of contingently acting, message-based phenomena be grouped under the heading of downwards causation? Arguably not, to the extent that we can think of the messages and their targets as lying at the same structural level, i.e. the molecular. On the other hand one might argue that such cases go some way towards undermining the concept of levels altogether. In an animal, for example, signalling molecules – a hormone say – might originate from the cells of one organ, such as the pituitary gland, and travel around the body to the cells of other organs where – if those cells bear the relevant receptors – they set in train molecular processes that might ultimately result in the expression of a particular gene. The pituitary gland cells may have been stimulated to secrete the hormone as a result of the collective activity of neurons distributed across a variety of cortical levels, including those associated with higher (more abstract) cognitive skills – neuronal activity which could well have been triggered by the organism finding itself in a particular kind of environment.

Another example which shows how causal explanations for phenomena often cut across level boundaries involves protein unfolding rather than folding. It has been proposed that cells such as fibroblasts can exert enough force on matrix fibrils to stretch them, resulting in the partial unfolding of their constituent fibronectin molecules. This brings about the sequential exposure – creation would be a better word – of binding sites for specific signaling molecules. The extracellular matrix, rather than being a passive support, enters into dynamic patterns of reciprocal causation with the cells it surrounds. Cells are able to sense the state of the matrix (since they are coupled to it via cell membrane receptors which in turn connect with intracellular signalling pathways) and alter their patterns of gene expression in order to regulate their environment (Smith et al. 2007). It is hard to see how thinking in terms of levels helps much in these and many other kinds of case; more helpful is to try to track the flows of causal influence and see how they propagate and act through space and time. And it is sometimes useful, in that regard, to think in terms of inward versus outward causation.

Agents, environments and parallel processes

In recent years the concept of emergence has frequently arisen in the context of debates about systems consisting of multiple agents and the collective behaviours to which

their interactions give rise. Termites, for example, are individually rather simple, with important aspects of their behaviour being interpretable in terms of a relatively small number of rules that key their actions to contexts. Despite this individual simplicity termites are collectively capable of sophisticated adaptive behaviours, such as the construction of complex mound structures (Clark 1996). Cellular processes – including, or perhaps especially, the fundamental capacities of metabolism, genome replication and cell division – may be thought emergent in a sense that is rather like this, for they result from the combined operation of numerous apparently independent activities. However, it is becoming increasingly apparent that many processes are in fact highly integrated, as was noted in the previous chapter. In Chapter 2 I discussed how those activities are implemented by molecular mechanisms – 'protein machines' – of sometimes surprising sophistication. A common feature is that the actions of individual agents or mechanisms appear not to be directed in any readily apparent manner towards the realization of the ultimate ends to which they do in fact contribute. An analogy is with a musical performance in which multiple instrumental parts are played simultaneously (Noble 2006). The overall melodic, harmonic and rhythmic effects for which the composer has striven arise – sometimes, and to an extent that varies according to the composer and the composition – out of the simultaneous performance of the different parts, but cannot necessarily be identified with any part in isolation.

Patterns and surprise

Musical analogies suggest another aspect of the phenomena we describe as emergent: the appearance of patterns or order of some kind. Sometimes the order is straightforwardly visual or spatial, as in the Belousov-Zhabotinsky reaction, in which the existence of alternative reaction pathways gives rise to spatial patterns that are analogous to the wave-like growth patterns exhibited in certain circumstances by the slime mould *Dictyostelium discoideum* (Peacocke 1989, pp.40-41). Sometimes however it takes the form of mathematical regularity or lawfulness that is apparent only under certain forms of analysis. The fact that the frequencies of words in a natural language corpus are inversely proportional to their rank order in a table of word frequencies (a relationship also referred to as Zipf's Law[76]) is an example of this kind of 'emergence under analysis'.

[76] Thus the most commonly occurring word will occur roughly twice as frequently as the second most commonly occurring word, which will occur twice as often as the fourth most common word, and so on. The law is named after George Zipf, who discussed the phenomenon in his *Human Behavior and the Principle of Least-Effort* (1949).

Why the appearance of order, symmetry, repetition or lawfulness in phenomena should strike us as significant is a fertile topic for speculation. Perhaps there is an entropic explanation: entropy favours disorder, and a lack of disorder implies the operation of some organizing force or causal principle that if understood or mastered might confer powers to predict or control. However, some spontaneous processes tend to result in homogeneity and isotropy (the dissolution of a substance in a solvent is an obvious example), and this could be characterized as a high degree of symmetry. In many cases little significance is attached to the symmetry associated with such uniformity; it is as if the significance we attach to patterns and symmetries reflects what we know or feel justified in assuming about their etiology. When and why significance is attached to order and pattern are thus somewhat problematic issues and it is perhaps unwise to generalize from specific cases. Nonetheless, it seems mistaken to say that emergence attributions necessarily depend on the appearance of patterns. The oscillation of a single pendulum exhibits a simple kind of order that we would not think of as demonstrating emergence. However, when a second pendulum is coupled to the first the regular periodicity is disrupted, to be replaced by chaotic oscillatory behaviour – and that behaviour probably would be described as emergent.

The pendulum case shows how an element of unexpectedness is often one of the hallmarks of phenomena we describe as emergent, leading some to talk of emergence as 'the science of surprise' (Casti 1994, 1997). Often, however, the surprise is of only an initial or second order character. The configurations of cell occupancy seen in John Horton Conway's well-known 'Game of Life' cellular automaton (Gardner 1970), or the patterns characteristic of the Belousov-Zhabotinsky reaction, occur reliably subject to certain conditions being met, and so we come to expect them irrespective of our capacity to explain them. Much of the surprise is connected with the *nature* of what occurs, be it the appearance of order or its disruption. However, when rule-following agents interact with novel environments and emergent behaviours are observed, as can occur in robotics (Ronald and Sipper 2001), perhaps the surprise is rather *that* something unforeseen occurs.

Divergent and convergent processes

Some authors have used the term emergence to describe the development of material complexity in the universe, concomitant with its expansion following the Big Bang

and the decrease in energy density this has brought about. Nagel sees this 'cosmogonic' sense of emergence as being fundamentally different from the sense of emergence implicit in interpretations that seek to tackle issues arising from the way in which systems are organized (Nagel 1961, pp.366-380). Often, however, it is unclear in which precise sense a phenomenon should be judged emergent, in part because it is difficult to give either sense a precise articulation. Nonetheless it is clear enough that when an author writes about 'The Emergence of Everything' (Morowitz 2002) an interpretation that connects closely with Nagel's cosmogonic sense is intended.

This sort of perspective overlaps with the work of the emergentists of the early twentieth century such as Samuel Alexander ('Space, Time, and Deity', 1920) and C. Lloyd Morgan ('Emergent Evolution', 1923). Emergence here can be seen as a conflation of appearance (as per the earlier example of the rabbit appearing at the burrow entrance) with novelty – echoing the point I made at the outset about Holland's Jack and the beanstalk example. But in addition the emergence claim implies that the novel phenomena are more complex than those that have already appeared, and there is an implication too that the novelty is of a metaphysically radical kind (Van Gulick 2007). Alexander regarded consciousness as an example of this kind of radical novelty. I agree about that phenomenon's distinctiveness, although it is not clear to me how this should – or whether it usefully can – be related to issues of material complexity in non-mental systems.

It may be possible and / or important (so far as understanding underlying causal mechanisms goes) to distinguish here between 'divergent' and 'convergent' processes. The outcomes of processes of the former sort are sensitive to the particularity of events, and evolution appears to demonstrate this property. If a population is reduced to a single potential breeding pair then the fate of the species and the possibility that it will give rise to successor species turns on the fates of just two individuals, which may in turn depend on a host of contingent environmental factors. Gould has asked what would happen if the 'tape' of evolution were 'played' again (Gould 1989), and if nothing else the question tests our intuitions about determinism.[77] However, it seems (to me at least) easier to entertain the idea that the outcome of each 'run' will be significantly different than the possibility that we will repeatedly observe the same outcomes, unless conditions are identical in all respects for all runs or unless the total system in which evolution takes place is extremely small and

[77] An extensive debate has grown up around this question and the issue of whether there are laws in biology (see e.g. Fontana and Buss 1994 and Beatty 2006), but as its twists and turns are not central to my present concerns I sidestep them here.

simple. The idea of there being identical conditions between runs might make no sense, if for example there were a possibility of random fluctuations at the quantum level 'leaking upwards' to give rise to macroscopic variations. The second condition – that the system is simple – effectively appeals to an entropic argument: if the state space that evolutionary processes explore is sufficiently small then the probability that the same configuration will be generated within a certain number of runs may be non-negligible. Obviously, the chances are increased as the number of runs increases. It is in this sense – of extreme sensitivity to initial conditions – that the emergence to which evolution gives rise may perhaps be considered divergent.[78] Convergent processes, on the other hand, are those that are robustly indifferent to minor perturbations, and they include many of those that figure in the life of the cell and the individual organism. Development of the individuals of a species, for example, typically results in the attainment of highly characteristic forms despite taking place in sometimes quite markedly different environments.

It is by now I think clear that emergence is a term that is applied to very different kinds of phenomena, and it is not at all obvious what it is that unites them. Perhaps further work *will* result in a unified framework that succeeds in encompassing the diversity of emergent phenomena while referring solely to properties of systems that are independent of the psychological capacities exercised when we contemplate or interact with them. The phenomena to be subsumed by such an account include the interactions of numerous, individually rather simple, entities at one extreme, and the complex interactions of a small number of highly structured, adaptive agents with a structurally and dynamically rich environment at the other. The account will have to handle, in biology, the play of historical contingency that over time gives rise to the increasingly differentiated and complex structures generated by evolutionary processes, the constancies and repetitions of metabolism, and the canalised unfolding of form seen in morphogenesis. The project of developing an ontologically unified account of emergence capable of addressing all these diverse phenomena is not one I shall attempt, however, since I think there is a simpler, more satisfactory account to be given. Although this account has a strong epistemological component, my starting point is the same as that from which such a more exclusively ontological project would probably proceed: consideration of the causal underpinnings of emergence.

[78] The intuitions I have expressed here are presumably susceptible to testing by in silico modelling and experiments in artificial life.

Causal comprehension and imaginative simulation

It is presumably an uncontroversial claim about emergence to say that it relates in some way to our capacity to account for phenomena in terms of antecedent, underlying or surrounding conditions. As Mark Bedau has observed, emergent phenomena often exhibit a paradoxical duality, being both dependent on and yet apparently autonomous from their causal foundations (Bedau 1997). This can make levels talk, for all its difficulties (already noted), hard to avoid. The paradox, frequently, is the existence of macro-level phenomena that appear to conform to different ordering principles from those obtaining at the micro-level (one might say the levels are defined by the distinct sets of ordering principles). Shifting attention from one level to the other involves making something like a gestalt switch, since the connection between the macro-level phenomena and their micro-level underpinnings is not apparent.[79]

Our seeming inability readily to comprehend in an explanation-supporting way the causal processes underlying emergent phenomena – in general despite the visibility of the entities and processes on which they depend – contrasts rather strikingly with our ability to provide quite satisfactory accounts of human behaviour across diverse conditions. We suppose that such behaviour is underpinned by the workings of a neuronally instantiated mind, but our account-giving capacity works despite the fact that the relevant neurological processes are invisible to us. Instead we are frequently able to make some sense of the behaviour of others by ascribing certain beliefs, desires, and so on, and by relating these to what we know about, for example, an individual's past experiences. Being able to explain someone's behaviour depends on that behaviour being consistent with what we know about the person in question and with what we know about human nature in general. Sometimes we will describe someone as having a particular type of personality, or as exhibiting a specific behavioural trait. Attributions made on the basis of these kinds of inferred dispositional structures – which I take to be a reasonable way of thinking about personality types and traits – act as significant modifiers or determinants of the kinds of behaviour we see as making sense in relation to a specific individual (or class of individuals) in a particular set of circumstances. The whole conceptual framework by which we rationalize and account for the behaviour of others in terms of 'propositional attitudes' such as beliefs and desires is often referred to in philosophical circles as folk psychology.

[79] I am grateful to Francesco Guala for this observation.

Debates of the 1980s and early 1990s concerning the nature of folk psychological explanation (see e.g. Davies and Stone 1995) have in recent years evolved into debates about what has been referred to as 'mindreading' (Nichols and Stich 2003). There are substantial elements of continuity to the debates, although the emphasis has shifted. Where once there was a preoccupation with whether and in what sense folk psychology could be considered a theory (Churchland 1988) there is now much greater interest in the role and status of mental simulation, imagination and pretence (Goldman 2006; and see also Nichols 2006). One of the probable reasons for this shift is that in the intervening time philosophy of mind has been reshaped by the development of a new interest in consciousness (Searle 1992; Chalmers 1996) and in issues of embodiment and emotion in relation to cognition (Damasio 1994, 1999; Noë 2004). Additional impetus, especially on the simulation side, has come from the discovery of 'mirror neurons', first in monkeys and then in humans, that fire both when a subject performs a motor activity and when the subject views another individual performing the same activity (Gallese and Goldman 1998). [80]

Foreshadowing the debates about imagination and mental simulation to which such findings have given new impetus are earlier speculative accounts concerning the part played by imagination in our ability to interpret other people's actions within a folk psychological framework. Morton (1980) suggests that an important strategy for making sense of someone's behaviour is to situate their actions within a behavioural schema that shows how they result from being in a particular psychological state. The schema provides a basis for internal simulation, aspects of the schema being adjustable to achieve congruence between the simulation and what is perceived. It is not entirely clear how simulation here should be understood, but it seems reasonable to assume that an involuntary, non-deliberative process is intended. Congruence between the simulated behaviour that underpins our interpretation of another person's behaviour and that person's actual behaviour is possible, I take Morton to be proposing, in virtue of the fact that both simulated and actual behaviours are causally structured in accordance with the same psychological schemas. This sort of interpretative mechanism presumably thus exploits similarities between our psychological structures and those of our fellow humans, and no doubt also involves tight corroborative loops of language, action, and perception.

[80] The existence of mirror neurons in humans is still a contested area, however (Lingnau, Gesierich and Caramazza 2009; Kilner et al. 2009). So too is the issue of how such 'resonance systems' might adjudicate between simulation- versus theory-based theories of mind (Gallagher 2007).

Cautiously we can speculate that behavioural interpretation, understanding and prediction involves a range of mechanisms, often working in parallel and varying in terms (for example) of degree of representational abstraction and the extent to which a mechanism's operation is associated with conscious experience. The discovery of mirror neurons appears to lend weight to the idea that some of these mechanisms are based on capacities to simulate the behaviour of others, or 'put ourselves in their shoes' (see e.g. Buckner and Carroll 2006). In the present context I want to entertain a related but more general idea: that making sense of the world often depends on organized cognitive dispositions – schemas – of various kinds that allow us to model its entailment structures by mentally simulating them.

Imagination and scientific thought

Let us suppose then that abilities to imagine and simulate play a part in comprehending not just the behaviour of our fellow humans – and the behaviour of other species too, to the extent that it conforms to schemas that stand in some reasonably direct relation to folk psychology – but phenomena more generally. A plausible conjecture then is that scientific understanding is sometimes a matter of being able to imagine the behaviour of material structures and systems through processes that are underpinned by cognitive schemas of various kinds. Scientific communication can be seen as reflecting this. Much scientific writing clearly aims at evoking images in the mind of the reader or conveying patterns of causality. One way of viewing mathematical expressions in physics is as communicable data structures associated with the operation of schemas for encoding and manipulating the quantitative relationships that obtain within actual or counterfactual scenarios.

Further evidence of a role for imagination in scientific thought comes from a number of directions. First there is plain introspection: when I think about how a protein might bind to DNA, say, or about the migration of endoplasmic reticulum to the cell surface, verbal processes seem very much secondary to visuospatial ones. This introspective experience resonates with the first-person accounts reported by Jacques Hadamard (1945) that accord a central role to a visuospatial cognitive mode in mathematics, and with recent work in the epistemology of mathematics (Giaquinto 2007). In the course of his discussion of the role of visuospatial and mechanical thinking in engineering Eugene Ferguson presents evidence gathered from a variety of sources across a

range of sciences (Ferguson 1992). He specifically mentions Galton, Boltzmann, Einstein, Bohr, and Heisenberg in connection with visual thinking, and notes the tendency of Faraday, Kelvin, and Maxwell to think in terms of 'models and mechanical analogies' (pp.44-45). Regarding the latter tendency he discusses Pierre Duhem's suggestion that British and French physicists of the nineteenth century adopted different styles of thought, reflecting wider cultural differences of cognitive approach.

Duhem identified in the British mind an 'extraordinary facility for imagining very complicated collections of concrete facts ... and an extreme difficulty in conceiving abstract notions and formulating general principles'. In contrast, the French mind he saw as 'strong enough to be unafraid of abstraction and generalization but too narrow to imagine anything complex before it is classified in a perfect order' (Duhem 1954/1991, p.64). These distinct styles were reflected in different approaches to, for example, electrostatic physics:

> This whole theory of electrostatics constitutes a group of abstract ideas and general propositions, formulated in the clear and precise language of geometry and algebra, and connected with one another by the rules of strict logic. This whole fully satisfies the reason of a French physicist and his taste for clarity, simplicity, and order.
>
> The same does not hold for an Englishman. These abstract notions of material points, force, lines of force, and equipotential surface do not satisfy his need to imagine concrete, material, visible, and tangible things.
>
> (Duhem 1954/1991, p.70)

Duhem goes on to note that in Oliver Lodge's treatise on the theory of electricity 'there are nothing but strings which move around pulleys, which roll around drums, which go through pearl beads, which carry weights...' Despairingly he says that 'We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory' (p.71). It may be that Duhem's distinction can itself be thought of as a manifestation of the very tendency he articulates so sharply, but at the very least it highlights the possibility that scientific activity is compatible with different styles of thought. That the scientists Ferguson mentions in connection with a visual-mechanical style of thought are (Galton apart) mostly physicists and physical chemists if anything lends additional significance to the reports. For if visual thinking plays a part even in physics, where thought is often assumed primarily to be mathematical, its role in biology – in which the visualization of phenomena is often the principle aim of research – must (it could be argued) be considerable. Of Einstein, Ferguson notes that he said how he ' rarely thought in words at all; his visual and "muscular" images had to be translated "laboriously" into conventional

verbal and mathematical terms' (p. 45). The phrase 'muscular images' here echoes MDC's proposal that the intelligibility of a phenomenon is not exclusively a visual matter but may incorporate other aspects of bodily experience.

How are first-person reports of a visuospatial component to scientific thought to be understood? A variety of findings in cognitive psychology point towards possible neuropsychological underpinnings. Roger Shepard famously noted that the time it takes for an individual to judge whether the objects shown in two pictures have the same shape is proportional to the angular difference in depicted orientation of the objects (Shepard and Metzler 1971). The two depicted objects have roughly the same shape, and it is as if the subject answers the question by mentally attempting to rotate one of the objects in order to bring its major morphological features into the same orientation as those of the other so that they can be superimposed and compared. From this kind of work has arisen a major strand of research in cognitive science and psychology concerning the part played in cognition by imagery and visual thought (e.g. Kosslyn 1994; Cornoldi et al. 1996; de Vega et al. 1996; Stenning 2002). Such work can be seen in part as working alongside fundamental research in neuropsychology concerning the participation of different brain regions in particular mental processes. The evidence that the rear portion of the neocortex is involved in visual processing is now overwhelming, for example, and much of the neuronal basis of visual perception is well understood (see e.g. Zeki 1993).[81] Against this neuropsychological background cognitive psychological models of visual thought can be regarded as a layer of conjecture capable of – and necessary for – guiding research by prompting particular questions which can then be tested by experiment.

Given the conjectural character of the ideas I have outlined to do with simulation and imagination it is necessary to preface any tentative philosophical theory of emergence based on it with some words of qualification. In the first place there is little doubt that individuals differ – irrespective of their nationality! – in terms of the reliance they place on different cognitive modes in different situations (Galton 1883; Hadamard 1945; Reisberg et al. 2003). Secondly, in performing different tasks we presumably switch between, blend or integrate imagistic, mechanistic, linguistic and other cognitive styles as necessary; and sometimes we appear to make use of mental models that combine aspects of logical and abstracted visuospatial thought (Johnson-Laird 1983). And thirdly it should be noted that I do not mean to imply that we can take the findings and ideas from cognitive psychology in

[81] To the point where it is increasingly possible to 'read' a subject's perceptual experience from brain fMRI scans (Haynes and Rees 2005; Kay et al. 2008).

these areas for granted: the imagery debate remains alive and well after several decades of argument and counter-argument (see for example Kosslyn et al. 2001; Pylyshyn 2003 and references therein). However, proceeding on the basis that a certain view may be roughly right at least exposes that view to further testing; and success in clarifying specific problems might, on an abductive basis, count in the view's favour. We should not suspend trying to make sense of the issues that interest us until, for example, the nature of the mental representations that underlie phenomenal experience has been resolved. The resolution of such questions may actually be facilitated by work that incorporates assumptions about the likely answers in attempting to make sense of other problems.

Cognition and material systems

It is a truism that our cognitive abilities are not unlimited, but still it is arguable that we lack a clear sense of the shape of those abilities and of the directions and degrees of limitation. George Miller (1956) famously noted constraints on the number of entities, variables or dimensions we can process simultaneously. These constraints are generally interpreted in terms of the limited capacity of working memory. Cognitive psychologists are exploring further the nature of many of our cognitive tendencies and biases, for example in relation to the attribution of functional properties (see e.g. Lombrozo and Carey 2006). A significant constraint on our ability to reason about complex systems appears to be what Resnick has referred to as a 'centralized mindset' (Resnick 1996), or the tendency to seek causal foci and concentrated causal sequences of events. It is, I take it, an open question to what extent it is within our powers to overcome this tendency. Sometimes we reveal other biases when thinking about complex systems, for example the tendency to think of causal power as dissipating as it propagates through complex systems such as ecosystems (White 1998). Perhaps we do so on the basis of inappropriate analogies with phenomena like, say, thermal conduction or sound propagation. Such systems clearly have the potential to tax our imaginative and inferential capacities heavily, and often exhaust them altogether. In contrast, certain sorts of system are much more amenable to comprehension.

Amongst the more epistemically tractable of material systems are the the machine-type artefacts I discussed at some length in Chapter 2. To recapitulate, these consist on the whole of solid-state components in well-defined spatial relationships, capable of moving relative to one another in a limited number of ways defined by points, lines and planes of articulation established by stable part shapes. The number of degrees of freedom is

extremely small relative to the number existing in the equivalent amount of matter in gaseous or liquid form. Psychological experiments suggest that we are able to simulate mentally the operation of simple mechanisms such as cogs and pulleys in order to trace how causal influences propagate through them (Hegarty 2004). It seems reasonable to speculate that the simulation process in such cases draws upon the same kinds of mental processes as those putatively involved in Shepard's work on the mental rotation of objects. Simple localized transformations of parts – principally rotations and translations – can be imagined and we can be confident, because of the properties of the solid state, that these do not have non-local consequences. Thus (it can be conjectured) the capacity of working memory is not exceeded.

Simulation and causal schemas

Many issues remain to be worked out in the picture I have outlined so far, which is obviously sketchy. Quite what is meant by mental simulation could do with clarification, for example. In talking about imagination, to what extent do we mean a faculty that is associated necessarily with conscious experience? Is it exclusively pictorial or imagistic, or can it draw upon or involve different dimensions, relating perhaps to specific perceptual or motor skills? (Einstein's talk of 'muscular images' suggests a likely answer to this question.)

These are hardly trivial issues, and it would probably require a book-length treatment to do them justice. But I can describe the sort of philosophical and psychological framework that would need to be fleshed out if the ideas outlined here were to be developed into a comprehensive account. The central notion is that our interactions with each other and the world are continuously guided by presumed patterns of entailment or schemas of various kinds, and these schema-conformant thoughts represent, or are associated with, expectations about how events in the world will unfold. (Research such as that reported by Bar (2007), Kveraga et al. (2007) and Schubotz (2007) provides inspiration and evidence for the plausibility of this line of thinking.) Schemas may be instantiated directly by the perceptual detection of features in phenomena, indirectly by the association of such perceived features with memories which then trigger schemas, or may be triggered purely by reflective – imaginative – processes. Causal schemas model patterns of entailment, and our minds are continually attempting to fit experience to schemas to guess the future (Sloman 2005; Frith 2007). Surprise, on this kind of account, is associated with a failure to find causal schemas that generate accurate predictions.

Morton's account of behavioural interpretation by imaginative synthesis is about the attempt to fit social experience to folk psychological schemas. As he says, this is largely a subconscious matter, but overt imaginative and reflective processes can also play a part. So it is, I suspect, with the schemas we have for making sense of the material world. As we move about in, interact with and observe the world we are guided (I conjecture) by streams of expectation generated by the causal schemas triggered by what we experience from moment to moment. Expectations are largely invisible, so long as they are gratified by experience. The causal schemas are, I assume, acquired through play and learning, although some may be innate. As we get older we learn progressively more abstract schemas, come to see how these connect with more primitive ones, and automatize associations between them.

Being knowledgable, on this kind of account, amounts not so much to having sets of justified true beliefs as possessing apt schemas, and perhaps we could say, à la Nozick (1981), that apt schemas are ones that track the truth. Expectation takes the form of being oriented towards the world in ways that presuppose the development of events in a particular manner. An important indication of a certain sort of understanding then consists just in not being surprised by events. This is not the passive affair it might sound like, however, but a hard-won state based on keeping track of sensed experience and comparing it with what we assume or imagine will happen. Mental simulation I think can be regarded as a process by which we find a causal schema relevant to an imagined system (e.g. a configuration of material elements) and use it as a basis for generating an expected future system state. That new state is then the basis for selecting (or just triggers the instantiation of) a new schema – if the first one is no longer relevant – in order to generate the next state, and so on. These various stages of selection and generation proceed, presumably, automatically and subconsciously, but may generate visual outputs (and perhaps can utilize the results of visual operations as inputs). A phenomenon is imaginable if we can trace through it, and is intelligible if we can keep in step with it, using the schemas at our disposal in order to coordinate the states of the system at different times.

I said earlier that I would return to the rabbit at its burrow entrance. What makes that case non-banal is that we cannot actually claim to know that the rabbit at the entrance was a rabbit when it was in the depths of the burrow. It may have been a dragon, and the burrow may not be what we suppose either; these things are matters of belief. But we get through life by making hosts of assumptions, on the basis of what seems most likely and

most consistent with our perceptions, memories and various interpretative schemas. We imagine – as a matter of cognitive parsimony – that before we saw it at the entrance the rabbit was in the burrow, being a rabbit. Hence we do not interpret its appearance as an instance of any technically interesting sense of emergence.

Emergence, simulation and epistemic prostheses

These rather sketchy ideas about schemas, simulation and expectation provide a basis for thinking about emergence. In a nutshell, the concept of emergence can be seen as pertaining to phenomena that we have difficulty connecting with initial or underlying conditions. The difficulty arises when we are unable to find or construct a causal schema that fits the case, and when we describe something as emergent one of the things we are doing, in effect, is providing a report on our own cognitive state: we are acknowledging that we lack the cognitive resources to trace through, see connections within, or simulate the causal structure of some aspect of the world.[82] It may be that the requisite representations are too complex, for example because they contain more adjustable parameters than we are capable of manipulating simultaneously (because working memory becomes overloaded, perhaps). Or, as with the simple chaotic oscillator, behaviour evolves in ways that we are unable to parallel adequately in folk physical or other imaginable terms because we cannot represent or manipulate the physically important quantities in a sufficiently precise manner. An analogy is with the generation of an exception by a computer program, and we can of course speculate about the nature of the exception-raising mechanism(s). Perhaps it is simply not possible to derive a viable causal schema on the basis of what is known about the phenomena – there is nothing for a putative simulator to 'run'. On the other hand maybe a schema can be found, but when run it 'crashes' the simulator because of inconsistencies of some kind. The possibilities seem limited mainly by our ingenuity at constructing psychological models.

Whatever the psychological facts, we can say that to the extent that we can see how they arise, we tend not to think of phenomena as emergent. We can see how phenomena arise if we can access causal schemas that are compatible with our cognitive constraints.

[82] Perhaps Nagel's cosmogonic sense of emergence (p.133) can be understood in terms of an inability to foresee the consequences of causal structure, often on account of highly contingent and non-linear relations amongst events, while many other uses of the term advert to an inability to model the causal connections between observed outcomes and prior or underlying conditions. An additional constraint on this kind of ascription of a non-cosmogonic sense of emergence to a phenomenon may be commitment to the idea that all the relevant causal factors lie within a particular spatiotemporal range.

And such schemas can be pretty rough-and-ready yet still be conducive to a sense of understanding if by entertaining them we can avoid leaving glaring explanatory gaps. (See Keil (2003) for discussion of our ability to 'paper over the cracks' in our understanding.) This kind of perspective does not provide an objective, observer-independent method of classifying phenomena as emergent or not. Instead it suggests that if different individuals categorize in the same way then it is because they share the same perceptual and cognitive propensities and capacities – and because they share similar sensitivities to linguistic and conceptual cultural conventions.

Ontological emergence

An objection to the perspective just presented is that it seems to cast emergence in an entirely epistemic light, inasmuch as emergence attributions are seen to be made in accordance with a psychological criterion. But this is not to say that emergent phenomena have no ontological grounding, if we grant that ontological weight is conferred by the possession of irreducible causal capacities (Silberstein and McGeever 1999, p.182). Here we should think back to the discussion of molecular dynamics (MD) simulations in Chapter 3. Imagine that one could perform a colossal MD simulation in which there were polypeptide chains, DNA, water molecules, and so on and so forth. Imagine too that we could run the simulation for as long as we liked, with time steps as small as we liked, and employing completely accurate interatomic potential functions, realistic control of thermodynamic parameters, etc. Given these provisos we would presumably expect that, ceteris paribus, the polypeptide chains would fold into native protein structures, and maybe we would actually see them binding to the DNA.

This example may be far-fetched, but it is not absurd. The point is that such a simulation – which I take it amounts to a virtual re-enactment of what we suppose happens in the physical world – would presumably give rise to evidence for equivalents of many of the causal capacities we would expect to see manifested by this kind of molecular system. An important qualification is that the results would be subject to the limitations imposed by the simulation's boundedness, to acknowledge the possibility of downward/inward causation. Sometimes, for some phenomena to be generated, the spatial bounds will have to be set so large as to make simulation over anything exceeding extremely small times scales impracticable. And an MD simulation would only go so far in mimicking reality, since it would not without suitable modification model the occurrence of chemical

reactions involving covalent bond breaking and making. To do that would (I take it) require the incorporation of quantum mechanical theory into the simulation, and even a very small simulated system might exhaust currently available computational resource. (It might even be that the only practicable system for modelling a particular physical system would be the physical system itself.) If we imagine that we had all the computational resource we needed, we generally reckon (I suggest) that scientific theory would enable the simulation of the molecular events that occur in biological systems. Thus we have confidence in our theoretical understanding, even though we cannot foresee what phenomena the über-simulation we are envisaging would generate (just as it rapidly becomes hard – and then impossible – to foresee future cell configurations in a Game of Life from knowledge of the current state and Conway's rules as the number of cells increases).

Mark Bedau has developed a computationally informed account of weak emergence, and it is interesting to compare it with the ideas about emergence sketched here. A phenomenon is weakly emergent, according to Bedau, if it can be predicted only by simulation.

> If *P* is weakly emergent, it is constituted by, and generated from, the system's underlying microdynamic, whether or not we know anything about this. Our need to use a simulation is due neither to the current contingent state of our knowledge nor to some specifically human limitation or frailty. Although a Laplacian supercalculator would have a decisive advantage over us in simulation speed, she would still need to simulate. Underivability without simulation is a purely formal notion concerning the existence and nonexistence of certain kinds of derivations of macrostates from a system's underlying dynamic.
>
> (Bedau 1997, p.379)

In a recent paper Bedau attempts to go further and show that emergent phenomena are 'real and objective phenomena' (Bedau 2008). If they are 'just in the mind', as he puts it, then emergent phenomena 'are not real and objective phenomena; they have no independent ontological existence; they have no independent causal power, they have no objective reality outside the mind' (Bedau 2008, p.444). He argues that emergence arises from when the network of micro-causal interactions on which macro-level phenomena depend becomes sufficiently complex that no formal short-cuts exist for deriving the macro from the micro:

> weak emergence results from incompressible macro-level structure in the network of micro-level causal connections. This causal web is embodied and brought to life in real ontological substances with real causal powers, and it really generates certain macro-level ontological and causal phenomena.
>
> (Bedau 2008, p.451)

Hence 'weak emergence is not merely in the mind' (p.457), which for Bedau is equivalent to saying that it is not merely epistemological (p.451).

Bedau's position and mine overlap substantially, although it should be obvious that he is much more disinclined than I am to say that emergence reflects our cognitive limitations. And the position against which he musters his argument, which is the view that emergence might be considered to be 'merely in the mind', has something of the character of a straw man. Surely no one ever thought that emergent phenomena were *just* in the mind? For if they were merely in the mind then they would amount to little more than fantasies or delusions, and we might never agree amongst ourselves about what looks emergent. If emergence were just in the mind presumably anything could look emergent, which is plainly not the case. We apply the term to particular phenomena, and the problem is to understand the basis on which we pick out some phenomena and not others. I have already noted the extreme ontological diversity of emergent phenomena, which ranges variously across properties, structures, dynamics, behaviours, patterns, and laws. This is potentially a problem for Bedau's account: as he notes, his interpretation of the phrase 'weak emergence' does not apply to some of the phenomena that have been described as weakly emergent by others (p.445, n6). He also grants that complexity-related weak emergence is a matter of degree (p.447). But in this case one has to ask what determines the point along the ontological scale – if indeed there *is* some unitary scale – at which we switch from not seeing emergence to seeing emergence.

My proposal is that the principal basis on which we pick out emergent phenomena is a negative one: an inability cognitively to model a pattern of entailment within a system. When the complexity of a phenomenon in some respect exceeds a particular level, our cognitive schemas become incapable of tracking or paralleling its causal structure (or as Bedau might put it, they become insufficient for 'crawling the micro-causal web' (Bedau 2008, p.446)). *This* is when we see emergence. A virtue of my account is that it provides the means to account for the attribution of emergence to radically different kinds of phenomena, in respect of which no ontological unity need exist. This will be so if making sense of the causal structure of phenomena involves multiple sorts of cognitive resource, working either in tandem or independently. Imagine that causal comprehension of a phenomenon of ontological kind P1 necessarily draws on cognitive resource C1, while comprehension of phenomena of ontological kind P2 draws on resource C2. Then P1 will look emergent if the capacity of C1 is exceeded, and P2 will look emergent if C2 is likewise

over-stretched. But note that P1 and P2 need have nothing in common, ontologically speaking. Such an account has both epistemic and ontic aspects, in that we classify certain phenomena as emergent because of how their objective physical natures interact with our psychological capacities. The emergence attributions of others are intelligible to us when we make sense of the causal structure of the world using the same or similar cognitive schemas.

I also think that Bedau is mistaken (1) to downplay the epistemic status of simulation, as he appears to do in his (1997) when he argues that the Laplacian supercalculator 'would still need to simulate', and (2) to objectify ('a purely formal notion') the nature of 'certain kinds of derivations'. I take the latter to be primarily logico-mathematical derivations, which as he argues are unobtainable in the case of emergent phenomena. I suggest, however, that we simply do not know enough about mathematical cognition to base an argument on the idea that there is a class of phenomena in the world involving relations that are amenable to formal treatment and another class involving relations that are not so amenable, and on the idea that the amenability in question is a mind-independent matter. Logico-mathematical reasoning is presumably underpinned by neural processes that correlate with particular patterns of thought, and as a contingent matter these cognitive entailment structures or schemas are capable of (and useful for) paralleling particular phenomena in the world or expressing relationships between them. But the contingency is key to the present case: it is again whether or not we have schemas that are projectible onto phenomena in the world that determines how we individuate and classify those phenomena.

A jointly epistemic and ontic account of emergence along the lines proposed here is not necessarily incompatible with the possibility of developing a less 'internalist' account of emergence (i.e. one framed in more exclusively mind-independent terms), based perhaps on some agreed causal complexity metric. But it seems unlikely that, programmed with such an account, some sort of automated system analyser would diagnose emergence in just those circumstances where humans tend to see emergence. For these are situations in which a phenomenon arises in a way that we are incapable of simulating or modelling on the basis of causal schemas that I suggest are numerous, diverse, and not necessarily systematically related to each other.

7. Functional attributions and the causal status of the genome

Introduction

Attention was drawn in the previous chapter to the complexity of the relationships that sometimes hold between ontology and epistemology, for in it I argued for an account of emergence framed jointly in ontic and epistemic terms. If I am right then the phenomenon can be seen as arising from the way in which how things are in the world interacts with our psychological constitution. This means that emergence as we experience it may be deemed a largely epistemological matter, yet reassuringly it is at the same time not without any ontological basis.

In this chapter the accent remains on the metaphysics of biological explanation. The principal topic is the nature of functional attribution, which I investigate by considering the causal status of the genome. The classic view, which has structured work in molecular biology for roughly five decades, is often identified with the 'Central Dogma' proposed by Francis Crick, which in its original formulation states that

> once 'information' has passed into protein *it cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein.
>
> (Crick 1958, p.153; italics in original)

The Dogma, it can be argued, encourages the view that DNA is the dominant causal player in the life of the cell and the organism. The idea is that DNA segments called genes store information that flows outwards from the genome to direct and structure cellular events in order to realise certain phenotypic outcomes. A relatively straightforward correspondence between genes and phenotypic features or traits is assumed – or in other words there are genes 'for' traits. On this kind of basis some have attempted to place the molecular gene at the heart of a comprehensive biological worldview (Dawkins 1976).

In recent decades, however, genocentric perspectives on biological causation have increasingly been called into question (e.g. Strohman 1994; Moss 2003; Van Regenmortel 2004; Carrier and Finzer 2006; Morange 2006b). Quite what the concept of the gene equates to materially has come to look problematic (Stotz, Griffiths and Knight 2004), and it is increasingly clear that the relationship between individual genes (when we can decide what they are) and particular phenotypic traits is complex. Associated with these developments in biological understanding and consequential doubts about genetic determinism has been the rise of new perspectives on biological causation such as developmental systems theory (DST) (Oyama et al. 2001). DST's proponents see the genome as merely one causal agent amongst many: it is the collective interaction of multiple and diverse organismic and environmental structures, processes and agencies that generates biological phenomena. Therefore it is argued that accounts of biological explanation must be even-handed as regards the assignment of causal responsibility.

A strategy for combating the idea that causal primacy resides in the genome, and one that is often employed by those sympathetic to DST-style approaches to biological explanation, is to call into question the utility or even the sense of information talk in biology. It is argued that to the extent that information concepts can be clearly formulated they are as applicable to non-genetic entities and processes as to genetic ones. The project of eradicating information talk from biology is part of a broader approach that could be termed 'metaphysical abstinence'. I suggest that the price of such abstinence is, for many purposes, explanatory impotence. Biological explanations are typically freighted with examples of functional attributions that go beyond what is physically given, and (relatedly) frequently they involve perspectivism, or epistemic stance-taking on the basis of what is effective for knowledge-building within a particular kind of context. I explore the idea that when biologists use information talk in relation to the genome they are engaged in the attribution of function in an explanatory but metaphysically laden way. One of my conclusions is that information talk does not necessarily commit one to an unacceptably strong form of genetic determinism. I argue that it is possible to view genomic involvement in cellular processes as representing something akin to information processing, but to do so while emphasizing the causal parity of genomic and extra-genomic factors and the tightness of the coupling between them. I also outline another perspective on the informational genome, according to which it can be viewed as a device for storing – often only very approximately – details about an organism's phenotype. This is not what the genome is 'for' as such, but given certain patterns of material organization and change (within organisms and across lineages of organisms) the functional schema of genome-as-storage-

device is often capable of framing phenomena in molecular and cell biology in an explanatory way.

Functions in biology

I suggested in Chapter 1 that a significant difference between biological explanations and explanations in the physical sciences is that the former tend to make heavier use of functional terminology and concepts than the latter (Hempel 1965, p.297; Ruse 2000). Finding effective strategies for decomposing and localizing functions in complex systems such as the cell constitutes a major element of biological research, and frequently presents the researcher with major practical and conceptual challenges (Bechtel and Richardson 1993). Certainly it is not hard to find evidence of biologists' interest in attributing functions to biological structures and processes, sometimes on a somewhat metaphorical basis. A recent example is the description of a photosynthetic pigment–protein complex, the Fenna–Matthews–Olson (FMO) complex, as 'a rectifier for unidirectional energy flow from the peripheral light-harvesting antenna to the reaction center complex' (Ishizaki and Fleming 2009). A rectifier is a component commonly found in electrical power supplies that converts alternating current to direct current in virtue of the fact that the diodes from which it is made prevent current from flowing in one direction but permit it to flow in the other. Another example comes from a paper reporting the development of an accurate mathematical model of the aspartate-derived amino-acid pathway in plants (Curien et al. 2009). The authors highlight in the abstract of the paper a 'crucial result': 'the identification of allosteric interactions whose function is not to couple demand and supply but to maintain a high independence between fluxes in competing pathways'. In other words to identify the functions, qua causal roles, played by processes in their systemic context is frequently an important research objective.

Why should function talk be abundant in biology while in physics it is rare if not entirely absent? I suggested in Chapter 1 that the answer probably lies in the more abstracted character of the latter. I conjectured that in general the ontological elements (particles, laws, properties, and so on) discovered or postulated by physicists are not 'for' anything in particular, are functionless, because they are conceived as being fundamental and universal, and hence are viewed in a context-independent way. Biologists on the other hand are interested in a subset of the specific material configurations that do occur under terrestrial conditions, and when talking about such systems it often proves possible and

explanatorily useful to associate particular parts and processes with specific causal roles within a system, i.e. with functions in Wouters' function$_2$ sense (as discussed in Chapter 2). This style of usage is underwritten by implicit counterfactual possibilities regarding what would happen in the absence of the system component in question. So it is common when considering the cell to explicate its contents (structures and processes) in terms of the functional roles they perform. Sometimes this is done through metaphor, as illustrated by the example above from the recent literature. We speak of mitochondria as the 'powerhouses' of the cell, since they produce 'fuel' in the form of ATP molecules; the cytoskeletal proteins are said to act as a molecular scaffold; lysosomes perform recycling and waste disposal functions; DNA polymerase functions as a chromosome replicator; potential examples are legion. More controversially, it has for several decades been conventional to think of the nucleus as the control centre of the (eukaryotic) cell, and of the genome it contains as an information store that bears the instructions needed to make the cell – and indeed the host organism – operate correctly. What is the nature of the link between genome-related information talk and biological causation? Does the ascription of informational properties to the genome necessarily imply that it has causal priority over events and processes? These are the questions I shall address after first reviewing some of the relevant historical background.

The determining genome

Much contemporary molecular and cell biology accords to the genome very high status as a locus of causal power in the cell, and for the past half century this power has been understood to reside primarily in genes construed as discrete molecular entities. The concept of the gene was initially understood not materially, however, but more abstractly in terms of the transmission of traits from adult organisms to their offspring. Through the work of H.J. Muller and others in the early years of the twentieth century it gradually became clear that the inheritance of traits was somehow bound up with the chromosomes, but for some time it was unclear whether genes were to be identified materially with the protein or nucleic acid component of the chromosomes. (Avery published results in 1943 which indicated that it is the latter that has genetic significance, but even this did not completely overcome entrenched beliefs in favour of proteins (Morange 2000, chapter 3).) In 1953, famously, Watson and Crick worked out the structure of DNA, and this breakthrough gave rise to the research programme with which molecular biology would be pre-occupied in the following decade: the elucidation of the molecular basis of gene action.

It became clear that the distinctive causal properties of the genome are indeed associated largely with the sequence of the DNA base pairs, inasmuch as genetic differences correspond to differences of sequence.

Many aspects of molecular biology's early disciplinary character can be attributed to the influx of physicists to biology that took place after the Second World War (Keller 1990). Schrödinger's 1944 monograph 'What is Life?' was an important influence on some of those who would become the new discipline's leading figures (Watson 1966/2007, p.239), for in it he brought the theoretical approach of the physicist, and persuasive reasoning about quantum mechanics and thermodynamics, to bear on fundamental biological problems such as heredity and biological organization. Especially significant was his proposal that the chromosomes contained 'some kind of code-script':

> It is these chromosomes, or probably only an axial skeleton fibre of what we actually see under the microscope as the chromosome, that contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state. ... In calling the structure of the chromosome fibres a code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or a speckled hen, into a fly or a maize plant, a rhododendron, a beetle, a mouse or a woman. ... But the term code-script is, of course, too narrow. The chromosome structures are at the same time instrumental in bringing about the development they foreshadow. They are law-code and executive power – or, to use another simile [sic], they are architect's plan and builder's craft – in one.
>
> (Schrödinger 1944, p.22-23)

The 'axial skeleton fibre' would of course later turn out to be DNA (even if the structural organization of the chromosomes owes as much or more to histone proteins). It is hard to disagree with those to whom the idea that the genome 'contains ... the entire pattern of the individual's future development' appears excessively preformationist, or in other words appears to attribute to the genome the power to specify an organism's form (Moss 2003). (Although it might be argued that the term 'code-script' is somewhat ambiguous.) Another source of difficulty is the idea of the Laplacian 'all-penetrating mind' (APM) and how that is to be conceived. If such a mind were able to comprehend a snap-shot of all the particles in the universe, and were cognisant too of their properties and relevant facts about their instantaneous dynamics, then could it not track the future development of the universe, 'the individual's future development' included? Following that line of thought soon leads to difficulties concerning deterministic chaos and the corresponding necessity to store infinitely long representations of numerical quantities. Or, given that a particulate view of

the universe is physically unrealistic, we are confronted with the issues concerning indeterminism that are associated with understanding of the quantum mechanical phenomena that occur at what relative to us we regard as the microlevel. It must be assumed that the APM is introduced here then merely as a conveniently suggestive device for talking about the way in which a series of events unfolds. But even leaving quantum mechanical indeterminacy to one side the device as invoked is deeply problematic: we need to know much more about the amount of information regarding context the APM takes into account, and how it does so, as it attempts to determine from the structure of the code-script how the egg would develop.

Ambiguities notwithstanding, the tendency amongst philosophers of biology has been to see the quoted passage as representing a strong form of genetic determinism, and this interpretation has the virtue of being compatible with the strong impact the essay made on a number of future molecular biologists, amongst them Watson. For if genes were such powerful determinants of organismic development and function then they were of course a highly attractive object of study for ambitious proto-biologists. The conflation of developmental and mature-functional roles for the genome that Schrödinger envisaged can be related to a more subtle point about the involvement of gene action in cell processes. He saw causal direction as being relevant as much to mature function (which can be thought of as metabolism) as to development because the mature state is not a fixed structure – a configurationally invariant destination towards which dynamic developmental processes proceed – but rather must be thought of as a dynamic process in its own right. This is what Schrödinger makes clear elsewhere in the essay when he discusses the way in which the organism avoids 'the rapid decay into the inert state of "equilibrium"' (p.75) by feeding on 'negative entropy' through metabolism. This rather abstract conception of life in terms of the establishment and control of particular kinds of thermodynamic process by physical structures was no doubt of immense appeal to the early post-war molecular biologists whose backgrounds lay in the physical sciences. It reinforced the conviction amongst some of them that biology would be amenable to, and even demanded, a new more theoretical way of thinking about biological problems.

Schrödinger's idea of a hereditary code-script located within the chromosomes resonated strongly with ideas that developed in the 1940s and 1950s and which were associated with developments in computing that followed the ground-breaking pre-war and war work of such figures as Alan Turing and John von Neumann (Hodges 1983/1992; Kay 2000). New computational possibilities stimulated the development of new fields such as

cybernetics (Wiener 1948; Pask 1961) and information theory. In relation to the latter the exposition by Shannon and Weaver of the 'Mathematical Theory of Communication' (MTOC for short) was especially influential (Shannon and Weaver 1949). The concepts of cybernetics and information theory shaped research in a number of areas, but in biology they can be seen, in certain respects, to have pulled in opposite directions. MTOC described the properties of an abstract model communication system consisting essentially of a transmitter emitting a signal via a channel to a receiver. This abstract linear schema was simple enough to admit of formal mathematical treatment, allowing MTOC to talk in quantitative terms about the relationships between information, channel capacity, noise and so on. Cybernetics on the other hand dealt with the behavioural characteristics of non-linear systems featuring looping causal connections. Shannon declared that their intention was to develop a very general model:

> The word communication will be used here in a very broad sense to include all of the procedures by which one mind may affect another. This, of course, involves not only written and oral speech, but also music, the pictorial arts, the theatre, the ballet, and in fact all human behaviour. In some connections it may be desirable to use a still broader definition of communication, namely, one which would include the procedures by means of which one mechanism (say automatic equipment to track an airplane and to compute its probably future positions) affects another mechanism (say a guided missile chasing this airplane).
>
> (Shannon and Weaver 1949/1998, p.3)

But not long afterwards he says:

> The word information, in this theory, is used in a special sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning.
>
> (p.8)

The special sense referred to relates to the statistical properties of a signal in the context of a set of possible signals. Specifically, information is equated with a quantity referred to as entropy, by analogy with thermodynamics (Shannon and Weaver 1949/1998, p.12).

Other factors besides the prominence of cybernetics and MTOC are relevant to understanding the relationship between biology and information concepts. For example the development of X-ray crystallography, the principle technique of structural molecular biology, took place in tandem with (and depended on) that of computing, on account of its propensity to generate large quantities of numerical data and the requirement to process

that data mathematically (de Chadarevian 2002, chapter 4; see also Powell and Dupré 2009, p.58). The overall effect was that computational, informational and cryptological concepts and terminology were familiar to many scientists of the 1950s, molecular biologists included. It seemed possible to view DNA as the molecular realization of something akin to Schrödinger's code-script, and informational language provided molecular biologists like Francis Crick and Sydney Brenner with a way of distinguishing their outlook from that of biochemists. Brenner later noted how

> ... in the early days of molecular biology, it was an evangelical movement. Most people were against us. Most of the biochemists didn't understand the nature of the problems that we thought were interesting and important. They had a completely different set of attitudes. ... I can remember meetings at which it was impossible to get across to people the idea that the most important thing in protein synthesis was how the order of the amino acids got established. They said, 'That's not the important problem. The important problem is, where does the energy come from to join the amino acids?' Well, we have written, on many occasions, that the sequence is the important thing, and never mind the energy, it'll look after itself. And really, this is what this part of molecular biology brought. It said that the flow of information can be studied at the chemical level. I don't think biochemists actually understood the importance of information at that level. It wasn't information theory, it was the flow of messages, and we tried to seek for explanation in terms of the molecules.
>
> (Brenner, in Wolpert and Richards, 1988, pp.101-2)

It was the spirit of the new computing-related fields that informed the molecular biologists' use of informational terminology more than the original technical senses of the relevant concepts (Kay 1997). (As Brenner himself notes, the informational language of molecular biology 'wasn't information theory'.) Despite the ingenious schemes devised by figures such as Gamov, cryptological concepts proved to be of only limited relevance to genetic biology, although of course simple correspondence relations were found to underlie the encoding of the amino acid sequences of proteins by nucleic acid sequences (Kay 2000). With the elucidation of the basic mechanisms of gene expression – transcription of DNA into RNA and translation of RNA into proteins – Crick was able to encapsulate the informational perspective succinctly in terms of the Central Dogma and the Sequence Hypothesis (Crick 1958). It is of course the latter, discussed in Chapter 3, to which Brenner is referring in the above passage.

The Central Dogma and Sequence Hypothesis together constituted the theoretical foundation of molecular biology. The Central Dogma asserts that information flows from DNA to RNA to protein, and once there it cannot pass back again into the DNA. Exactly

what is meant by information in this context is not spelt out, however. The information of the Dogma is often conflated with causation or causal power, and indeed it is clear that the spirit if not the letter of the Dogma is that the important causal vectors governing cell processes run from DNA to the rest of the cell rather than in the reverse direction. (Arguably, the concept of a message is basically causal, and the term 'messenger RNA' trades on this.) Since its original formulation a number of discoveries have shown the Dogma to be not entirely exceptionless. Retroviruses possess the means to incorporate their RNA into DNA host genomes, and methylation of DNA base pairs can be seen as constituting an epigenetic inheritance mechanism, i.e. a mechanism capable of modulating genomic contents and stabilizing particular epigenetic cell configurations across generations. Nonetheless if the Central Dogma is interpreted narrowly in terms of the template-driven specification of protein structure from sequence then it clearly contains an important kernel of truth (Šustar 2007).

Ambiguities to do with the meaning of informational terminology, and the conflation of information with causation, have been important aspects of genetic determinism. By this last term I mean the thesis that there are things called genes and that these direct the development, behaviour and other phenotypic characteristics of an organism. Concomitant with this is the belief that for any particular trait there is likely to be a straightforward genetic explanation.[83] Preformationism, the doctrine that an organism's form is encoded in its genome, represents an extreme manifestation of genetic determinism. A point to note is that genetic determinism is inevitably in part about scale and the direction of causation, since the genome is localized within the cell to the nucleus (in eukaryotic cells) and in any case, to a first approximation, to the DNA. Causal influence, so the story goes, radiates outwards from genetic structures located on the chromosomes to the cell cytoplasm and (in metazoans) thence – for what amount to mereological reasons – to the macroscopic organism as it interacts with the world. This (to recapitulate some ideas discussed in Chapter 1) is reductionist in two senses: in terms of the dependence of the macro on the micro, and also as regards the emphasis the account places on the role of a single kind of factor (i.e. the genetic) in explaining the phenotypic characteristics of organisms. It is also worth noting that while normally defined in terms of the action of genes, the status and even existence of genes could be disputed while not denying genetic determinism, if it were re-defined in terms of the causal primacy of a cell's DNA

[83] Sometimes the traits in question are capacities of various kinds, e.g. for evolvability, or – as often posited by evolutionary psychologists – psychological dispositions to act in particular ways.

component. This point becomes potentially important as the concept of the gene becomes more complex and problematic (Griffiths and Neumann-Held 1999; Morange 2001).

Information talk

Because of its associations with genetic determinism and preformationism the 'information talk' introduced into biology by molecular biologists by way of the Central Dogma has become a particular focus for anti-reductionist criticism. The idea that DNA is informational remains pervasive, however, despite the qualifications to the Dogma that have been necessitated by findings in the very field it helped to define (Maynard Smith 2000; Nurse 2008). How can we make sense of the ubiquity of information talk in biology, and what is its explanatory import? Is it possible to maintain that genes and genomes play informational roles whilst acknowledging that many biological processes in which they participate, such as organismic development, depend on the interplay of a variety of causes, many of them non-genetic / non-genomic? The question probably appears to most biologists to be largely rhetorical, but nevertheless it seems important to be able to demonstrate the coherence of the position of one who wishes to answer it in the affirmative.

A route into debates about information talk in biology is provided by looking at what sometimes, in the guise of DST (developmental systems theory), is advanced as a contrasting perspective. DST has been described as an attempt to do biology without the dichotomies of nature or nurture, genes or environment, biology or culture, and as not so much a theory as 'a general theoretical perspective on development, heredity and evolution, a framework both for conducting scientific research and for understanding the broader significance of research findings' (Oyama, Griffiths and Gray 2001, pp.1-2). A core principle of DST is that causes should be seen within their systemic contexts. Causal responsibility for biological phenomena is typically seen as being distributed across multiple entities and processes, and rarely (DST proponents tend to argue) can causal primacy be attached to a single class of entity or process. Griffiths and Gray (2005) note how 'DST stresses the delicate dependence ('contingency') of development on a rich matrix of factors 'outside' the genome' (p.419). There are hints here of a rather holistic perspective on biological causation (Godfrey-Smith 2001, p.289). The status of information talk in biology plays an important part in DST-related debates, since the idea of the informational gene, as well as being associated with preformationism (Moss 2003), is regarded by some as

representing the privileging of genetic causes in biological processes such as development. The editors of the canonical articulation of DST perspectives state that

> We believe that the heuristic value of the idea of developmental information in certain contexts is more than outweighed by its misleading connotations. Locating information in a single type of developmental resource obscures the context-dependency of causation by localizing control.
>
> (Oyama, Griffiths and Gray 2001, p.5)

In his (2001) Griffiths develops the attack on information talk in order further to undermine the idea that genetic causation should be privileged over the other kinds of causal factors that impact on biological processes such as development. This is an exemplification of what he terms the 'parity thesis', which is a general thesis about biological causation to the effect that genetic and non-genetic factors play equally important parts in developmental processes (Griffiths and Knight 1998). It does not, its advocates argue, 'imply that there is no difference between the particulars of the causal roles of genes and factors such as endosymbionts or imprinting events. It does assert that such differences do not justify building theories of development and evolution around a distinction between what genes do and what every other causal factor does' (Oyama et al. 2001, p.3). The purpose of the thesis is 'to prevent these empirical differences [e.g. between nucleic acids and natural languages] turning into a kind of scientific metaphysics, as happens when genes are identified with information (or even "form") and everything else in development as mere matter. This distracts attention from the many ways in which non-genetic resources sometimes play biological roles more usually associated with genes. It also leads to the inappropriate lumping together of very different non-genetic resources (the "environment")' (Griffiths 2001, p.406).

I shall argue that these concerns about information are misplaced, and that the 'particulars of the causal role of genes' are in fact capable of supporting informational interpretations that are neutral on the issue of causal priority. If I succeed then the onus is on Griffiths to explain why information talk is so ubiquitous and why it is associated in particular with genetic and genomic processes. I begin by outlining, and then questioning, Griffiths' position on information, before outlining an alternative position that accounts for the appeal of information talk.

Griffiths begins his argument with the claim that '[g]enetic causation is interpreted deterministically because genes are thought to be a special kind of cause. Genes are

instructions – they provide information – whilst other causal factors are merely material' (2001, p.395). Thinking in this way is mistaken, however, he argues. Contra Maynard Smith (2000), information talk in biology has a substantive basis, Griffiths contends, only when it adverts to the idea that there is a genetic code 'by which the sequence of DNA bases in the coding regions of a gene corresponds to the sequence of amino acids in the primary structure of one or more proteins'. The rest of information talk in biology

> is on a par with the claim that the planets compute their orbits around the sun or that the economy computes an efficient distribution of goods and resources. It is a way to talk about correlation that, in some cases, allows a useful application of the mathematical theory of communication and in others plays no theoretical role but merely reflects the current cultural prominence of information technology.
>
> (Griffiths 2001, p.395)

Thus important aspects of Griffiths' position are the strong distinction he draws between metaphor and theory, and the disparaging view he holds of the former. If genes and genomes are informational in only a metaphorical sense then they are not *really* informational, it is implied.[84] Yet this kind of claim is at odds with the intuition one might have that a metaphor, or indeed any sort of analogy (I take it that it makes sense to think of metaphors as especially abstract analogies or similarity-stressing comparisons), will have explanatory value just in case it establishes a correlation of some kind between features of the object to which it is applied and features of the object that is the metaphor's basis. If what we are interested in is scientific explanation then analogical relationships may be rather important, and it is not obvious a priori that our explanatory practices do not depend heavily on the possibility of pointing out such relationships. Merely to reflect 'the current cultural prominence of information technology' may be of considerable epistemic value. A second but related worry is that the distinction between metaphor and theory is in any case overdone. Theories are idealizations, and so the resemblance between a particular empirical phenomenon and the theory that is regarded as its best scientific characterization may be somewhat tenuous.

If biological information talk is to be given a *theoretical* footing then the choice, Griffiths argues, is between two major alternative conceptions of information: causal interpretations (with MTOC usually being regarded as the prime exemplar) and semantic (or intentional) construals. Concerning the latter, there is widespread agreement that the only viable options are so-called teleosemantic accounts, in which biological meaning is

[84] Levy (2007) argues that information can be understood in biological contexts as an explanatory metaphor.

cashed out in terms of evolutionarily selected function (Millikan 1984). A signal S carries a certain meaning M for an S-detecting system, for example, if it is in virtue of S having been interpreted to mean M that the S-detecting system has survived to the present day. (Systems that treat S as meaning something other than M are winnowed out by natural selection. If S indicates 'big and stripy with pointed teeth' and this is taken – or not – to mean M 'a tiger that might eat me' then it is easy to see in principle how this might work.) As this sort of example hints at, information in this teleosemantic sense can be associated with many things besides genes and genomes. But I am in any case sceptical that when biologists talk about genetic information they are appealing to a sense that is freighted with semantic burdens of an intentional kind. Rather, such meaning as might be implied I take to relate to causality rather than intentionality. A number of significant pitfalls attend the explication of this causal sense of aboutness, however.

One rather simplistic proposal for what it is to say that a gene G is 'about' (or 'means') a particular phenotypic feature P is that statistically, in a particular cellular or organismic context[85],

> (A) the occurrence of G correlates with the occurrence of P and the non-occurrence of G correlates with the non-occurrence of P (i.e. [if G then P] & [if ~G then ~P]),

or

> (B) the occurrence of G correlates with the non-occurrence of P and the non-occurrence of G correlates with the occurrence of P (i.e. [if G then ~P] & [if ~G then P]).

The idea here is that for any given G and P, only (A) or (B) can hold if we are to say that G and P are meaningfully related. The need to cater for both patterns of correlation, direct and inverse, arises from the different molecular causal possibilities that can potentially arise. Some phenotypic traits depend on the presence of particular molecule-dependent functions, and these functions will be disrupted if the relevant kinds of molecules are absent or are dysfunctionally abnormal. An example would be the requirement for a particular kind of enzyme to be present to catalyse a specific reaction. This would conform

[85] One could add environmental context here, but the environment generally affects gene action through its effects at the molecular and cellular levels so it is redundant.

to (A) above. Other traits may require the *absence* of certain kinds of molecule. The presence of a gene coding for an inhibitor of the enzyme just mentioned, for example, would abolish the P in question (as per (B) above). (There is a parallel here with the negative account of emergence given in the previous chapter, in that sometimes interest centres on phenomena that arise as a result of what is missing rather than what is present – the relevant causal schemas in the case of emergence, an absence of the relevant molecules in the case of some phenotypic outcomes.)

Complicating this simple picture are possibilities for multiple realizability (Fodor 1974): different genes can give rise to the same phenotypic outcome, and similar outcomes can arise in very different ways. If a certain cellular function depends on the formation of a complex of six proteins, say, then the absence or abnormality of any of the six might abolish the function associated with the complex. On the other hand several different variants of one of the proteins may exist, each encoded by a different gene and each capable of forming a functional complex with the other five proteins.

However, even this more qualified picture cannot be the complete story concerning a causal sense in which a gene could be said to be 'about' a phenotypic feature, or even the major part of it, since it is agnostic about the direction of causality between G and P (i.e. P because G, but also – mistakenly on a causal basis, even in the absence of possibilities for multiple realizability – G because P). The account must be supplemented by a belief component, to the effect that such correlations as may be apparent are taken to be underpinned by a molecular narrative spanning G and P. This asserts that in principle there is a causal-mechanical account to be given of how, in a particular set of circumstances, G gives rise to P by participating in processes which potentially involve numerous causal factors besides G. But note that it may be untrue to say that P could not occur in the absence of G (again, even in the absence of possibilities for multiple realizability of the kind described above), for the cell might sense the effects of the lack of G and compensate by expressing another gene such that P does in fact result. This conclusion is at odds with (A) above.

Moss (2003, 2006) explicates these complications in terms of his distinction between two senses of gene, one pertaining to phenotypic predictability (Gene-P) and the other to developmental processes (Gene-D). These distinct concepts can be related to the inverse and direct relations between DNA segments and phenotypic features discussed above: 'Gene-P phenomena are based on the absence of an otherwise normal protein or

other resource and what bodies will predictably do, for better or for worse, in the absence of that normal resource' (2006, p.529). It is the Gene-P sense that is associated with scenarios of type (B) above. A Gene-D ('the sense of a gene when it is defined by a nucleic acid sequence that provides the template resource (or information) for some set of potential downstream polypeptide and/or RNA products') on the other hand is indeterminate with respect to phenotypic outcome – 'just because between variable splicing, co- and post-translational modification, targeting, and many other contextual factors, the same Gene-D could be a contributing factor to entirely different, even antithetical, phenotypic outcomes' (2006, p.529). Scenarios of type (A) above involve the Gene-D sense.[86]

The upshot of all this is that although sometimes P does allow us to infer G, and G may correlate with P, G/P relations can be more obscure than this. Whether or not empirically G and P do show up as being correlated may be highly context-dependent, and associations between G and P may not be readily apparent (Rosenberg 1985). Sometimes it is only when large-scale studies are conducted, and the data subjected to sophisticated statistical analysis, that the relevant correlations are detected (see e.g., in relation to Alzheimer's disease, Harold et al. 2009 and Lambert et al. 2009).

Notwithstanding the complexity of genotype—phenotype relations, the possibilities these reflections indicate of relating the semantic to the causal have the effect of undercutting Griffiths' strong dichotomy between two basic types of information account. ('If there is a relationship between intentional information and causal information it is a complex and distant one' (2001, p.397).) However, it should be recalled that the basis for Griffiths' antipathy towards the possibility of construing genes as informational in a way that is inapplicable to non-genetic factors is that it is believed to lend support to the privileging of genetic causation. This position makes sense if information is interpreted in semantic terms, for it is semantic construals that are most vulnerable to charges of varying degrees of context-disregarding genetic determinism right up to full-blown preformationism. It is this that bothers Griffiths:

[86]A Gene-D is not necessarily indeterminate with respect to phenotypic outcome, however, despite what Moss asserts – just as it is not necessarily the case that 'any gene that is a "gene for" ... would count as a Gene-P', i.e. would involve the absence of a protein (Moss 2006, p.529). In other words a DNA sequence may be transcribed and translated to create a protein that instantiates a positive capacity, e.g. to metabolize compound X, and it makes sense to say that the gene is the gene for that capacity with which it positively correlates. Indeed Moss' general claims about the indeterminacy of genotype—phenotype relations that he attempts to capture through the concepts of Gene-P and Gene-D must be seen as contingent and subordinate to empirical findings, rather than as being necessary in any strong sense.

> The intuitive notion of information is a semantic notion, carrying the implication that genes, unlike other causal factors, are about, or directed at, the outcomes they help to produce. Little wonder, then, that the gene-trait relationship seems intuitively more context-independent than the relationship between traits and other causes.
>
> (2001, p.396)

Seen non-semantically, on the other hand, it is not clear why information cannot just be viewed as pertaining to a variety of causation which, whilst it might make genetic action distinctive in some respect or other, is not obviously inimical to the interests of the anti-reductionist. Turning then to causal construals of information, MTOC is standardly appealed to as the best candidate for a full-blown theory of information, and this is a line from which Griffiths appears not to depart. We might thus expect MTOC to contrast sharply and obviously with mere metaphors, and yet the theory articulates in simple formal terms some highly abstract ideas. In fact it is an account so simplified and so abstract that there is a good case for saying that it can only ever bear an analogical relation to its exemplifications in actual phenomena. Moreover there is (as I shall discuss shortly) a sense, concerning entropy and order, in which MTOC gets things exactly the wrong way around when applied to genetic information. However, while these kinds of detail are important for an appreciation of the place of information concepts in biology they are to some extent incidental to understanding Griffiths' position, the central feature of which is that whichever sense of information is chosen, genetic factors have the same informational status as a variety of non-genetic causal factors involved in development. The implication is that information talk is otiose and therefore dispensable. The mystery then, however, is why information talk is so widespread, yet is associated mostly with the genome, when currently elaborated information accounts are either inapplicable to biological phenomena or are applicable beyond the genome. I conjecture that this is a mystery that can be solved only by reconceptualizing information.

My aim then in the remainder of the chapter is to identify and articulate a view of information talk that makes sense of its ubiquity in relation to genomic and genetic processes whilst recognising the causal parity of those and other processes and events. In seeking to do this I am taking seriously the possibility that information talk performs a useful and substantive explanatory function in molecular and cell biology, and seeing such putative utility as the factor that accounts for its persistence. The fact that in order to make themselves understood to diverse audiences it is apparently unnecessary for users of information talk to spell out exactly what they mean by information in the context of

genetic processes implies the existence of tacit inter-subjective agreement about the concept's projectibility. If this common understanding is more widespread or entrenched than that of genetic concepts and mechanisms then information is potentially a valuable basis for explaining those concepts and mechanisms. The approach I shall take is to treat the property of being informational as a kind of functional attribution, much like many others that biologists routinely make (invoking such notions as control, signalling, sensing and so on). This promises to shift the emphasis, as with emergence in the previous chapter, from ontology to epistemology. Whether some new conception of information entails the idea that genes have causal priority will depend on its details, but prima facie, given what I have already said about cellular complexity (Chapters 4 and 5) and downward causation (Chapter 6), it seems unlikely that such an entailment will be found.

Informational structures and informational roles

To get started it is helpful to note how difficult it is, on reflection, to identify some object or class of objects that could be said to be informational in some 'true' or definitive sense. There is in fact, I suggest, no list of criteria which when fulfilled make something definitively the bearer of information, and there is no class of paradigmatically informational objects in relation to which other objects stand in some merely metaphorical relation as regards the bearing of information. We can probably agree, however, that music CDs contain information, as do novels, cookbooks, train timetables, radio broadcasts, and much else besides. We might say that clocks impart information, but do not contain it. The state of the clock watcher's mind seems as important here as that of the information-imparting artefact. If a clock indicates a time that is in rough agreement with our expectations then we presume that the time indicated by the clock is approximately the right time, but if there is a large discrepancy then we tend to infer that the clock is wrong or (depending on what grounds we have for believing the clock to be indicating the right time) we introspect in order to figure out why we were so mistaken about the time. We might think of a computer program as constituting a kind of information capable of directing the operations of mechanisms that act on another kind of information we refer to as data, and we might also note that the program can be regarded as data in the context of the operations specified by the computer's operating system. This suggests a relational aspect to informationhood.

Thinking about the processes in which the information objects we commonly recognise as such participate leads to the thought that they usually involve the interaction of an interpretative or reading mechanism with the information object. As a result of the interaction the state of the interpretative or reading mechanism, or some other system to which it is coupled, changes in ways that are keyed to the nature of what is read. The state of what is read, on the other hand, is generally unchanged by the reading process. These kinds of idea in turn lead to the thought that associated with – and perhaps partially constitutive of – the concept of information are a variety of informational roles, including (besides reading / interpreting) production, storage, transmission, and transformation / translation. My task now is to investigate whether it makes sense to see the genome as a structure that fulfils informational roles such as storage and transmission.

If it were to prove possible to attribute informational roles to the genome then this would distinguish it in functional terms from cell components that play structural (i.e. force-resisting) or catalytic roles, say, or those that act as containers, or sources of energy. My immediate working hypothesis is that the genome represents a storage device that is in some ways analogous to magnetic tape or some form of computer memory. This interpretation of the genome's function is, some biologists may think, a rather obvious one. But it is philosophically interesting to see how it might work and how it squares with the views of those like Griffiths for whom information talk is, if not anathema, to be regarded with deep misgivings. I approach the concept of genome as storage device from several directions. The first approach is to focus on the nature of the DNA molecule, and involves recognising a point that has been made by a number of molecular biologists at different times concerning its structure. The view to which this gives rise connects with another informational perspective, according to which the genome participates in processes that stand comparison with computer algorithms. A rather different approach is to think about the patterns of material stability and change with which genomes are involved across organismic lineages. These different approaches yield subtly different, but related, pictures of the genome as a storage device, and separately as well as jointly these are compatible with, and go a long way towards making sense of, information talk in biology. In addition, and significantly, I believe that they should not arouse the antipathy of anti-reductionists.

To consider first the structure of the DNA molecule, Watson and Crick wrote:

> The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely

> that the precise sequence of the bases is the code which carries the genetical information.
>
> (Watson and Crick 1953, p.965)

Wilkins, reflecting later on the impact of Watson and Crick's structure, noted that

> ... the really impressive feature of the structure was the extraordinary way in which the two kinds of base pairs had exactly the same overall dimensions and shape. ... I was rather stunned by it all – the exactness and the replication idea, and the resolving of the paradox that DNA was so regular (even crystalline) and yet contained the complex and irregular genetic message in the sequence of base pairs.
>
> (Wilkins 2003, p.212)

The eminent crystallographer David Blow, discussing the contrasting structures and roles of DNA and proteins, said in an early overview of the then new field of molecular biology:

> There are two main types of polymer to be considered. One type is used for the storage and transfer of information, and for this purpose a polymer is required whose properties are virtually independent of the information written on it, just as one page of a book is very much like another. Enough effect is required to enable the information to be 'read', and no more. ... A quite different type of polymer is needed for the fabric and working material of the cell. Here we require the largest possible diversity of properties consistent with the continuity of a single type of chain.
>
> (Blow 1962, p.179)

Common to all three passages is a sense of the significance of the fact that DNA has an overall structure that is substantially independent of base sequence. This is a highly unusual property for a hetero-polymer (one made up of different monomers) to have. In general different monomers have different sizes and chemical properties, and these impose different steric and other constraints on the polymer chain, such that distinct monomer sequences are associated with different polymer conformations. Indeed, proteins exemplify this to a high degree: it is the idea that the amino acid sequence essentially determines a protein's conformation that the Sequence Hypothesis enshrines.

The linear arrangement of an arbitrary number of symbolic tokens selected from a limited number of types (nucleotide bases A, C, T, G) that is a constitutive feature of DNA is strongly analogous – in non-semantic respects – to the way in which a text is built up by selecting letters from an alphabet. The co-linear relationship between 'one-dimensional' nucleic acid sequences and the polypeptide sequences they encode has been described in terms of 'template correlative determination', where that phrase relates to a concept

explicated in terms of the different causal characteristics of nucleic acids and proteins (Šustar 2007). Protein folding processes mean that one-dimensional nucleic acids are capable, in the context of the machinery of gene expression, of giving rise to three-dimensional protein molecules that are structurally and hence functionally uniform (within a given context) on a type-specific basis. If it becomes possible, through use of (say) the computational techniques discussed in Chapter 3, routinely to predict the structure of a protein from its amino acid sequence then in theory it also becomes possible to trace determinate connections running from DNA base sequence to protein structure. In practice the production of many proteins involves RNA processing steps that complicate the picture somewhat, but perhaps not insuperably (see David and Manley 2008).

These distinctive structural properties make for strong analogies between DNA and a variety of data storage media. Indeed, novel uses for DNA have been proposed that capitalize on its data storage capability.[87] The regular bulk structure of a substrate supporting arbitrary sequences of readably different monomers enables generic 'read', 'write' and 'error-correction' mechanisms to act on DNA, rather in the way that a magnetic tape can be scanned by a tape-head because of the uniformity of the tape's form and the independence of that form from the specifics of the recorded content. The image of a DNA molecule being 'read' by other molecular devices, such as RNA polymerase, is a diagrammatic commonplace of textbooks on molecular and cell biology. Such images are presumably explanatory, but how do they work? This is a matter for psychological conjecture, but I suspect that one of the things they do – provided that we understand how to interpret the diagrams in question – is help us build cognitive models that provide a basis for imagining particular phenomena. A diagram showing DNA being translated may contain arrows which show the direction of travel of the polymerase molecule, and their purpose is to indicate how to turn the static image into an internal animation. The arrow says: imagine this feature in the diagram (i.e. the polymerase molecule) travelling along this feature (the DNA molecule). The understanding that the diagram confers is thus connected with the capacity to simulate and imagine a pattern of events involving a variety of molecular structures. Being able to imagine a pattern of events in this way is to have a form of causal knowledge.

On the basis of these reflections we might say that DNA is informational in this respect: it is capable of supporting the localized storage of tokenized encodings of entities with causal properties that are specific given particular environmental conditions. This

[87] http://www.nytimes.com/2007/06/26/science/26DNA.html (last accessed 22 September 2009).

sense clearly does not imply any kind of preformationism, but it is informational to the extent that there are structural and functional similarities with a range of objects we unproblematically regard as non-semantically informational. If I am right that the structural properties just discussed do underpin a tendency to ascribe informational properties to DNA then perhaps it is also a tendency that is encouraged by other factors. Among them is the apparently critical importance of sequence. A variety of mechanisms serve to ensure that sequences are copied accurately during DNA replication, errors are corrected, breaks and mismatches in DNA strands are repaired, and sequence variation is introduced at controlled levels during particular key genomic processes (see e.g. Lodish et al. 1999, pp.472-492). These processes include the recombination events that occur during meiosis (involved in gamete production) and following the fertilization of an egg by a sperm. It is hard not to conclude from the existence of these diverse and tightly regulated sequence-focused processes and properties that sequence has profound biological significance.

Not only is the genome stable, on account of the native structural stability of DNA molecules in tandem, in cellular contexts, with the active maintenance of sequence and structure just described, but it is highly ordered. By this I mean that while many base sequences of a given length are possible in theory[88] (and all sequence possibilities are, apparently, physically compatible with DNA's double helical structure), very few are maintained and perpetuated in (are causally compatible with) a particular cellular or organismic configuration. This points to a potential confusion concerning MTOC that I have already alluded to. The Shannon/Weaver construal equates information with entropy, such that a high degree of information is formally equivalent to a *high* degree of entropy, whereas an important sense in which DNA can be considered informational relates not to a high degree of disorder, but of order in a sense closer to that invoked by Schrödinger.[89] This is order qua improbability, i.e. *low* entropy, and it is largely a structural matter to do with the factors just discussed: the surprising independence of the bulk form of DNA from base sequence, the stability of the molecule, and the way in which sequence is maintained and transmitted by specialized mechanisms.

[88] For a DNA molecule of length *N* base pairs the number of possible sequences is 4^N. (So 16 sequences of two base pairs are possible.)

[89] Stonier (1990) discusses the relationship between information and entropy in MTOC and in other accounts in more detail. He argues that Shannon's theory, taken to its logical conclusion, 'would mean that pure noise, which contains the greatest amount of entropy, would contain the greatest amount of information' (p.56). Perhaps to resolve the issue requires distinguishing in some explicit way between on the one hand the sense of entropy involved in the spatial constraint of nucleotides as they exist in a DNA molecule as against their free state and on the other issues to do with the arbitrariness of base sequences and the fact that any given sequence of a certain length is a member of a much larger set of sequences of that length. It is on the latter issue, to do with probability of selection from a set, that the Shannon/Weaver account focuses.

At the same time the genome appears to have few other functions besides storing a particular sequence of base pairs: it is not itself the catalyst of a particular class of reactions, it does not itself pump protons, it does not itself repair cell walls, and so on. This combination of, positively, the genome's apparent indispensability related to sequence storage, and, negatively, a lack of other obvious functions, make the genome what could be termed an attractor for functional attributions. It seems to be doing something of functional significance, i.e. playing some sort of causal role, and epistemically we feel the need to identify that role and find ways to speak about it. The ideas of information storage and transmission I take to be attempts to capture that role.[90]

I suggest – and perhaps it is because informational habits of thought are so ingrained that this sounds somewhat banal – that DNA can be regarded as a storage structure in the proximate and structural sense just discussed, and that what are stored are representations of causal entities. How does this set of ideas relate to DST? One of the dangers as I see it of banishing information talk as advocated by many of DST's proponents is that we loose access to a valuable explanatory resource: the idea of causal potential. Current philosophical accounts of causation are incapable of dealing with the storage of specific causal possibilities. If we want, for well-motivated explanatory purposes, to refer to potential causation and to related notions of causal memory and buffering then we need a different or supplementary – and more metaphysically laden – concept. This is part of what the concept of information gives us. When we see or conceive of a DNA molecule in a cell, with that molecule bearing particular promoter and other sequences – and when in addition we see (or feel warranted in assuming) the presence of specific molecules associated with gene expression – then it is surely inevitable that we recognise the latent causal possibilities associated with the DNA. Granted, those causal possibilities are stored within the total system of DNA plus expression apparatus, all within a cellular context, but the way in which DNA participates in the processes that realise those possibilities makes it natural, I contend, to talk in terms of information storage – and to identify that function with the DNA rather than with the other cellular apparatus.

Some cautionary remarks are in order. One is that the causal properties of the proteins that issue from the genome in response to the dialogue between it and the rest of the cell are contextually contingent. When we think we understand a context well we regard the causal properties of entities within it as determinate because we believe we can accurately imagine how they will behave in that context – although empirical testing has the

[90] A transmission sense of information has recently been explicated by Bergstrom and Rosvall (2009).

potential to disconfirm the assessments we make of our understanding. Also, the structural—functional analogy between DNA and other informational objects is part of the basis on which we attribute informational roles to DNA, but it holds in the main only schematically and proximately to the structure of DNA in relation to specific operations in which it participates such as transcription and error correction. When the larger system is considered we see significant differences with respect to the properties and behaviour of other information objects as well as similarities, and in the next section I describe some of the relevant systemic issues. The case of DNA and information storage shows, however, how promiscuous our function-attributing tendencies can be. Even a rather schematic structural—functional analogy can serve as basis for an attribution that has explanatory utility, especially when supported by additional factors (which in the present case include the apparent biological significance of sequence).

Networks

So far I have emphasized simply the fact that a DNA molecule stores a particular base sequence, and that in a cellular context this sequence is actively maintained and modified by a variety of mechanisms and processes. I have discussed one of the major ends to which this sequence is put: the storage of what I referred to as tokenized encodings of entities with specific, albeit context-dependent, causal properties – i.e. protein molecules. Now I want briefly to consider another way in which DNA sequences participate causally in cell processes. It involves thinking not just about the reading or maintenance of particular portions of sequence in isolation but rather about how such sequence-focused operations can form elements within larger processes. These processes involve what are commonly referred to as gene expression networks, which can be succinctly defined as 'co-expressed functional groups of genes' (Lelandais et al. 2006). Particular cellular phenotypes are associated with specific molecular populations and the particular cell structures and behaviours to which these give rise, and these in turn are associated with particular patterns of gene expression. Comprehending these patterns involves appreciating the roles of both genomic and non-genomic factors and the mutuality of their interactions. The expression of certain DNA segments produces proteins with specific DNA binding properties, i.e. specificity for certain DNA sequences. These so-called transcription factors, which effectively represent a modular mechanism for altering the specificity of generic DNA transcription devices, result in the expression of particular sets of genes. This yields proteins having the particular properties that collectively, and in conjunction with other cellular structures, establish or maintain particular forms of cellular life.

Regulation of gene expression, for example through feedback mechanisms that depend on the ability to sense the levels of particular expressed proteins or the levels of particular metabolites, couples genomic activity to cellular processes. The overall picture that results is one in which the genome and the rest of the cell form an integrated causal complex, with the genome storing a set of encoded representations of structures that includes a subset that acts back on the genome to determine which of the total set are physically instantiated as a function of current cellular activity. Events within the cell that arise as a result of the molecular activities of its contents, and the effects that sensed external events have on those activities, are capable of modulating the action of the molecules that act on the genome to regulate gene expression. Hence cellular life as it is shaped by specific environments might be said to determine its own future by selecting the operative patterns of gene expression. This, clearly, is not a picture in which genes and the genome could be said to have causal priority over non-genomic structures and processes. But is it one in which the genome could be said to be playing an informational role? Yes, I contend, in the storage sense I have discussed in terms of the structural and entropic factors that underwrite its attribution. However, thinking about the regulation of gene expression and gene expression networks reminds us that the genome does more than simply store the encoded representations of protein molecules. Its regulatory sequences (i.e. sequences that are the targets of regulatory molecules such as promoters and repressors of gene expression) serve to bind it tightly to cellular processes in a highly context-dependent way. Sometimes it is suggested that the resulting process architecture makes the cell akin to a computer program, for example because particular functions are implemented, through the expression or repression of particular genes, when specific conditions hold. The parallel here is with the role of conditional statements in software, which have the form *if [conditions] then do [actions]*. If a particular transcription factor, a promoter say, is activated when a particular cellular condition is sensed (for example when the cell cycle reaches a certain stage), then the genes regulated by that factor will be expressed and a particular molecular function, qua role-fulfilling activity, will be instantiated.

To push the analogy further, when a computer program runs the values of its variables are updated in accordance with the scheme of logical operations, or algorithm, its statements define, and at any given moment the state of any particular 'run' of the program is reflected in a particular set of variable values. A cell state is defined in terms of the distribution of a population of molecules representing a snapshot of a similarly dynamic process, which can be seen as having a logical architecture defined in part by the placement

of regulatory sequences in relation to coding sequences. Thus it seems possible to draw a parallel between the state of a cell and the state of a program run, and to see both cells and programs as reflecting the workings of a logical mechanism qua a system of entailments (in both cases reflecting the occurrence of sequences of symbols in a particular kind of context). But there are striking differences too. For example, the hardware/software distinction that is readily made where man-made computers are concerned appears to have little place in the cellular case. The algorithmic logic of the cell, to the extent that such language is defensible, is distributed across the 'hardware' of both genomic and extra-genomic cell structures. Moreover, the 'logic' of cellular events and processes is not determined solely by issues of sequence occurrence and placement: it is also dependent in part on spatiotemporal aspects of genomic and non-genomic processes (e.g. molecular diffusion rates), which have no obvious software equivalent.

The algorithmic analogy points to an additional sense in which the genome and genetic processes can be thought of as broadly informational, building on the structural sense outlined earlier which relates to proximate DNA-reading and sequence-correcting mechanisms. Those mechanisms suggest an analogy with other information storage structures such as CDs and magnetic tapes, but when viewed systemically the context-dependence of the causal capacities of the entities DNA encodes becomes clear. The overall effects those entities have depend as much on the composition and character of the extra-genomic cell as they do on their own individual properties, and the timing and order of transcription events matters. The position of a coding sequence within the genome, not just linearly in sequence terms but also in terms of three-dimensional genome structure, may be an important factor influencing its transcription. (Whether a transcription complex is bound at one sequence may for steric reasons influence whether transcription is possible at a spatially adjacent – though perhaps sequence-distant – site, for example.)

These reflections on the systemic role of the genome and genetic processes show how DNA is informational in a robust structural—functional sense that depends on schematic overlap with other informational objects involved in data storage, and which is reinforced by entropic and metaphysical factors. Thinking about the participation of DNA sequences in the larger context of cellular processes shows, however, that genome and cell are tightly coupled and that causal priority attaches to neither. Equally, a computer's hard disk drive could not be said to be causally prior to the rest of the hardware in the context of the running of a software application. Rather it is a place that may be used to store

specific sequences of byte values that play particular functional roles in the context of the dynamic process economy of a particular application.

Genomes across generations

The construal of DNA as informational on the several bases just outlined is, I suggest, compelling in its own right, but it gains additional force from its compatibility with another idea. This is the second major sense in which the genome can be seen as a repository of information. Appreciating this sense depends on shifting one's view from the proximate causal processes that take place within an individual cell to processes spanning lineages of organisms. Focusing on the replication and transmission of genomes across generations of organisms reinforces the sense engendered by thinking about how genomic processes couple with extra-genomic ones via regulatory networks that the genome has a causally ambivalent character. Sometimes – within the context of an individual organism, when we focus narrowly on the expression of an individual gene – it appears to be situated at the start of a chain of events (and thus looks like the locus of significant causal responsibility for subsequent events in the cell as implied by the Central Dogma). However, at other times – over evolutionary timescales and in the context of lineage descent – it looks as much like the passive outcome or instrument of processes as it does their active driver. The claim I now want to defend is that these issues are bound up with the idea that genomes represent, in various ways, encoded partial descriptions of their organismic hosts.

To appreciate this it is useful to imagine several alternative replication scenarios. The first case to consider is that in which an organism gives rise to offspring that are exact copies of itself. All individuals in the lineage have the same phenotype P. Whether a member M of such a lineage survives beyond a certain generation N depends on the fitness of P relative to M's environment at generation N. The species will survive only so long as there are individuals living in environments to which P is well suited. Evolutionary adaptation is clearly not possible and environmental change continually threatens to extinguish the species altogether. Suppose that P is entirely determined by a genome G, and that P is invariant from parent to offspring on account of the perfect copying and transmission of G. In these circumstances I contend that it makes sense to think of G as a description of P.

The second scenario to consider is one in which an organism gives rise to offspring having phenotypes that vary from its own in seemingly random and extreme ways, irrespective of the genome they inherit. Let us suppose that as before the genome is copied and transmitted with complete fidelity, only now assume that the organism's phenotype is causally independent of genotype. (Perhaps it is keyed, exclusively and in a fine-grained way, to the details of its developmental environment.) Now environmental change appears to pose less of a threat to the lineage, inasmuch as there is always a chance of variation yielding a context-fit phenotype. However, a fit phenotype is no more heritable in this case than an unfit one, and a priori it seems unlikely that the generation of random, extreme, but non-heritable variations can be beneficial for lineage survival. (The effects of rare but highly beneficial variations are likely to be undone quickly. The relative dynamics and qualitative character of phenotypic variation and environmental change are presumably key to determining the chances of long term lineage survival.) The genome in this case, assuming that it is invariant from generation to generation, serves as a useful marker or label for the lineage – being perhaps the only common feature by which to identify organisms as lineage members. But it is not in any richer sense a description of members of the lineage. If, however, the genome were to vary slightly from parent to offspring then it could still function as a lineage label, but it would become even more useful since it would now potentially provide a guide to an individual's path of descent. This is the sense in which the genome represents an information store to which Zuckerkandl and Pauling (1965) drew attention (and see also Sommer 2008).

The third case is, I take it, what we in fact find in nature: organisms bearing genomes that at least in part determine phenotype and that vary slightly from parents to offspring. Now evolutionary adaptation is possible, because there is a mechanism for generating variation that stands a good chance of persisting in a population for multiple generations to provide the raw material on which selection can act, generating increasingly well-adapted organisms against a backdrop of modest environmental change. As in the previous case the genome can potentially provide evidence regarding the historical path of descent of an individual organism. In addition, however, the causal coupling of genotype and phenotype – which is a major factor in making phenotypic variation heritable – means that there is an additional sense in which we can think of the genome as an information store.

As well as revealing the historical sense in which genomes can be informational – i.e. as a record of the path of descent – thinking about generational transitions as I have

just done provides a fresh perspective on the causal relationship between genotype and phenotype. For it becomes possible to view the genome as a package of resources assembled to equip the next generation with much of the wherewithal to survive and prosper (reproduce) in environments like those encountered by the current generation. This teleological stance has the effect of turning the direction of causality on its head, making the genome appear if anything to be the instrument of the organism.

Conclusions

In this chapter I have defended the idea that one can think of genes and genomes in informational terms without thereby being committed to the view that they have causal priority over other cellular (or indeed extracellular) structures and processes. I have done so by adopting a functional approach. Rather than seeing information as being the basic ontological category at issue, I have framed my account in terms of the informational roles that structures and processes can play within larger systemic contexts. These roles include storage, reading, writing, and transmission. To identify these different roles with the parts of a system requires us to make a functional decomposition of the system, and I have suggested that this is a matter of recognising how genomic / genetic structures and processes can be accommodated by a variety of functional schemas. The structural distinctiveness of DNA, and the profound biological importance of sequence to which the existence of a variety of sequence-focused cellular mechanisms attests, are key. These features, I suggest, fit some of the functional decompositions we make of systems we think of as storing and processing information – even though, as I argued, it is hard to articulate any clear folk sense of information.

This approach to biological information has one especially important consequence: a particular entity or process can potentially occupy an informational role in multiple system decompositions, and these can be decompositions of very different systems. For example, a particular segment of DNA may occur in two different systems and be read by reading units that are connected to different coupling structures or processes in the two systems. These coupling units might in turn connect with functionally quite distinct processes, so that very different ends are realized by the two systems. In both cases, however, we are justified in saying that the DNA segment is fulfilling an information storage role. Some may wish to argue that the segment 'means' different things in the two systems, but semantics is not really the issue. We are misled into thinking that a book, say,

has a meaning that inheres solely in the text itself just because our minds are structured so similarly (whether by nature or nurture is incidental) that the text gives rise to similar effects in the minds of different readers. What matters for the ability to associate an informational storage role with the book is the reasonableness of the assumption of readers and writers, i.e. of an overall functional system architecture that includes within its functional decomposition the role of information storage.

These reflections on informational roles within systems lead back to the nature of functions and functional attributions. They suggest a somewhat holistic picture: an individual function is a component within a larger structure, namely an overall functional decomposition into which an entire system fits. If this is so then maybe sometimes whether a particular part of a system can be assigned a function depends on the possibility of fitting other parts of the system into the overall decomposition of which that function is a component. (But if a good fit exists within one region of the decomposition then a looser fit might be tolerated elsewhere.) The fact that explanation in chemistry and physics seems to have no need for functional ascriptions could then be related to the fact that they abstract phenomena from specific systemic contexts in their quest for generality. Those are circumstances in which functions can play no part.

My approach to biological information represents an attempt to answer the question that anti-informationists leave hanging in the air: why it is that information talk remains pervasive in biology, and why it is associated in particular with genetic and genomic phenomena. At the same time it shows that their concerns about causal priority are unfounded.

8. Mind in biology

In this thesis I have investigated how a number of related philosophical topics intersect with several connected areas of research in molecular and cell biology, in order to gain a better sense of what explanation and understanding amount to in those areas and more generally, as well as obtain insight into what factors constrain them. In this final chapter I recapitulate the main flow of the argument to highlight and reflect further on some of the issues it raises, before closing with some more general speculations and indications of possible directions for research.

To begin with it is worth restating the main areas of claim-making that the preceding chapters have encompassed. As was advertised in Chapter 1 these fall under five distinct but related headings:

1) mechanism in molecular and cell biology;
2) protein folding and the epistemology of molecular dynamics simulation;
3) cellular complexity;
4) emergence;
5) functional attribution and information talk in biology.

The overall perspective that emerges from what I say about these topics is one that emphasizes the potential significance of epistemic – in the sense of psychological – factors for our understanding of them and the philosophical concepts with which they connect. As extensive discussion of a variety of biological phenomena has made plain, however, it has not been my intention to downplay the importance of objective and external considerations. This ontic/epistemic even-handedness argues for the possibility of integrating internalist and externalist thinking in order to create what could be described as a systems approach to epistemology.[91] By this I mean an approach that regards the total

[91] Surprisingly little explicit overlap exists between what I have said in this thesis and the views presented by Godfrey-Smith (1994). When he talks about internalism and externalism he seems generally to have in mind different perspectives on why organisms come to have the organization and capacities they do. However, his argument that the complexity of minds reflects the need for organisms to cope with life in complex environments is highly compatible with what I say. A more direct connection with his argument could be effected by combining my claim that phenomenal complexity is epistemically relative to our cognitive capacities with another idea. This is that the complexity of those capacities is such as to position the boundary between what seems complex (i.e. seemingly threatens to overwhelm an organism's cognitive abilities) and what does not at just that point which, minimally and at the population level, is compatible with survival.

system of mind in the world as potentially relevant to philosophical understanding of topics bearing on scientific explanation, since sometimes the relevant concepts operate at (and serve to locate) the boundary of mind and world in the sense illustrated by my treatment of emergence.

I began by outlining the legacy of logical positivism, in the guise of the logical empiricist programme that dominated philosophy of science in the aftermath of the Second World War. Philosophers of science around this time devoted much of their energy to the construction of formal structures representing the relationships between scientific observations and theoretical statements and generalizations. The belief was that these structures would collectively provide an objective view of how science works – how it explains the world – without making reference to metaphysically problematic notions such as causation. The idea was articulated in part via the notion of theoretical subsumption, or reduction in the classic sense articulated by Oppenheim and Putnam which I discussed in Chapter 1, in terms of which it was hoped that it would be possible to explicate the 'logic of science'.

The problem with explanation as reduction is its requirement that a scientific field's conceptual content be represented (representable) in formal terms, i.e. as a set of nomic generalizations or laws to which its phenomena conform. Such laws are rare in many areas of biology, and this is related in large part to the causal complexity of biological phenomena. In Chapters 4 and 5 I discussed complexity at the cellular level in part in terms of fluidity and the diverse nature of the causal processes that sustain cellular life. Rosenberg has argued that the fact that biology is a 'nomological vacuum' is a reflection of the fact that natural selection filters populations by function rather than structure – so structural variation is tolerated if it is not dysfunctional and flourishes if it is adaptive (Rosenberg 2001). The connection between the cellular complexity I have described and biological diversity of the sort Rosenberg presumably has in mind is not obvious – even if there were no biological diversity the causal complexity of the cell would still militate against its simple formal description. Causal complexity at all scales is the usual state of affairs in biology, however, and where mathematical generalizations are possible they tend to be statistical in nature, concerning the collective behaviours of molecules, cells or organisms, or are of application only to specific systems and circumstances. When laws do apply to phenomena they are not usually held to be constitutive of the phenomena; rather they are merely contingently descriptive.

Even when the goal of mathematicization is abandoned, finding verbal generalizations that subsume and potentially explicate large classes of biological phenomena is often far from straightforward: for every putative rule there is (as a rule) a large set of exceptions (see e.g. in relation to genomics Dupré 2004, p.324). This biological lawlessness is problematic for accounts of explanation of the kind developed within the logical empiricist tradition, such as Hempel's D-N model, which construe explanations as arguments resting on nomological generalizations. In Chapter 1 I outlined some of the problems associated with these accounts, which relate in the main to the eschewal of causal notions. Intense philosophical activity focusing on causation, building on centuries-old debates, has revealed it to be a far from tractable concept, especially if it is approached from an exclusively externalist point of view. Hume famously argued that there is nothing given to us in phenomena as a basis on which to pick out causal relations beyond 'constant conjunction' and other regularities in the succession of events. Yet objective phenomena such as spatiotemporal event correlations are often ambiguous or misleading about the causal properties and powers of objects. Despite this, we very often *are* able to predict and explain phenomena in causal terms. How are our successes to be accounted for if they are not down to luck? In Chapter 1 I discussed a variety of accounts of causation that have been proposed, most of which break down when applied to particular kinds of case. This thesis has been shaped by the under-explored idea that causal understanding is grounded in cognitive models and schemas with which we are able to parallel phenomena in the world.[92] My suspicion was that this perspective on causation would be found to connect in an interesting way with the concepts of mechanism currently attracting the interest of philosophers of science.

Mechanism, complexity and emergence

This then was the background against which I discussed, in Chapter 2, those concepts of mechanism. I started by distinguishing between material and causal senses. Machines exemplify the former sense, and I argued that the alignment of structural and functional decompositions is one of their key features. But the two senses are related inasmuch as machines can be thought of as the material instantiation of particular causal structures or entailment patterns. The discussion of function in Chapter 2 revealed the concept to be philosophically problematic, since functions are not given to us in any direct

[92] Causal models could originate in a variety of ways. Some may be innate and others may be learnt (perhaps on the basis of innate capacities to acquire such models given particular forms of experience and interaction with the world), while still others may be discovered or constructed.

way by the phenomena we study scientifically. Where a man-made machine is concerned we see the functions of its parts in relation to the overall function that is attributed to the machine. The function of a machine is what it was designed to do, or the purpose to which its inventor imagined it being put. The overall function or goal of an organism we generally see as the maintenance of its own life and the generation of offspring, for it is apparently towards these ends that the organism's resources are directed.

After discussing the machine and causal senses of mechanism I looked in some detail at the MDC account that has been so influential in recent years in philosophy of biology and philosophy of science. This account as I see it stems from several motivations. Chief amongst these are to make sense of mechanism talk in biology, i.e. to understand the heuristic explanatory value of the conceptual content of such talk, and to articulate a more normative conception of mechanism capable of playing a major part within a philosophical account of explanation. I argued that the account is rather ambiguous, especially as regards the ontic twins of entities and activities and how these stand in relation to structure and function. On the other hand MDC make the interesting epistemological proposal that intelligibility is a matter of connecting a phenomenon with sensory experience, and suggest that mechanistic descriptions explain by 'showing how' phenomena are brought about. The understanding that mechanistic descriptions confer is not just a visual affair, however: it may involve other aspects of bodily experience such as proprioception. This epistemic aspect of the MDC account is I think also its most potentially fruitful, and it connects with much of what I have said in Chapters 4-6 in relation to complexity and emergence.

Before I got on to those topics, however, I explored what at the outset I had suspected would be an interesting intermediate case situated between mechanism and non-mechanism: the phenomenon of protein folding and what it means when scientists talk about the mechanism of protein folding. I argued that protein folding is mechanistic in only a schematic sense related to the causal sense I outlined in Chapter 2. As if to lend weight to the epistemic ideas of MDC concerning intelligibility I found that the science of protein folding makes extensive use of visual metaphors, generally of a topographical nature. Protein scientists talk frequently about potential energy landscapes, funnels, barriers, wells, troughs and peaks. This figurative language provides a means of visualizing some of the causally relevant characteristics of complex stochastic processes from which hard and fast structure—function relationships (of the sort found in true machines) are generally lacking. In addition protein scientists have long depended on techniques for visualizing the molecules that interest them. Simplified representations of various kinds –

ribbon models, for example – serve to emphasize particular features such as secondary structure. Interactive 3-D computer graphics techniques are especially important, and I think that it is helpful to think of them, like simulation methods, as a kind of cognitive prosthesis. Chiefly what they compensate for are constraints that exist in relation to our ability to view a 2-D image and infer from it – and retain in our minds – a 3-D structure. They do this by allowing the researcher to manipulate structures and view them from different orientations: the mere act of rotating a molecular model on screen in real time is sufficient to form an impression of the molecule's shape. The perceptual basis for this is the changes that occur under rotation in the relative positions of the molecule's atoms, in conjunction with depth cues that serve to distinguish between front and back and hence disambiguate the direction of rotation. The possibility of adding 'haptic feedback' underscores the idea that intelligibility can be more than a merely visual matter.

In Chapters 4 and 5 I turned from complex molecular structures to the nature of cell complexity, thinking initially that in the cell I would find clear-cut examples of non-mechanistic phenomena and indeed perhaps an instance of a thoroughly non-mechanistic system. To some extent this expectation was gratified, if (as I argued we might) we think of non-mechanistic phenomena as ones in which structure—function relationships are complex, opaque or possibly even non-existent. (Given the epistemic, subjective nature of functional attribution the last must be considered at least in principle a live possibility.) Metabolism was found to provide examples of phenomena in which properties and behaviours cannot be identified with parts (structures or processes) in isolation but rather arise as a result of the combined actions and interactions of multiple parts. Yet the fundamental properties of cells, such as genome replication and cell division, and the capability for adaptive behaviour and environmental compensation, seem to allow for the existence of an underlying overall functional architecture or causal logic. I discussed the idea that many cell processes depend on hyperstructures, and used this notion of a category of non-equilibrium and equilibrium structures that implement particular functions as the basis for a new perspective on mechanism. My aim was to build on the MDC account by recognizing the value of that account's neutrality about the relative causal status of structures (entities) and processes (activities, roughly speaking). At the same time I wanted to find a way to rehabilitate function, and provide the means to distinguish between equilibrium and non-equilibrium phenomena. Achieving the latter was important because it bears on a key difference between a machine's organization and that of a living system. I found that the best way to accomplish these goals was to jettison the entity/activity ontology and structure my account instead around a class of epistemic objects I refer to as

P-structures. These are the psychologically grounded structures and processes we individuate and use in our explanations of phenomena, and to which we attribute a variety of functions and properties.

Thinking of mechanisms epistemically in terms of P-structures provides the flexibility needed to subsume much of the diversity of mechanism talk in science, whilst avoiding the reification of functions or causal properties. (It will be recalled that Torres has argued that reification of the latter is one of the drawbacks of the MDC account, especially in relation to negative causation.) The result is to loosen the bindings of mechanistic ascriptions and descriptions to underlying physical processes, and in effect engineer a conflation of the machine and causal senses of mechanism discussed in Chapter 2. The concept of mechanism that results is I think best thought of as a cognitive 'resonance structure', by analogy with the representation of certain molecules as resonance structures that combine the properties of several distinct electronic configurations. This conceptual structure involves two components: a set of a functionally related P-structures and a pattern of entailment in phenomena, each of which provides a partial basis for ascribing mechanism:

[**M0**] set of P-structures ↔ pattern of entailment

The point of representing the concept in this way is that it is more cognitively realistic than the simple conjunctive definition of a mechanism as a set of P-structures that instantiate a pattern of entailment. Sometimes the set of P-structures involved in a mechanism will be understood only incompletely or schematically, although the pattern of entailment they give rise to is well defined. In other cases the structures may be understood in detail, but their causal consequences are less thoroughly appreciated. But when both the identity of the P-structures and an associated pattern of entailment are stable and well-defined then the ascription of mechanism is made confidently and readily. This resonance conception of mechanism can also be thought of as the combination of separate definitions emphasizing the two aspects:

[**M1**] a set of P-structures [instantiating a pattern of entailment]

[**M2**] a pattern of entailment [instantiated by a set of P-structures].

It can be seen that **M1** is akin to the machine sense in that it stresses a mechanism's materiality, while **M2** leads with the causal story (and if we bracket off 'instantiated by a set of P-structures' then indeed it reduces to the causal process conception). Because P-structures can incorporate the properties we associate with structures and processes, including collective properties such as cytoplasmic fluidity, these definitions embrace mechanism descriptions that incorporate such properties as explanatory factors. A mechanism in the sense of **M0, M1** or **M2** can thus be a good deal more schematic than the traditional machine conception, as I think is necessary if scientific language use is to be assimilated adequately.

One can have too much of a good thing, however, and it can rightly be objected that in relation to biological systems this perspective is so flexible as to subsume rather more than we might always wish. In concluding Chapter 5 I remarked that additional constraints would be needed in order to distinguish living from non-living systems, and these would probably relate to metabolic circularity, autopoietic capacity, or intrinsic teleology, or some combination of these. I suggested that interesting inter-relationships exist between all these concepts, although to work them out satisfactorily would require more time and space than has been available to me. Perhaps, however, it is feasible to give fuller and more explicit expression to thoughts concerning structure—function relations than I have given so far. I have alluded several times to the idea that if a system is mechanistic then it is possible to associate functions with system parts in a fairly straightforward way. And conversely I have said that non-mechanistic systems are generally ones in which structure—function relations are complex, problematic, opaque or non-existent. But to construe the degree of mechanisticity in this way seems if not at odds with the definitions given above and the perspective from which they arose then at least to stand in uncertain or orthogonal relation to them.

Perhaps these problems to do with structure—function relations can be solved by recognising different classes of mechanism according to the nature of such relations. The most straightforward cases are entailment systems in which specific functions can be identified for each and every P-structure in the system. Then we can imagine systems in which a pattern of entailment can be identified with a particular configuration of P-structures, only a subset of which can be associated with a particular function. (Perhaps. Would we not tend to assign coupling or connective functions to otherwise 'functionless' P-structures?) Different again will be entailment systems in which one or more functions are distributed over multiple system components, or in which in some other way a

functional decomposition fails to map straightforwardly to material structures and processes. My inclination is to see mechanism as being basically a causal matter, to do with the ability to see or imagine (and therefore specify) what (in terms of P-structures) brings something about and how they do it. When we can do that we can in principle (in a Woodwardian counterfactual or ideal sense) intervene in or manipulate the phenomenon. Perhaps sometimes – my intuitions here are rather weak – we can do this without attributing functions to parts of a system, and if this is so then it argues for providing a space for functional attributions in one's mechanism account (as I do), but not for making them mandatory.[93] But then a lack of mechanisticity has to be cashed out in terms other than structure—function correspondence, such as haziness of the association between entailment and P-structure identity.

This raises another, perhaps related, problem: how we should view the relationship between mechanism, emergence and cognition. I argued in Chapter 6 for a negative account of emergence, whereby a phenomenon is emergent if we lack the cognitive schemas needed to trace through or model the entailment structures giving rise to it. When such schemas are to hand (or, rather, mind) phenomena strike us as intelligible, explicable and only to be expected. If functions are highly distributed across numerous P-structures rather than highly localized then prima facie this seems likely to make entailment modelling cognitively taxing, in line for example with what Resnick has said about our bias towards a 'centralized mindset' (Resnick 1996). We need to be clear about what we mean by distribution, however.

As a thought experiment, suppose one were somehow to reify an 'exploded view' of the clock discussed in Chapter 2, say, and draw apart the parts that this reveals so that they are highly separated in space. (We must imagine the parts suspended in space, in defiance of gravity.) Imagine then surrounding each part with a set of sensors and actuators that communicate (by radio frequency waves, perhaps) with the sensors and actuators surrounding just those other parts with which the part is normally in mechanical contact. We can imagine the system of sensors and actuators being configured so as to establish the same pattern of causal inter-relationships amongst the exploded set of clock parts as normally exist amongst the parts of the unexploded clock. In this case there would be a sense in which the exploded clock and the normal clock exhibit a similar functional architecture, although the parts are more distributed in the sense of being more spatially

[93] However, I suspect that the processes of functional attribution and P-structure individuation are often closely and synergistically coupled.

dispersed. (Note that even the relative dispositions of parts could be altered without changing the functional architecture, provided that the sensor/actuator system implemented the same pattern of inter-relationships as exists in the unexploded clock.) For each part in the normal clock there is a functionally equivalent part in the exploded clock, albeit now associated with a system of sensors and actuators.

Davies (2008) argues that the concept of causal distribution versus localization serves to capture important characteristics of certain biological problems and approaches, and helps to account for our bias towards simple causal analyses. Thinking about the exploded clock helps us to interpret him correctly, however. Such examples show that distribution qua dispersal in space does not necessarily make for epistemic impenetrability: exploded views of complex artefacts are intended, after all, to facilitate comprehension by showing the parts that make them up and how they relate spatially. What matters more for the purposes of Davies' argument is that sometimes causal responsibility for a function is distributed in the sense of being divided across multiple structures or sub-systems. It is the combined activities of these – their partial contributions – that implement the whole function, and sometimes this functional distribution confers redundancy in that not all of the structures are required for functional implementation. Exactly how distinct the issues of spatial distribution and functional division are, and how they relate to functional architecture, is not entirely obvious, however. Perhaps the distribution of a function over multiple entities implies an increased burden on working memory, relative to the load presented by phenomena in which functions are concentrated within fewer entities. But this is more a matter of having to keep track of multiple entities in order to monitor system state than it is of spatial distribution.

The implication of all this is that from the point of view of epistemic tractability as much interest potentially attaches to the comparison of different functional architectures that fulfil the same overall goals or which exhibit the same overall causal properties as it does to different patterns of functional distribution. A system in which each part contributes towards each and every function of the system's functional decomposition would presumably be deemed more complex than one in which the same functions were associated with parts in a more modular fashion. But an overall system function might be implemented in accordance with quite different functional decompositions (think about the different ways in which electricity can be generated, for example), some of which may be more intelligible than others.

Irrespective of issues of this sort to do with functional architecture, on a Cummins-style (causal role in a system) account of function there is a close relationship between causal schemas and functional schemas. I noted in Chapter 2 that functions can be associated with sets of counterfactual structures, any member of which would fulfil the function to enable normal system operation. We can therefore think of a functional schema as a meta-causal schema – a causal schema at an additional degree of schematic abstraction. For this reason something like the account of emergence I outlined in Chapter 6 subsumes cases in which what we lack is not a causal schema as such but a functional schema, i.e. a set of counterfactual causal schemas, for a phenomenon. To be functionally emergent goes beyond being non-mechanistic in the sense of lacking a functional analysis that maps neatly onto P-structures. It means instead fulfilling the more stringent condition of having no obvious functional decomposition at all.

In some ways the account of emergence outlined in Chapter 6 represents the complement of what I have said about mechanism. The latter concept pertains, I have argued, to the association of a pattern of entailment with a more or less determinate structural or processual basis, which I explicated in terms of P-structures. An emergence attribution on the other hand points to a failure to correlate a pattern of entailment with any such material basis, through lack of a cognitive schema with which to model the entailment structure. The difference is thus one of cognitive connectability, and perhaps there is a link here with reduction and reducibility. If so then the conceptual relationships are not straightforward, however. I argued that emergence is associated with a lack of connectability in the sense, typically, of imaginability or simulability – although I was equivocal about the extent to which the latter should be thought to involve conscious experience or deliberate cognitive activity. Sometimes we do visualize how a process occurs – for example, knowledge of a little chemistry enables us to generate internal images in response to the invitation to imagine an ethane molecule and then imagine rotating the two ethyl groups around the bond that joins them. That is feasible enough, but imagining the effects of rotating around several of the rotatable main-chain bonds of a polypeptide structure is not, I assume. (Does this make protein folding an emergent process? The answer is perhaps not clear-cut: to the extent that the process is unimaginable, yes; but to the extent that we know what the relevant physical factors are and how they operate schematically to yield a folded polypeptide, arguably no.) On the other hand I think many emergence attributions are to be understood as arising from a sub-conscious inability to trace or follow the causal connections underlying a phenomenon. The involvement of

mirror neurons and similar neurological mechanisms in behavioural comprehension might count as evidence in favour of – or at least compatible with – this kind of view.

The topic of genetic determinism discussed in Chapter 7 is an interesting case to consider in relation to the topics of reduction and cognitive and conceptual connectability. Up until the 1970s (when the part genes play in development became better understood) molecular biologists propounded the idea that an organism's genotype specifies its phenotype, and seemed to imply that it does so in a linear way, gene by gene. The Central Dogma, with its simple readily visualized schema, helped to consolidate genocentric patterns of thinking by deflecting attention from the contextual factors that influence gene action and focusing just on DNA-proximate processes. These processes can certainly be thought of as mechanisms in my sense as far as the production of mRNA goes, for we can readily identify the relevant entailments and P-structures. In conjunction with knowledge of the RNA splicing 'machinery' our mechanistic picture of gene expression can be extended a little further out into the cell, where it comes into contact with other pockets of mechanistic understanding focusing on regulation, metabolism, signalling and so on. But whatever the prospects are for integrating all these functionally specialized sub-mechanisms into an overall mechanistic picture of the cell, the vision that straightforward connections in general exist between genotype and phenotype must be discarded.

As I argued in Chapter 7, however, this is not to say that the concept of information, which has often been associated with genetic determinism, has no part to play in how we conceptualize the cell. The reified view of information encouraged by the Central Dogma, which construes it as something that flows outwards from the genome, was surely misleading. But the alternative perspective I have presented, emphasizing concepts of information storage and transmission, makes sense of the stability of chromosomal DNA sequences. Moreover it is consonant with the structural and functional schemas in terms of which we interpret the working of a variety of objects and products we regard as unambiguously informational, such as books, CDs and magnetic tapes.

Explanation and understanding in biology

To return almost to where I began, in Chapter 1 I reflected on some of the ways in which biological fields are liable to be distinctive as regards their explanatory character. I speculated that particular areas of biology are notable in three respects: their comparative

lack of law and theory, their heavy emphasis on the visualization of phenomena as a primary research objective, and the important part played by functional concepts. On the basis of the preceding chapters it is reasonable to conclude that the second two characteristics are ones that molecular and cell biology share. More doubt attaches to the status of laws and theories, in that contemporary systems and synthetic biologists argue that they are discovering general abstract principles about the workings of cells – expressed for example in terms of network motifs (Alon 2007a) – and indeed laws (Boogerd et al. 2007, p.332). But even if this is so, it is I think reasonable to say that laws and theories of the kind we know from physics have not played a major part in organizing or codifying (and still less constituting) the knowledge acquired in mainstream molecular and cell biology.

Given these findings we can ask how de Regt's account of understanding, also discussed in Chapter 1, fares in relation to mainstream work in molecular and cell biology. It will be recalled that his view is based on a semantic conception of theories, in which models occupy a mediating position between phenomena and theory (Morgan and Morrison 1999). Skill and judgement must be exercised in order to develop perspicuous explanatory models, and a premium is placed on the intelligibility of theories. As a basis for explicating understanding in molecular and cell biology this sort of account seems on the face of it to be altogether too complex and theory-oriented. (And it is disconcerting that de Regt endorses a broadly Hempelian view of explanation.[94]) The putative laws and general principles of systems and synthetic biology apart, most work in biology that focuses on the cell has traditionally involved identifying and correlating macromolecular structures, processes and functions. This is frequently a visual business, as the case of protein science well illustrates, and I now want to examine the implications of the larger claim that understanding in these areas of biology has traditionally been largely a matter of visualizability and perhaps other forms of 'sensibility'. In service of this goal research has emphasized the characterization of macromolecular structures and elucidation of their origins, interactions and fates. Commonly the structural and interactional knowledge furnished by these activities allows functions to be correlated with particular structures and processes, and when this is done a mechanistic 'picture' emerges of how a particular cellular phenomenon occurs. The picture here is not a simple static image, however. Rather I propose that it can be thought of as a set of sensory 'frames' (whose contents are visualizable or otherwise sensible structures or schematic representations of processes, i.e.

[94] 'I agree with the basic Hempelian idea that explanations are arguments which attempt to fit a phenomenon into a broader theoretical framework' (de Regt 2009, p.25).

derivatives of the P-structures described earlier) associated with dispositions to connect frames together in ways that model the patterns of causality seen in molecular and cell biological phenomena, often on the basis of properties we attribute to frame contents. These visuo-sensory frames plus associated connective dispositions or schemas constitute the cognitive mechanisms by which we are able to imagine phenomena by paralleling them in our minds, and on the basis of which we can generate verbal descriptions when asked to explain them such that our interlocutors can 'see' what we are saying.[95] These cognitive mechanisms are explanatory because they tell us what structures and processes participate in a phenomenon (frame contents) and how the phenomenon comes about (via the dispositions to connect frames in causality-modelling ways). (These ideas are closely related to the conjectures I made in Chapter 6, in that we can say that sometimes visuo-sensory frames are generated at 'run time', so to speak, by simulation schemas.)

This preliminary, and somewhat dispositional, visuo-sensory account of understanding is consistent with – and unpacks a little further – MDC's proposal concerning intelligibility, according to which it is a matter of making connections between a phenomenon and sensory experience. This is not necessarily incompatible with the perspective de Regt outlines, although it is unclear where theories fit in the account, which involves just phenomena and cognitive mechanisms for paralleling them. Should we say that these mechanisms are theory and model in one, or that the cognitive mechanisms in my sketch are to be equated straightforwardly just with the models of the semantic view (with theories being factored out of the picture)? Then again to what extent are the models of semantic accounts to be thought of as logico-linguistic structures, and what is their relationship to the visuospatial models that feature in molecular modelling?[96] Perhaps one of the lessons of Chapter 3, concerning protein folding and MD simulation, is that there *is* a theoretical foundation to the molecular phenomena of the cell, but it is a micro-scale physico-chemical one (to do with inter-atomic interactions and so on) that is cognitively disconnected from the biological understanding obtained from visualization on account of the causal complexity of the relevant processes. (Note that this is compatible with my rendering of downward causation as inward causation, in which causes can converge on a region from distant and widely distributed points to act in highly non-linear ways. It is thus not a microreductionist view.)

[95] They can also perhaps be thought of as examples of the 4-D representations for which Griesemer argues (2004, p.434).
[96] Giere (1988) for one emphasizes the non-linguistic nature of models.

To answer such questions requires us to be clearer about what models are, and the difficulty is that the concept has been interpreted in a variety of ways (Morgan and Morrison 1999, Chapters 1 and 2; Morrison 2007, esp. Section 2). The received or syntactic view of models as abstract deductive structures is I take it of very little relevance to the sensory perspective on understanding I am outlining. Van Fraassen's semantic account on the other hand incorporates the idea of a state space, which Morrison and Morgan explicate in these terms:

> If we think of a system consisting of physical entities developing in time, each of which has a space of possible states, then we can define a model as representing one of these possibilities. The models of the system will be united by a common state space with each model having a domain of objects plus a history function that assigns to each object a trajectory in that space. A physical theory will have a number of state spaces each of which contains a cluster of models. For example, the laws of motion in classical particle mechanics are laws of succession. These laws select the physically possible trajectories in the state space; in other words only the trajectories in the state space that satisfy the equations describing the laws of motion will be physically possible. Each of these physical possibilities is represented by a model.
>
> (Morrison and Morgan 1999, p.4)

There are potential points of contact here with both the account I gave of protein folding science in Chapter 3 and the particle tracking ideas I discussed at the start of Chapter 4. A polypeptide's conformational space is the state space within which folding takes place, and the laws of interatomic interaction are what in Morrison and Morgan's terminology 'select' the physically possible trajectory of a polypeptide through that space. Each physically possible trajectory in Van Fraassen's account is represented by a model which, as Morrison and Morgan put it, incorporates 'a history function that assigns to each object [in the model] a trajectory in that space'. Thus a phenomenon such as the folding of a particular polypeptide is represented by a *family* of models rather than by a single model. It is this family of models that seems closest to my conception of a cognitive mechanism, although there are difficulties with this idea in that the models incorporate more physics than we are capable of cognizing. Van Fraassen's models must for this reason be considered more as ontological constructs than as epistemic structures we put to practical use in the ways I have in mind for cognitive mechanisms.

A more straightforward idea is to think of both models and cognitive mechanisms as representations of phenomena (Morgan and Morrison 1999, Section 2.3; Giere 2004; Morrison 2007, pp.210-217). Skill and judgement enter into de Regt's account of

understanding chiefly in relation to the development of perspicuous models from theories. In the perspective I outlined above understanding and explanation are two sides of the same coin. The sensory frames and connective schemas that together instantiate the visuo-sensory type of understanding that I suggest is central to the epistemology of molecular and cell biology are explanatory because they represent phenomena so that we can model them cognitively. This enables us to make predictions, construct hypotheses, design interventions that test our knowledge, and generate verbal and pictorial descriptions that communicate our knowledge to others. When we have an adequate cognitive mechanism we can display our understanding of a phenomenon by inspecting and interrogating frame contents or by imagining an entailment pattern according to our dispositions to connect frames in causality-modelling ways. Differences in understanding on this view could come about through differences in the facility with which structures or processes can be visualized (or other sensory associations activated) and in the capacity to coordinate frames with connective dispositions.

Mind in biology

The principal aim of this thesis has been to illustrate how epistemic and ontic factors interact to shape our understanding of biological systems and the explanations we give of them. This is perhaps seen most clearly in relation to emergence: I have argued that phenomena are described as emergent when and because they fail to fit our causal cognitive schemas. But in relation to mechanism and complexity my position is similar, in that what strikes us as a mechanism and what seems complex are also epistemically relative matters. The concept of function has played a central part in my discussion of these concepts, yet it remains elusive and difficult to work with. One possibility – which Ratcliffe (2000) points towards – is that this is because functional talk is as epistemic an affair, is as symptomatic of internal cognitive states, as are (say) emergence attributions. Perhaps, however, we can say that one of the functions of functional attribution is to act as an orienting device. When we know the function of something we are primed with a schematic set of causal expectations about how it will behave in particular circumstances, and about the likely systemic or contextual effects of its removal or disruption.[97] This I suggest is what underlies the importance biologists attach to determining the functions of

[97] Lombrozo and Carey (2006), in their account of 'Explanation for Export', propose the broadly compatible idea that the function of explanation is 'to provide the kind of information likely to subserve future intervention and prediction'. They argue that functional explanations point to general causal patterns of potential predictive utility (pp.195-197).

the structures and processes that make up living systems, as was evident in the several recent examples from the literature mentioned in Chapter 7. There I also noted how reflecting on the role of the genome in transmitting phenotypic characteristics from one generation to the next encourages a teleological mode of explanation. That topic is too complex for a detailed exposition to be attempted here, but further work would I think show it to connect with my overall perspective in potentially fruitful ways.

The teleological mode of explanation works, it will be recalled, by treating phenomena as though they come about on account of events and processes that are directed at bringing them about or that intend them to come about.[98] This seeming end-directedness I propose is a matter of phenomena fitting cognitive or behavioural schemas of some sort, and again (as with emergence) perhaps these schemas can be highly diverse. Some forms of adaptive or purposive behaviour, such as those of the hunting protozoans (and Grey Walter's robotic turtles) described by Bray (2009, Chapters 1 and 2), presumably conform to the kinds of folk psychological schema I discussed in Chapter 6. The connotation of psychological characteristics in respect of non-psychological systems that results from this has led to teleological explanation acquiring a slight air of philosophical disreputability. Another class of teleological explanations involves the adoption of a 'design stance' or 'artefact model', in which a system is described as if it were the creation of a rational designer (Lewens 2005). (It has been suggested that this kind of explanation is especially common in elementary-level presentations such as 'popular science' publications (Lewens 2000, pp.100-102). In this regard it is interesting to note research suggesting that this is a basic explanatory mode to which young children default, before education instils increasing capacities to explain phenomena in causal-mechanical terms (Keleman 1999).) The idea here is that the system engages the cognitive and action schemas that would be invoked were we to be tasked with constructing a system to fulfil the function we attribute to it. Sometimes, however, perhaps a system is teleologically explicable when its behaviour is the result of a seemingly coordinated set of processes. Here it may that we project ourselves into the driving seat, so to speak, and find consonance between on the one hand the inter-relations between system processes and the overall system state that results and on the other the pattern of processual coordination that we would seek to establish were that system state a goal we were responsible for achieving.

[98] Psychological work suggests that it is a mode of explanation to which we are strongly disposed, in relation not only to human behaviour (Rosset 2008) but also to natural phenomena (Keleman and Rosset 2009).

The picture that results from these reflections and the more detailed discussions of the preceding chapters is a naturalistic one that associates the content of philosophical debates with our embodied cognitive situation in the world. The above conjectures about teleological explanation show the diverse nature of our explanatory capacities, and reinforce the point illustrated by my treatment of emergence that they can often be understood in terms of cognitive schemas of various kinds. I have argued that explanation and understanding in molecular and cell biology make heavy use of the visual and other sense-proximate forms of cognition that have been relatively neglected by philosophers of science, but it is clear that other forms of cognition are also important. James Clerk Maxwell noted the relevant distinction:

> There are ... some minds which can go on contemplating with satisfaction pure quantities presented to the eye by symbols, and to the mind in a form which none but mathematicians can conceive. There are others who feel more enjoyment in following geometrical forms, which they draw on paper, or build in the empty space before them.
>
> (Maxwell 1890, quoted in Cat 2001, p.407)

Maybe the new more abstract perspectives of systems and synthetic biology represent a major shift in the nature of biological knowledge in the direction of a framing in terms of 'pure quantities presented to the eye by symbols'. It is probably still too early to assess the prospects for developing an all-embracing 'theory of the cell', but if any such thing proves attainable it seems likely that it will be something akin to a schematic functional algorithm rather than a theory of the kind we know well from physics, expressed in terms of succinct formulae in which time and/or space appear as independent variables. It would be extremely interesting for this reason to investigate the relationships between sense-proximate thought, language, and more abstract and logically based forms of reasoning.[99] A fundamental issue in this regard is likely to concern the plasticity of different cognitive capacities, in terms of the extent to which they are 'hard-wired' into the brain.

It seems plausible to suppose that sense-proximate thought, including the capacity to simulate phenomena visually, is more developmentally entrenched and less plastic than capacities for more abstract patterns of thought. Neuroanatomical differences between different brain regions tend to support this view. Logical reasoning and abstract thought

[99] The tension between visual and abstract thought surfaces even in relation to quantum mechanics. Schrödinger wrote that he 'felt discouraged, not to say repelled' by Heisenberg's transcendental algebraic approach, 'which appeared very difficult to me, and by the lack of visualizability'. Heisenberg meanwhile said that 'the more I reflect on the physical portion of Schrödinger's theory the more disgusting I find it. ... What Schrödinger writes on the visualizability of his theory ... I consider trash' (quoted in Yourgrau 2007, p.83).

are associated with the more recently evolved areas of the neocortex, whereas sensory areas correspond to some of the evolutionarily oldest regions of the brain. The neurons in these older regions are more heavily myelinated than those of the neocortex, and myelin sheath formation occurs earlier, suggesting a relative lack of developmental plasticity (Fields 2008). What are the implications of this? Does understanding necessarily require the participation of sense-proximate mental processes? What are the constraints on abstract thought, and how constrained is it, if it depends on more plastic neurological structures? Can any sort of philosophical connection usefully be made between the sense-proximate versus abstract/logical distinction and the ideas of, say, Sellars (1963/1991) or McDowell (1996)?

Another set of directions for research flows from the fact that I have discussed my chosen themes from an obviously individualistic point of view, focusing on aspects of the cognitive psychology of humans as isolated epistemic agents. One natural course would be to broaden scope to consider the role of sociological factors, whilst another would be to attend much more to the practices and technologies surrounding scientific problem solving. These approaches and themes are already well represented, of course, in thriving programmes in philosophy and sociology of science. The relevance to such programmes of the ideas discussed in this thesis stems from the fact that explanation comes to ground in individual understanding, and often explanations function, in a surprisingly literal way, by making sense of things.

Bibliography

Achinstein, P. (1983) *The Nature of Explanation* (New York: Oxford University Press).

Alberts, B.M. (1984) The DNA enzymology of protein machines. *Cold Spring Harbor Symposia in Quantitative Biology* **49**: 1–12.

Alberts, B.M. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**(3): 291-4.

Albiges-Rizo, C., et al. (2009) Actin machinery and mechanosensitivity in invadopodia, podosomes and focal adhesions. *Journal of Cell Science* **122**: 3037-3049.

Alon, U. (2007a) *An Introduction to Systems Biology: Design Principles of Biological Circuits* (London: Chapman and Hall).

Alon, U. (2007b) Simplicity in biology. *Nature* **446**: 497.

Amar, P., et al. (2002) Hyperstructures, genome analysis and I-cell. *Acta Biotheoretica* **50**: 357-373.

Anfinsen, C.B. (1972) Studies on the Principles that Govern the Folding of Protein Chains. Nobel Lecture, available at http://nobelprize.org / nobel_prizes / chemistry / laureates/ 1972 / anfinsen-lecture.html. (Last accessed 15 October 2009; omit spaces from URL).

Baker, D. (2000) A surprising simplicity to protein folding. *Nature* **405**: 39-42.

Baldwin, R.L. (1995) The nature of protein folding pathways: the classical versus the new view. *Journal of Biomolecular NMR.* **5**(2): 103-9.

Bar, M. (2007) The proactive brain: using analogies and association to generate predictions. *Trends in Cognitive Sciences* **11**(7): 280-289.

Barabási, A.-L. and Oltvai, Z.N. (2004) Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **5**: 101–113.

Bash, P.A., Singh, U.C., Langridge, R. and Kollman, P.A. (1987) Free Energy Calculation by Computer Simulations. *Science* **236**(4801): 564-568.

Beatty, J. (2006) Replaying Life's Tape. *Journal of Philosophy* **103**: 336-362.

Bechtel, W. (2006) *Discovering Cell Mechanisms* (Cambridge: Cambridge University Press).

Bechtel, W. and Richardson, R. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research* (Princeton: Princeton University Press).

Bedau, M.A. (1997) Weak Emergence. *Nous* **31** (Supplement: Philosophical Perspectives, 11): 375-399.

Bedau, M.A. (2008) Is Weak Emergence Just in the Mind? *Minds & Machines* **18**: 443-459.

Bergstrom, C.T. and Rosvall, M. (2009) The transmission sense of information. *Biology and Philosophy*. doi: 10.1007/s10539-009-9180-z.

Berkholz, D.S., Shapovalov, M.V., Dunbrack Jr., R.L. and Karplus, P.A. (2009) Conformation Dependence of Backbone Geometry in Proteins. *Structure* **17**: 1316-1325.

Berman, H. (2008) The Protein Data Bank: a historical perspective. *Acta Crystallographica A* **64**: 88–95.

Blomberg, C. (2006) Fluctuations for good and bad: The role of noise in living systems. *Physics of Life Reviews* **3**: 133-161.

Blow, D.M. (1962) The Molecular Approach to Biology. *Contemporary Physics* **3**: 177-193.

Boogerd, F.C., et al. (2005) Emergence and Its Place in Nature: A Case Study of Biochemical Networks. *Synthese* **145**: 131-164.

Boogerd, F.C., Bruggeman, F.J., Hofmeyr, J.-H.S. and Westerhoff, H.V. (2007) *Systems Biology – Philosophical Foundations* (Amsterdam: Elsevier).

Boorse (1976) Wright on Functions. *The Philosophical Review* **85**(1): 70-86.

Bray, D. (1998) Signaling complexes: Biophysical constraints on intracellular communication. *Annual Reviews of Biophysics and Biomolecular Structure* **27**: 59–75.

Bray, D. (2009) *Wetware: A Computer in Every Living Cell* (New Haven, CT: Yale University Press).

Brewer, S.H., et al. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proceedings of the National Academy of Sciences USA* **102**(46): 16662-16667.

Broad, C.D. (1925) *The Mind and Its Place in Nature* (London: Routledge and Kegan Paul).

Bromberger, S. (1966) Why-Questions. In *University of Pittsburgh Series in the Philosophy of Science* Vol. LLI (Pittsburgh: University of Pittsburgh Press).

Brooks, B. and Karplus, M. (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences USA* **80**(21): 6571-5.

Bruce, V., Green, P.R. and Georgeson, M. (2003) *Visual Perception: Physiology, Psychology and Ecology* (4th Edition) (Hove: Psychology Press).

Buckner, R.L. and Carroll, D.C. (2006) Self-projection and the brain. *Trends in Cognitive Sciences* **11**(2): 49-57.

Bukowski, R., Szalewicz, K., Groenenboom, G.C. and van der Avoird, A. (2007) Predictions of the Properties of Water from First Principles. *Science* **315**: 1249-1252.

Campbell, D.T. (1974) 'Downward causation' in hierarchically organised biological systems. In F. Ayala and T. Dobzhansky (eds.) *Studies in the Philosophy of Biology* (Berkeley: University of California Press).

Canguilhem, G. (1952/2008) *Knowledge of Life* (New York: Fordham University Press).

Carnap, R. (1955) Logical Foundations of the Unity of Science. In O. Neurath, R. Carnap, C. Morris (eds.) *International Encyclopedia of Unified Science, Volume 1* (Chicago: University of Chicago Press), pp.42-62.

Carrier, M. and Finzer, P. (2006) Explanatory Loops and the Limits of Genetic Reductionism. *International Studies in the Philosophy of Science* **20**(3): 267-283.

Cartwright, N. (1983) *How the Laws of Physics Lie* (Oxford: Oxford University Press).

Cartwright, N. (1999) *The Dappled World. A Study of the Boundaries of Science* (Cambridge: Cambridge University Press).

Cartwright, N. (2004) Causation: one word, many things. *Philosophy of Science* **71**: 805-819.

Casti, J.L. (1994) *Complexification* (New York: HarperCollins).

Casti, J.L. (1997) *Would-Be Worlds* (New York: John Wiley & Sons).

Cat, J. (2001) On Understanding: Maxwell on the Methods of Illustration and Scientific Metaphor. *Studies in History and Philosophy of Modern Physics* **32**(3): 395–441.

de Chadarevian, S. (2002) *Designs for Life: Molecular Biology after World War II* (Cambridge: Cambridge University Press).

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).

Chater, N. (1999) The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology* **52A**: 273-302.

Chavez, L.L., Onuchic, J.N. and Clementi, C. (2004) Quantifying the Roughness on the Free Energy Landscape: Entropic Bottlenecks and Protein Folding Rates. *Journal of the American Chemical Society* **126**: 8426-8432.

Chen, Y., et al. (2008) Protein folding: Then and now. *Archives of Biochemistry and Biophysics* **469**: 4-19.

Chiti, F. and Dobson, C.M. (2006) Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry* **75**: 333–66.

Christof, J., Gebhardt, M. and Rief, M. (2009) Force Signaling in Biology. *Science* **324**(5932): 1278-1280.

Churchland, P.M. (1988) *Matter and Consciousness* (revised edition) (Cambridge, MA: MIT Press).

Clark, A. (1996) *Being There – Putting Brain, Body and World Together Again* (Cambridge, MA: MIT Press).

Clark, A.C. (2007) Protein folding: Are we there yet? *Archives of Biochemistry and Biophysics* **469**: 1-3.

Clayton, P. and Davies, P. (2006) *The Re-Emergence of Emergence* (Oxford: Oxford University Press).

Connolly, M.L. (1983) Analytical molecular surface calculation. *Journal of Applied Crystallography* **16**(5): 548–558.

Cornish-Bowden, A. and Cárdenas, M.L. (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. and Cárdenas, M.L. (2007) Organizational Invariance in (M,R)-Systems. *Chemistry & Biodiversity* **4**: 2396-2406.

Cornoldi, C., et al. (1996) *Stretching the Imagination* (Oxford: Oxford University Press).

Craik, K. (1943/1967) *The Nature of Explanation* (revised edition) (Cambridge: Cambridge University Press).

Crampin, E.J., Hackborn, W.W. and Maini, P.K. (2002) Pattern formation in reaction-diffusion models with nonuniform growth. *Bulletin of Mathematical Biology* **64**(4): 747-769.

Craver, C. (2001) When mechanistic models explain. *Synthese* **153**: 355-376.

Crick, F.H.C. (1958) On Protein Synthesis. *The Symposia of the Society for Experimental Biology* **12**: 138-163.

Cruz-Neira, C., Langley, R. and Bash, P.A. (1996) VIBE: A virtual biomolecular environment for interactive molecular modeling. *Computers & Chemistry* **20**(4): 469-475.

Cummins, R. (1975) Functional Analysis. *The Journal of Philosophy* **72**(20): 741-765.

Curien, G., et al. (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Molecular Systems Biology* **5**, article no. 271. DOI: 10.1038/msb.2009.29.

Damasio, A.R. (1994) *Descartes' error: Emotion, reason and the human brain* (New York: Putnam).

Damasio, A.R. (1999) *The feeling of what happens: Body and emotion in the making of consciousness* (New York: Harcourt Brace).

Darley, V. (1994) Emergent Phenomena and Complexity. In R.A Brooks and P. Maes (eds.) *Artificial Life IV – Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (Cambridge, MA: MIT Press), p.411.

Das, R. and Baker, D. (2008) Macromolecular Modeling with Rosetta. *Annual Review of Biochemistry* **77**: 363-382.

David, C.J. and Manley, J.L. (2008) The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes and Development* **22**: 279-285.

Davies, M. and Stone, T. (eds.) (1995) *Folk Psychology: The Theory of Mind Debate* (Oxford: Blackwell Publishers).

Davies, J.F. (2008) *Epistemologies and Ontologies of Genomics: Two Approaches to the Problems of Biology*. Unpublished PhD thesis, University of Exeter.

Dawkins, R. (1976) *The Selfish Gene* (Oxford: Oxford University Press).

Deguet, J., Demazeau, Y. and Magnin, L. (2006) Elements about the Emergence Issue: A Survey of Emergence Definitions. *Complexus* **3**: 24-31.

De Regt, H.W. (2009) Explanation, Understanding and Intelligibility. In H.W. de Regt, S. Leonelli and K. Eigner (eds.) *Scientific Understanding: Philosophical Perspectives* (Pittsburgh, PA: University of Pittsburgh Press).

Dickerson, R.E. and Geis, I. (1969) *The Structure and Action of Proteins* (Menlo Park: W.A. Benjamin, Inc.).

Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nature Structural Biology* **4**(1): 10-19.

Di Paolo, E.A. (2005) Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* **4**(4): 429-452.

Dobson, C.M. (1992) Unfolded proteins, compact states and molten globules. *Current Opinion in Structural Biology* **2**: 6–12.

Dobson, C.M. (2003) Protein folding and misfolding. *Nature* **426**: 884-890.

Dodson, E.J. (2007) Protein predictions. *Nature* **450**: 176-177.

Duhem, P. (1954/1991) *The Aim and Structure of Physical Theory* (paperback edition) (Princeton, NJ: Princeton University Press).

Dupré, J. (1993) *The Disorder of Things* (Cambridge, MA: Harvard University Press).

Dupré, J. (2004) Understanding Contemporary Genomics. *Perspectives on Science* **12**(3): 320-338.

Dupré, J. (2008) *The Constituents of Life* (Assen: Van Gorcum).

Dyer, R.B. (2007) Ultrafast and downhill protein folding. *Current Opinion in Structural Biology* **17**: 38-47.

Elgin, E. (2006) There May Be Strict Empirical Laws in Biology, After All. *Biology and Philosophy* **21**(1): 119-134.

Ellis, R.J. (2001) Macromolecular crowding: Obvious but under-appreciated. *Trends in Biochemical Sciences* **26**(10): 597–604.

Ferguson, E.S. (1992) *Engineering and the Mind's Eye* (Cambridge, MA: MIT Press).

Ferry, G. (1998) *Dorothy Hodgkin – A Life* (London: Granta Publications).

Fields, R.D. (2008) White Matter Matters. *Scientific American*, March 2008 issue.

Fodor, J. (1974) Special Sciences: Or the Disunity of Science as a Working Hypothesis. *Synthese* **28**: 97-115.

Foley, J.D. and Van Dam, A. (1982) *Fundamentals of Interactive Computer Graphics* (Reading, MA: Addison Wesley).

Fontana, W. and Buss, L.W. (1994) What Would be Conserved if 'the Tape were Played Twice'? *Proceedings of the National Academy of Sciences USA* **91**(2): 757-761.

Francoeur, E. and Segal, J. (2004) From Model Kits to Interactive Computer Graphics. In S. de Chadarevian and N. Hopwood (eds.) *Models: The Third Dimension of Science* (Stanford, CA: Stanford University Press), pp. 402-429.

Friedman, M. (1974) Explanation and Scientific Understanding. *The Journal of Philosophy* **71**(1): 5-19.

Frigg, R. and Reiss, J. (2008) The philosophy of simulation: hot new issues or same old stew? *Synthese* **169**(3): 593-613.

Frith, C. (2007) *Making Up the Mind: How the Brain Creates Our Mental World* (Oxford: Blackwell Publishing).

Fromm, J. (2005) Types and Forms of Emergence. Available at http://arxiv.org/ftp/nlin/papers/0506/050628.pdf (last accessed 5 November 2009).

Gallagher, S. (2007) Simulation trouble. *Social Neuroscience* **2**(3-4): 353-365.

Gallese, V. and Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* **2**: 493-501.

Galton, F. (1883) *Inquiries into Human Faculty and Its Development* (Macmillan).

Ganti, T. (2003) *The Principles of Life* (Oxford: Oxford University Press).

Gardner, M. (1970) The fantastic combinations of John Conway's new solitaire game 'life'. *Scientific American* **223**(4) (October 1970): 120-123.

Giaquinto, M. (2007) *Visual Thinking in Mathematics* (Oxford: Oxford University Press).

Giere, R.N. (1988) *Explaining Science: A Cognitive Approach* (Chicago: University of Chicago Press).

Giere, R.N. (2004) How Models Are Used to Represent Reality. *Philosophy of Science* **71**: 742-752.

Gijsbers, V. (2007) Why Unification Is Neither Necessary Nor Sufficient for Explanation. *Philosophy of Science* **74**(4): 481-500.

Glennan (1996) Mechanisms and the nature of causation. *Erkenntnis* **44**: 49-71.

Glennan (2002) Rethinking Mechanistic Explanation. *Philosophy of Science* **69**: S342-S353.

Godfrey-Smith, P. (1994*) Complexity and the Function of Mind in Nature* (Cambridge: Cambridge University Press).

Godfrey-Smith, P. (2003*) Theory and Reality – An Introduction to the Philosophy of Science* (Chicago: The University of Chicago Press).

Golding, I. and Cox, E.C. (2006) Physical nature of bacterial cytoplasm. *Physical Review Letters* **96**: 098102-1–098102-4.

Goldman, A.I. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (New York: Oxford University Press).

Gorter, E. and Grendel, F. (1925) On bimolecular layers of lipoids on the chromocytes of the blood. *The Journal of Experimental Medicine* **41**: 439-443.

Gould, S.J. (1989) *Wonderful Life: The Burgess Shale and the Nature of History* (New York: Norton).

Gregory, R. (1981) *Mind in Science* (Harmondsworth: Penguin).

Gregory, R.L. (1997) *Eye and Brain: The Psychology of Seeing* (5th Edition) (Oxford: Oxford University Press).

Griesemer, J. (2004) 3-D Models in Philosophical Perspective. In S. de Chadarevian and N. Hopwood (eds.) *Models: the Third Dimension of Science* (Stanford, CA: Stanford University Press), pp.433-441.

Griffiths, P.E. (2001) Genetic information: a metaphor in search of a theory. *Philosophy of Science* **68**: 394-412.

Griffiths, P.E. and Gray, R.D. (2005) Discussion: Three ways to misunderstand developmental systems theory. *Biology and Philosophy* **20**: 417–425.

Griffiths, P.E. and Knight, R.D. (1998) What is the Developmentalist Challenge? *Philosophy of Science* **65**(2): 253-258.

Griffiths, P.E. and Neumann-Held, E. (1999) The many faces of the gene. *BioScience* **49**(8): 656-662.

Hacking, I. (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge: Cambridge University Press).

Hadamard, J. (1945) *The Psychology of Invention in the Mathematical Field* (Princeton, NJ: Princeton University Press).

Hall, D. and Minton, A.P. (2003) Macromolecular crowding: Qualitative and semiquantitative successes, quantitative challenges. *Biochimica et Biophysica Acta* **1649**: 127–139.

Harold, D., et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics* **41**(10): 1088-1093.

Harold, F.M. (2005) Molecules into Cells: Specifying Spatial Architecture. *Microbiology and Molecular Biology Reviews* **69**(4): 544-564.

Harris, H. (1999) *The Birth of the Cell* (New Haven, CT: Yale University Press).

Hartwell, L.H. and Weinert, T.A. (1989) Checkpoints: Controls That Ensure the Order of Cell Cycle Events. *Science* **246**: 629-634.

Haynes, J.-D. and Rees, G. (2005) Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Current Biology* **15**: 1301-1307.

Hazen R.M., Griffin P.L., Carothers J.M. and Szostak J.W. (2007) Functional information and the emergence of biocomplexity. *Proceedings of the National Academy of Sciences USA* **104** Suppl 1: 8574-81.

Hegarty, M. (2004) Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences.* **8**(6): 280-285.

Helles, G. (2008) A comparative study of the reported performance of ab initio protein structure prediction algorithms. *Journal of the Royal Society Interface* **5**: 387-396.

Hempel, C.G. (1965/1970) *Aspects of Scientific Explanation* (paperback edition) (New York: Free Press).

Hempel, C.G. and Oppenheim, P. (1948) Studies in the Logic of Explanation. *Philosophy of Science* **15**: 135-75. [Reproduced in C.G. Hempel (1965) *Aspects of Scientific Explanation* (New York: Free Press).]

Higgins, D. and Taylor, W. (2000) *Bioinformatics: Sequence, structure, and databanks* (Oxford: Oxford University Press).

Ho, B.K., Thomas, A. and Brasseur, R. (2003) Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α-helix. *Protein Science* **12**: 2508-2522.

Ho, B.K. and Dill, K.A. (2006) Folding Very Short Peptides Using Molecular Dynamics. *PLoS Computational Biology* **2**(4)(e27): 0228-0237.

Hodges, A. (1983/1992) *Alan Turing: The Enigma* (paperback edition) (London: Vintage).

Holland, J.H. (1998) *Emergence* (Oxford: Oxford University Press).

Hume, D. (1739/1978) *A Treatise of Human Nature* (edited by L.A. Selby-Bigge, revised by P.H. Nidditch) (Oxford: Oxford University Press).

Huneman, P. (2008) Emergence Made Ontological? Computational versus Combinatorial Approaches. *Philosophy of Science* **75**: 595-607.

Hytönen, V.P. and Vogel, V. (2008) How Force Might Activate Talin's Vinculin Binding Sites: SMD Reveals a Structural Mechanism. *PLoS Computational Biology* **4**(2): e24.

Ingber, D.E. (2003a) Tensegrity I. Cell structure and hierarchical systems biology. *Journal of Cell Science* **116**: 1157-1173.

Ingber, D.E. (2003b) Tensegrity II. How structural processing networks influence cellular information processing networks. *Journal of Cell Science* **116**: 1397-1408.

Ishizaki, A. and Fleming, G.R. (2009) Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature. *Proceedings of the National Academy of Sciences USA* **106**(41): 17255–17260.

Jayachandran, G., et al. (2007) Local structure formation in simulations of two small proteins. *Journal of Structural Biology* **157**: 491-499.

Jeffery, C.J. (1999) Moonlighting proteins. *Trends in Biochemical Sciences* **4**(1): 8-11.

Johnson, S. (2002) *Emergence* (paperback edition) (London: Penguin).

Johnson-Laird, P.N. (1983) *Mental Models* (Cambridge: Cambridge University Press).

Jones, N.S. (2008) Using the Memories of Multiscale Machines to Characterize Complex Systems. *Physical Review Letters* **100**: 208702.

Jones, T.A. (1978) A graphics model building and refinement system for macromolecules. *Journal of Applied Crystallography* **11**: 268–272.

Judson, H.F. (1996) *The Eighth Day of Creation* (2nd edition) (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Kant, I. (1790/2007) *Critique of Judgement* (Oxford World Classics Edition) (Oxford: Oxford University Press).

Karplus, M. (1997) The Levinthal paradox: yesterday and today. *Folding & Design* **2** (supplement): S72.

Karplus, M. and McCammon, J.A. (2002) Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9**(9): 646-652.

Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (New York: Oxford University Press).

Kauffman, S.A. (1996) *At Home in the Universe: The Search for Laws of Self-Organization and Complexity* (London: Penguin Books).

Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* **14**: 1-63.

Kay, K.N., Naselaris, T., Prenger, R.J. and Gallant, J.L. (2008) Identifying natural images from human brain activity. *Nature* **452**: 352-355.

Kay, L.E. (1993) *The Molecular Vision of Life: Caltech, The Rockefeller Foundation, and the Rise of the New Biology* (New York: Oxford University Press).

Kay. L.E. (1997) Cybernetics, Information, Life: The Emergence of Scriptural Representations of Heredity. *Configurations* **5**(1): 23-91.

Kay, L.E. (2000) *Who Wrote the Book of Life? A History of the Genetic Code* (Stanford, CA: Stanford University Press).

Keil, F.C. (2003) Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences* **7**(8): 368-373.

Kelemen, D. (1999) The scope of teleological thinking in preschool children. *Cognition* **70**: 241-272.

Kelemen, D. and Rosset, E. (2009) The Human Function Compunction: Teleological explanation in adults. *Cognition* **111**: 138-143.

Keller, E.F. (1990) Physics and the Emergence of Molecular Biology: A History of Cognitive and Political Synergy. *Journal of the History of Biology* **23**(3): 389-409.

Kemeny, J.G. and P. Oppenheim (1956) On reduction. *Philosophical Studies* **7**: 6-19.

Kilner, J.M., et al. (2009) Evidence of Mirror Neurons in Human Inferior Frontal Gyrus. *The Journal of Neuroscience* **29**(32): 10153–1015.

Kim, J. (1999) Making Sense of Emergence. *Philosophical Studies* **95**: 3-36.

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

Kitcher, P. (1989) Explanatory Unification and the Causal Structure of the World. In P. Kitcher and W.C. Salmon (eds.) *Scientific Explanation* (Minnesota Studies in the Philosophy of Science, Volume 8) (Minneapolis, MN: University of Minnesota Press), pp.410–505.

Kleywegt and Jones (1996) Phi/Psi-chology: Ramachandran revisited. *Structure* **4**(12): 1395-1400.

Kosslyn, S. (2001) Neural Foundations of Imagery. *Nature Reviews Neuroscience* **2**: 635-642.

Krivov, S.V. and Karplus, M. (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. *Proceedings of the National Academy of Sciences USA* **101**(41): 14766-14770.

Kveraga, K., Ghuman, A.S. and Bar, M. (2007) Top-down predictions in the cognitive brain. *Brain and Cognition* **65**: 145-168.

Lambert, J.C., et al. (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature Genetics* **41**(10): 1094-1099.

Leach, A.R. (2001) *Molecular Modelling: Principles and Applications* (2nd Edition) (Prentice-Hall).

Lei, H. and Duan, Y. (2007) Two-stage Folding of HP-35 from *Ab Initio* Simulations. *Journal of Molecular Biology* **370**: 196-206.

Lelandais, G., et al. (2006) Comparing gene expression networks in a multi-dimensional space to extract similarities and differences between organisms. *Bioinformatics* **22**(11): 1359-1366.

Letelier, J.C., Marín, G. and Mpodozis, J. (2003) Autopoietic and (M, R) systems. *Journal of Theoretical Biology* **222**: 261-272.

Letelier, J.C., et al. (2006) Organizational invariance and metabolic closure: Analysis in terms of (M, R) systems. *Journal of Theoretical Biology* **238**(4): 949-961.

Levinthal, C. (1966) Molecular Model-Building by Computer. *Scientific American* **214**: 42-52.

Levinthal, C. (1969) How to fold graciously. In P. Debrunner, J.C.M. Tsibris and E. Münck, (eds.) *Mössbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House, Monticello, Illinois* (Urbana, IL: University of Illinois Press), pp.22-24.

Levy, A. (2007) Biological Information as an Explanatory Metaphor. Talk presented at ISHPSSB meeting, Exeter, UK, July 2007.

Lewens, T. (2000) Function Talk and the Artefact Model. *Studies in History and Philosophy of Biological and Biomedical Sciences* **31**(1): 95-111.

Lewens, T. (2005) *Organisms and Artifacts: Design in Nature and Elsewhere* (Cambridge, MA: MIT Press).

Lewontin, R. (1992) The Dream of the Human Genome. *New York Review of Books* **39**: 31-40.

Li, P., et al. (2009) Dynamics of one-state downhill protein folding. *Proceedings of the National Academy of Sciences USA* **106**(1): 103–108.

Lingnau, A., Gesierich, B. and Caramazza, A. (2009) Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences USA* **106**(24): 9925-9930.

Lipton, P. (2004) *Inference to the Best Explanation* (2nd Edition) (London: Routledge).

Lodish, H., et al. (1999) *Molecular Cell Biology* (4th edition) (New York: Freeman).

Lombrozo, T. and Carey, S. (2006) Functional explanation and the function of explanation. *Cognition* **99**: 167-204.

Lovell, S.C. and Papp, B. (2005) Bioinformatics: from molecules to systems. A Discussion Meeting held at The Royal Society on 4 and 5 April 2005. *Journal of the Royal Society Interface* **2**: 397-400.

McCammon, J.A. and Harvey, S.C. (1989) *The Dynamics of Proteins and Nucleic Acids* (Cambridge: Cambridge University Press).

McDowell, J. (1996) *Mind and World* (paperback edition) (Cambridge, MA: Harvard University Press).

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

MacKay, R.S. (2008) Nonlinearity in complexity science. *Nonlinearity* **21**(12): T273-T281.

Mackie, J.L. (1980) *The Cement of the Universe* (paperback edition) (Oxford: Oxford University Press).

Mainzer, K. (2007) *Thinking in Complexity* (5th Edition) (Berlin: Springer).

Manneville, P., Boccara, N., Vichniac, G.Y. and Bidaux, R. (1989) *Cellular Automata and Modeling of Complex Physical Systems* (Springer Proceedings in Physics 46) (Berlin: Springer-Verlag).

Marr, D. (1982) *Vision* (New York: W.H. Freeman).

Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989) Modelling antibody hypervariable loops: a combined algorithm. *Proceedings of the National Academy of Sciences USA* **86**: 9268-9272.

Matagne, A. and Dobson, C.M. (1998) The folding process of hen lysozyme: a perspective from the 'new view'. *Cellular and Molecular Life Sciences* **54**: 363-371.

Maturana, H.R. and Varela, F.J. (1980) *Autopoiesis and Cognition* (Boston Studies in the Philosophy of Science, Volume 42) (Dordrecht: Reidel).

Mayer, B.J., Blinov, M.L. and Loew, L.M. (2009) Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *Journal of Biology* **8**: 81.1-81.8.

Maynard Smith, J. (2000) The concept of information in biology. *Philosophy of Science* **67**: 177-194.

Mayr, E. (1961) Cause and Effect in Biology. *Science* **134**(3489): 1501-1506.

Michie, K.A. and Löwe, J. (2006) Dynamic Filaments of the Bacterial Cytoskeleton. *Annual Review of Biochemistry* **75**: 467-492.

Miller, G.A. (1956) The Magical Number Seven, Plus or Minus Two. *Psychological Review* **63**: 81-97.

Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories* (Cambridge, MA: MIT Press).

Minsky, M. (1987) *The Society of Mind* (London: Picador).

Mok, K.H., et al. (2007) A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* **447**: 106-109.

Morange, M. (2000) *A History of Molecular Biology* (paperback edition) (Cambridge, MA: Harvard University Press).

Morange, M. (2001) *The Misunderstood Gene* (Cambridge, MA: Harvard University Press).

Morange, M. (2006a) The Ambiguous Place of Structure Biology in the Historiography of Molecular Biology. In H. Rheinberger and S. Chadarevian (eds.) *History and Epistemology of Molecular Biology and Beyond* (Berlin: Max Planck Institute for the History of Science), pp.179-186.

Morange, M. (2006b) Post-genomics, between reduction and emergence. *Synthese* **151**: 355-360.

Morgan, D.O. (2007) *The Cell Cycle – Principles of Control* (London: New Science Press).

Morgan, G.J. (2009) Laws of Biological Design: A Reply to John Beatty. *Biology and Philosophy* 10.1007/s10539-009-9181-y.

Morgan, M. and Morrison, M. (1999) *Models as Mediators: Perspectives on Natural and Social Science* (Cambridge: Cambridge University Press).

Morowitz, H.J. (2002) *The Emergence of Everything – How the World Became Complex* (Oxford: Oxford University Press).

Morrison, M. (2007) Where Have All the Theories Gone? *Philosophy of Science* **74**(2): 195-228.

Morton, A. (1980) *Frames of Mind* (Oxford: Oxford University Press).

Moss, L. (2003) *What Genes Can't Do* (Cambridge, MA: MIT Press).

Moss, L. (2006) The Question of Questions: What is a gene? Comments on Rolston and Griffiths & Stotz. *Theoretical Medicine and Bioethics* **27**: 523–534.

Nagel, E. (1961/1979) *The Structure of Science – Problems in the the Logic of Scientific Explanation* (Indianapolis, IN: Hackett).

Neuweiler, H., Doose, S. and Sauer, M. (2005) A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proceedings of the National Academy of Sciences USA* **102**(46): 16650-16655.

Nichols, S. (ed.) (2006) *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction* (Oxford: Oxford University Press).

Nichols, S. and Stich, S.P. (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford: Oxford University Press).

Nilsen, T.W. (2003) The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* **25**: 1147–1149.

Noble, D. (2006) *The Music of Life: Biology beyond the Genome* (Oxford: Oxford University Press).

Noë, A. (2004) *Action in Perception* (Cambridge, MA: MIT Press).

Norris, V., et al. (2004) Questions for cell cyclists. *Journal of Biological Physics and Chemistry* **4**: 124–130.

Norris, V., Cabin, A. and Zemirline, A. (2005) Hypercomplexity. *Acta Biotheoretica* **53**: 313-330.

Norris, V., et al. (2007) Toward a Hyperstructure Taxonomy. *Annual Review of Microbiology* **61**: 309-329.

Nozick, R. (1981) *Philosophical Explanations* (Oxford: Oxford University Press).

Nurse, P. (2000) A long twentieth century of the cell cycle and beyond. *Cell* **100**: 71–78.

Nurse, P. (2008) Life, logic and information. *Nature* **454**: 424-426.

Ohgushi, M. and Wada, A. (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Letters* **164**(1): 21–24.

Olby, R. (1974/1994) *The Path to the Double Helix* (Dover Edition) (New York: Dover Publications).

O'Malley, M.A. and Dupré, J. (2005) Fundamental issues in systems biology. *BioEssays* **27**: 1270-76.

Oppenheim, P. and Putnam, H. (1958) Unity of Science as a Working Hypothesis. *Minnesota Studies in the Philosophy of Science* **2**: 3-36.

Oxender, D.L. and Fox, C.F. (1987) *Protein Engineering* (New York: Alan R. Liss).

Oyama, S., Griffiths, P.E. and Gray, R.D. (2001) *Cycles of Contingency: Developmental Systems and Evolution* (Cambridge, MA: MIT Press).

Pais, A. (1986) *Inward Bound – Of Matter and Forces in the Physical World* (Oxford: Oxford University Press).

Pask, G. (1961) *An Approach to Cybernetics* (London: Hutchinson).

Peacocke, A.R. (1989) *An Introduction to the Physical Chemistry of Biological Organization* (paperback edition) (Oxford: Oxford University Press).

Peitgen, H.-O., Juergens, H. and Saupe, D. (1992) *Chaos and Fractals: New Frontiers of Science* (Berlin: Springer-Verlag).

Pezzulo, G. (2008) Anticipation and Anticipatory Systems: An Introduction. Available at: http://www.istc.cnr.it/doc/1a_0000b_20080724d_anticipation.pdf (last accessed 31 October 2009).

Phillips, D.C. (1981) Concluding remarks. In R.H. Sarma (ed.) *Biomolecular Stereodynamics II*, (Guilderland, NY: Adenine Press), pp.497-498.

Platt, J.R. (1960) The Need for Better Macromolecular Models. *Science* **131**(3409): 1309-1310.

Powell, A. (1989) Thinking About Protein Folding. Unpublished manuscript.

Powell, A. and Dupré, J. (2009) From Molecules to Systems: The Importance of Looking Both Ways. *Studies in History and Philosophy of Biological and Biomedical Sciences* **40**(1): 54-64.

Psillos, S. (2002) *Causation and Explanation* (Chesham: Acumen).

Pylyshyn, Z. (2003) Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences* **7**(3): 113-118.

Ramakrishnan, C. and Ramachandran, G.N. (1965) Stereochemical Criteria for Polypeptide and Protein Chain Conformations. II. Allowed Conformations for a Pair of Peptide Units. *Biophysical Journal* **5**: 909-933.

Raman, S., et al. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function, Bioinformatics* **77**(Suppl. 9): 89-99.

Ratcliffe, M. (2000) The Function of Function. *Studies in History and Philosophy of Biological and Biomedical Sciences* **31**(1): 113-133.

Reisberg, D., Pearson, D.G. and Kosslyn, S.M. (2003) Intuitions and Introspections about Imagery: The Role of Imagery Experience in Shaping an Investigator's Theoretical Views. *Applied Cognitive Psychology* **17**: 147-160.

Reiss, J. (2007) Causation: An Opinionated Introduction. Available at http://jreiss.org/Causality%20Manuscript.pdf (last accessed 21 December 2009).

Resnick, M. (1996) Beyond the Centralized Mindset. *The Journal of the Learning Sciences* **5**(1): 1-22.

Reynolds, A. (2007) The cell's journey: from metaphorical to literal factory. *Endeavour* **31**(2): 65-70.

del Rio, A., et al. (2009) Stretching Single Talin Rod Molecules Activates Vinculin Binding. *Science* **323**(5914): 638-641.

Ronald, E.M.A. and Sipper, M. (2001) Surprise versus unsurprise: Implications of emergence in robotics. *Robotics and Autonomous Systems* **37**: 19-24.

Rose, G.D., Fleming, P.J., Banavar, J.R. and Maritan, A. (2006) A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences USA* **103**(45): 16623-16633.

Rosen, R. (1985) *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations* (Oxford: Pergamon Press).

Rosen, R. (1991) *Life Itself* (New York: Columbia University Press).

Rosen, R. and Kineman, J.J. (2005) Anticipatory Systems and Time: A New Look at Rosennean Complexity. *Systems Research and Behavioural Science* **22**: 399-412.

Rosenberg, A. (1985) *The Structure of Biological Science* (Cambridge/New York: Cambridge University Press.

Rosenberg, A. (2001) How is Biological Explanation Possible. *British Journal for the Philosophy of Science* **52**: 735-760.

Rosset, E. (2008) It's no accident: Our bias for intentional explanations. *Cognition* **108**: 771-780.

Ruse, M. (2000) Teleology: Yesterday, Today, and Tomorrow? *Studies in History and Philosophy of Biological and Biomedical Sciences* **31**(1): 213-232.

Ruse, M. (2005) Darwinism and mechanism: metaphor in science. *Studies in History and Philosophy of Biological and Biomedical Sciences* **36**: 285-302.

Ryan, A. (2007) Emergence is coupled to scope, not level. *Complexity* **13**(2): 67-77.

Salmon, W.C. (1989) *Four Decades of Scientific Explanation* (Pittsburgh, PA: University of Pittsburgh Press).

Salmon, W.C. (1998) *Causality and Explanation* (New York: Oxford University Press).

Sanger, F. and Thompson, E.O.P. (1953) The Amino-acid Sequence in the Glycyl Chain of Insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochemical Journal* **53**: 366-374.

Sarkar, S. (1998) *Genetics and Reductionism* (Cambridge: Cambridge University Press).

Schaffner, K.F. (1967) Approaches to Reduction. *Philosophy of Science* **34**: 137-147.

Schnell, S. and Turner, T.E. (2004) Reaction kinetics in intracellular environments with macromolecular crowding: Simulations and rate laws. *Progress in Biophysics and Molecular Biology* **85**: 235–260.

Schrödinger, E. (1944) *What is Life?* (Cambridge: Cambridge University Press).

Schubotz, R.I. (2007) Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences* **11**(5): 211-218.

Schulz, G.E and Schirmer, R.H. (1979) *Principles of Protein Structure* (New York: Springer-Verlag).

Scriven, M. (1962) Explanations, Predictions, and Laws. In H. Feigl and G. Maxwell (eds.) *Scientific Explanation, Space, and Time* (Minnesota Studies in the Philosophy of Science, Volume 3) (Minneapolis, MN: University of Minnesota Press), pp.170-230.

Searle, J. (1992) *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).

Sellars, W. (1963/1991) *Science, Perception and Reality* (Atascadero, CA: Ridgeview Publishing Company).

Service, R.F. (2008) Problem Solved* (*sort of). *Science* **321**: 784-786.

Shannon, C.E and Weaver, W. (1949/1998) *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press).

Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* **171**: 701-3.

Silberstein, M. and McGeever, J. (1999) The Search for Ontological Emergence. *The Philosophical Quarterly* **49**: 182-200.

Simon, H.A. (1996) *The Sciences of the Artificial* (3rd Edition) (Cambridge, MA: MIT Press).

Singer, S.J. and Nicolson, G.L. (1972) The fluid mosaic model of the structure of cell membranes. *Science* **175**: 720-731.

Sloep, P. and Van der Steen, W. (1991) Philosophy of Biology, Faithful or Useful? *Biology and Philosophy* **6**: 93-98.

Sloman, S. (2005) *Causal Models: How People Think About the World and Its Alternatives* (New York: Oxford University Press).

Smith, L.D., et al. (2000) Scientific Graphs and the Hierarchy of the Sciences: A Latourian Survey of Inscription Practices. *Social Studies of Science* **30**(1): 73-94.

Smith, L.J., Mark, A.E., Dobson, C.M. and van Gunsteren, W.F. (1998) Molecular Dynamics Simulations of Peptide Fragments from Hen Lysozyme: Insight into Non-native Protein Conformations. *Journal of Molecular Biology* **280**: 703-719.

Smith, M.L., et al. (2007) Force-Induced Unfolding of Fibronectin in the Extracellular Matrix of Living Cells. *PLoS Biology* **5**(10): 2243-2254.

Sommer, M. (2008) History in the Gene: Negotiations Between Molecular and Organismal Anthropology. *Journal of the History of Biology* **41**(3):473-528.

Spörlein, S., et al. (2002) Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proceedings of the National Academy of Sciences USA* **99**(12): 7998-8002.

Stenning, K. (2002) *Seeing Reason* (Oxford: Oxford University Press).

Stephan, A. (1999) Varieties of Emergentism. *Evolution and Cognition* **5**(1): 49-59.

Stonier, T. (1990) *Information and the Internal Structure of the Universe* (London: Springer-Verlag).

Stotz, K., Griffiths, P.E. and Knight, R. (2004) How biologists conceptualize genes: an empirical study. *Studies in History and Philosophy of Biological and Biomedical Sciences* **35**: 647-673.

Stretton, A.O.W. (2002) The First Sequence: Fred Sanger and Insulin. *Genetics* **162**: 527-532.

Strohman, R.C. (1994) Epigenesis: The missing beat in biotechnology. *Bio/Technology* **12**: 156–164.

Suppe, F. (ed.) (1974) *The Structure of Scientific Theories* (Urbana, IL: University of Illinois Press).

Suppe, F. (2000) Understanding Scientific Theories: An Assessment of Developments, 1969-1998. *Philosophy of Science* **67** (Proceedings): S102-S115.

Šustar, P. (2007) Crick's Notion of Genetic Information and the 'Central Dogma' of Molecular Biology. *British Journal of Philosophy of Science* **58**: 13-24.

Sutherland, I.E. (1963) SketchPad: A man-machine graphical communication system. *American Federation of Information Processing Societies Conference Proceedings* **23**: 323-328.

Tanford, C. (1978) The hydrophobic effect and the organization of living matter. *Science* **200**(4345): 1012-1018.

Thomas, K. (1971) *Religion and the Decline of Magic* (London: Weidenfeld and Nicholson).

Thompson, P. (1989) *The Structure of Biological Theories* (Albany, NY: SUNY Press).

Toettcher, J.E., et al. (2009) Distinct mechanisms act in concert to mediate cell cycle arrest. *Proceedings of the National Academy of Sciences USA* **106**(3): 785-790.

Torres, P.J. (2009) A Modified Conception of Mechanisms. *Erkenntnis* **71**: 233-251.

Tramontano, A. (2006) *Protein Structure Prediction – Concepts and Applications* (Weinheim: Wiley-VCH).

Trout, J.D. (2002) Scientific Explanation and the Sense of Understanding. *Philosophy of Science* **69**: 212-233.

Turing, A.M. (1952) The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London* **B 237**: 37-72.

Van Gulick, R. (2007) Reduction, Emergence, and the Mind/Body Problem: A Philosophic Overview. In P. Clayton and P. Davies (eds.) *The Re-Emergence of Emergence* (Oxford: Oxford University Press).

Van Regenmortel, M.H.V. (2004) Reductionism and complexity in molecular biology. *EMBO Reports* **5**(11): 1016–1020.

de Vega, M., et al. (1996) *Models of Visuospatial Cognition* (Oxford: Oxford University Press).

Wang, W., Donini, O., Reyes, C.M. and Kollman, P.A. (2001) Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. *Annual Review of Biophysics and Biomolecular Structure* **30**: 211-243.

Warwicker, J., Ollis, D., Richards, F.M., Steitz, T.A. (1985) Electrostatic field of the large fragment of Escherichia coli DNA polymerase I. *Journal of Molecular Biology* **186**(3): 645-649.

Waters, C.K. (2007) Causes that make a difference. *Journal of Philosphy* **104**: 551–579.

Watson, J.D. and Crick, F.H.C. (1953) A Structure for Deoxyribose Nucleic Acid. *Nature* **171**: 737-738.

Watts, D.J. (2003) *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton, NJ: Princeton University Press).

Weiss, K.M. (2005) The phenogenetic logic of life. *Nature Reviews Genetics* **6**: 36–46.

Weiss, M. (2003) Stabilizing Turing patterns with subdiffusion in systems with low particle numbers. *Physical Review E* **68**: 036213-1–036213-5.

Weiss, M., Elsner, M., Kartberg, F. and Nilsson, T. (2004) Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophysical Journal* **87**: 3518–3524.

White, P.A. (1998) The dissipation effect: A general tendency in causal judgments about complex physical systems. *American Journal of Psychology* **111**(3): 379-410.

Whitford, D. (2005) *Proteins: Structure and Function* (Chichester: Wiley).

Wiener, N. (1948/1961) *Cybernetics: Or Control and Communication in the Animal and the Machine* (2nd revised edition) (Cambridge, MA: MIT Press).

Wilkins, M. (2003) *The Third Man of the Double Helix* (Oxford: Oxford University Press).

Wimsatt, W.C. (2007) *Re-Engineering Philosophy for Limited Beings* (Cambridge, MA: Harvard University Press).

Winter, G., et al. (1982) Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **299**: 756-758.

Wolpert, L. and Richards, A. (1988) *A Passion for Science* (Oxford: Oxford University Press).

Wolynes, P.G. (2004) Latest folding game results: Protein A barely frustrates computationalists. *Proceedings of the National Academy of Sciences USA* **101**(18): 6837-6838.

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

Woodward, J. (2003b) *Scientific Explanation*. Stanford Encyclopedia of Philosophy.

Wouters, A. (2003) Four notions of biological function. *Studies in History and Philosophy of Biological and Biomedical Sciences* **34**: 633-668.

Wouters, A. (2005) The Function Debate in Philosophy. *Acta Biotheoretica* **53**: 123-151.

Wright, L. (1973) Functions. *The Philosophical Review* **82**(2): 139-168.

Wüthrich, K. (1995) *NMR in Structural Biology* (Singapore: World Scientific).

Ybe, J.A. and Hecht, M.H. (1996) Sequence replacements in the central β-turn of plastocyanin. *Protein Science* **5**(5): 814-824.

Yourgrau, P. (2007) *A World Without Time: The Forgotten Legacy of Gödel and Einstein* (London: Penguin Books).

Zeki, S. (1993) *A Vision of the Brain* (Oxford: Blackwell Scientific Publications).

Zhang, Y. (2009) Protein structure prediction: when is it useful? *Current Opinion in Structural Biology* **19**: 1-11.

Zolkiewski, M. (2006) A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Molecular Microbiology* **61**(5): 1094-1100.

Zuckerkandl, E. and Pauling, L. (1965) Molecules as Documents of Evolutionary History. *Journal of Theoretical Biology* **8**: 357-366.