Biological mechanisms: a case study in conceptual plasticity[1]

Alexander Powell
(ap.cybercraft@googlemail.com)

Introduction

In recent years the topic of mechanistic explanation has attracted considerable philosophical interest.[2] Works by Glennan (1996, 2002) and by Machamer, Darden and Craver (2000) in particular have spawned a proliferating mechanism literature that makes contact with debates about many of the subjects of greatest concern to philosophers of science and philosophers of biology.[3] These subjects include (besides explanation) causation, reduction, and function, as well as laws, theories and models. A specifically biological debate centres on whether evolution is to be understood in mechanistic terms (Skipper and Millstein 2005).

Two main factors account for the amount of philosophical attention devoted to the topic of mechanism. One is the failure, despite strenuous efforts, to give a philosophically satisfactory account of scientific explanation in terms of such concepts as laws and theories. In biology, where formal laws are in short supply, this failure is felt especially keenly. Secondly there is the obvious fact that mechanism talk is a common feature of biological explanations in particular and descriptions of scientific phenomena more generally, as well as being widespread throughout broader culture. Across a huge range of areas it is easy to find references to the mechanism by which X happens or to mechanisms for doing Y. Chemists talk about reaction mechanisms; physicists are liable to describe in mechanistic terms processes ranging in scale from the sub-atomic to the cosmic; cell biologists talk about cell signalling mechanisms, about the mechanism of protein synthesis, and about the mechanisms underpinning the immune response (to give but three examples); a zoologist might discuss camouflage mechanisms; biologists and psychologists investigate learning mechanisms and mechanisms of communication between individuals of a particular species; the list can be extended apparently without end.

Against this background it is natural to ask what mechanisms are, and to seek to establish their role in explanation. Are they objectively real, or should they be thought of as explanatory constructs we use to interpret and understand phenomena? Or are they perhaps one of the forms that understanding can take? Can we give an account of the concept that makes sense of its application across contexts that exhibit considerable ontological diversity? What is the basis

[1] Parts of this chapter first appeared in Powell 2009.
[2] The topic is hardly a new one, however. Gregory (1981) devotes a chapter to its significance in science and Mackie (1974) makes the connection between mechanism and causation from a more detached philosophical standpoint. The mechanistic tenor of much modern bioscience is prefigured in, for example, Monod's contention that life is machine-like at the molecular level (Monod 1971), and in Crick's ready adoption of mechanistic language (see, for example, his references to mechanism quoted in Olby 2009). More directly connected with the recent mechanism literature is Bechtel and Richardson 1993.
[3] See, for example: Bechtel 2006; Bechtel and Craver 2006; Bechtel and Abrahamsen 2005; Craver 2001; McKay Illari and Williamson 2010; Psillos 2004. Glennan (2008) provides a useful introduction to recent mechanism debates.

for a link between mechanism and explanation if, to exaggerate a little, almost anything can be regarded as a mechanism? [4]

Here I consider some of the principal contemporary philosophical perspectives. My stance in relation to these is somewhat equivocal. For example, the most influential account, that of Machamer, Darden and Craver (2000), suffers from problems of scope and reference. These stem perhaps from the arguably reductionist (but hardly unreasonable) desire for an account that subsumes much by way of little. On the other hand, however, the account does suggest possibilities for tightening our philosophical grasp on explanation and understanding. I discuss neo-mechanist philosophical perspectives in relation to several more or less commonsense ideas about mechanism, and whilst bearing in mind the diversity of molecular and cell biological phenomena. Towards the end of the chapter I explore further the connections between mechanism talk and psychological issues suggested by the account of Machamer, Darden and Craver. I argue that when we describe phenomena in the world mechanistically we inadvertently reveal the existence of cognitive entailment structures which mean that our thoughts are capable of paralleling, at least approximately, those external phenomena. The entailment structures in question are protean in nature, arising as they do from complex and interwoven conceptual, linguistic and possibly even emotional associations, and from diverse capacities to imagine, represent and reason. In large part this is why ontologically spare philosophical accounts struggle to capture the richness and variety of the mechanism talk we see in biology, in science generally, and in culture more widely.

Neo-mechanist accounts: vices and virtues

MDC: entities and activities

The most prominent contemporary philosophical account of mechanism is that of Machamer, Darden and Craver (2000). (Henceforth I shall use 'MDC' to refer to the paper, its authors and the ideas it articulates; context will indicate the intended meaning well enough.) The account defines mechanisms as '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 discussed from molecular biology and neurobiology is the case of DNA replication:

> 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)

[4] Confining ourselves to material processes and phenomena it is surprisingly unclear how great is the degree of exaggeration.

Distinctive features of the MDC account include (1) its dual ontology of entities and activities; (2) its apparent substitution of the notion of productivity for explicit reference to the concept of causation; (3) its de-emphasis of functional language; and (4) the importance attached to the idea of regularity. In addition we can join Psillos in noting that its authors strive for an account that is in tune with scientific practice (Psillos 2004, p.294).

Regarding the dual ontology of entities and activities, entities at least seem unproblematic. They are '[t]hings that engage in activities' (MDC, p.3), and presumably can be equated with material structures or parts of structures, such as macromolecules and macromolecular complexes. Activities are less straightforwardly understood in any ontologically weighty sense, however. MDC describe them as 'the producers of change', and note that '[a]ctivities usually require that entities have specific types of properties' (ibid.). The aim of MDC in emphasizing activities in their account is to reflect the importance, highlighted by process ontologists, of 'active kinds of changing'. As they put it, '[t]here are kinds of changing just as there are kinds of entities. These different kinds are recognized by science and are basic to the ways that things work' (MDC, p.5). Activities are 'constitutive of the transformations that yield new states of affairs or new products', and '[a]n activity is usually designated by a verb or verb form (participles, gerundives, etc.)' (MDC, p.4) – examples might include pushing, pulling, bonding, and so on.

It is worth noting right away an important issue here: whether 'active kinds of changing' should be regarded as things in or aspects of the world or rather as psychologically contingent aspects of the descriptions we give of things in the world. Torres (2009) pursues a course that is broadly compatible with the second interpretation. He argues that under the MDC account activities amount to reified causes. This is problematic, however, because molecular/cellular processes sometimes involve activities that can be construed as *negative* causes, whereas MDC construe activities as having a mandatory ontic component. The example he gives is neuronal long term potentiation (LTP), in which the removal of blocking magnesium ions enables (causes) calcium ions to diffuse throught the NMDA channel (Torres 2009, pp.242-247). In their conception of activities, Torres argues, MDC conflate *how* property changes are brought about with *what* it is that brings them about. Under his proposed 'descriptivist' interpretation, activity verbs such as 'push', 'stretch' or 'break' (or, more negatively, 'enable' or 'allow') are explanatory just in virtue of their capacity to express something of the manner in which changes are brought about. This business of describing how things happen is distinct from any claim-making about what the ontologically prior things (or, perhaps and more abstractly, properties) are that do the bringing about. Distinguishing between how and what in this way confers on Torres' account the characteristic that activities do not (necessarily) stand in need of an ontic referent, and hence, Torres claims, it is able to accommodate cases of negative causation.

Psillos also disputes the notion that entities and activities enjoy equal ontic status, describing it as 'wrong-headed' (Psillos 2004, p.312). He argues for the causal adequacy of an ontology of entities together with the dispositional capacities or active powers of those entities. On this view activities are ontologically otiose because they supervene on such capacities – 'even if, from an epistemic point of view, we need to attend to the (observed) activities in order to conjecture about the capacities' (pp.313-4). (Which is one of the bases on which MDC attempt to justify the primacy they accord to activities over properties.) Pertinent to the arguments of both Psillos and Torres are the computational molecular dynamics simulations employed by protein scientists to gain insight into peptide dynamics and protein folding (Karplus and McCammon 2002; Ho and Dill 2006). These work by computing the interactions between the atoms of the molecular system being studied, such as a peptide surrounded by water molecules, in order to calculate their motions. The 'atoms' in such simulations are represented very simply in terms of points in Cartesian space, their interactions governed by laws representing the different inter-molecular forces (e.g. an appropriately scaled inverse-square law to model electrostatic interactions). At each time step of the simulation the forces acting on each atom are computed and the atoms are then displaced accordingly, and this procedure is repeated many times. The point is that to carry out the simulation what we need to know about are the entities (atoms) and their properties (locations, laws of interaction, covalent bonding relationships). To the extent that distinct activities can be identified in the context of an MD simulation – the effects of one part of a molecule on another might be describable in terms of activity verbs such as pushing, stretching, rotating or constraining, for example – they arise as a result of a property-driven interplay of entities.

Glennan: parts and interactions

The ontological character of the MDC account contrasts with that of Glennan's account (1996, 2002). He proposes that:

> A mechanism underlying a behaviour is a complex system which produces that behaviour by the interaction of a number of parts according to direct causal laws. (Glennan 1996, p.52)

The emphasis here, ontologically speaking, is on parts (entities) and their interaction in ways that can be expressed in terms of laws; it is these law-conforming interactions that give rise to behaviours (which presumably approximate to the activities of the MDC account). This proposal, in terms of the direction and nature of the dependency between parts and behaviours, accords with the character of MD simulations described previously. However, I do not wish to make the claim that MD simulations should be considered as exemplifications of mechanism in some general sense appealed to by all the mechanism talk we see in science and elsewhere. I simply note that the MD simulation example and Glennan's account intersect in the idea that interacting parts or entities give rise to behaviours or activities.

Potential difficulties faced by Glennan's account centre on the status of causal laws and their ability to underwrite a general account of the explanations we give of phenomena, on the ontological nature of parts, and on what I take to be the fact that mechanistic explanations do not always require or admit of reinterpretation or 'bottoming out' in terms of fundamental entities or causal laws. Focusing just on the nature of mechanism parts, Glennan says that 'parts must be objects' whilst also noting that for his concept of mechanism 'to be sufficiently general it is important that a very wide variety of entities may be parts of mechanisms' (1996, p.53). For example, he suggests, '[p]arts may be simple or complex in internal structure, they need not be spatially localizable, and they need not be describable in a purely physical vocabulary. In certain contexts … one might wish to consider genetic mechanisms whose parts are genes or information processing mechanisms whose parts are software modules or data structures' (ibid). This seems right, and what we know about many biological phenomena suggests that any constraint that parts of mechanisms be objects in some mereologically and compositionally uncomplicated sense is too stringent.[5] But in his follow-up account Glennan invites us to take such a constraint seriously:

> [Parts] must have a relatively high degree of robustness or stability; that is, in the absence of interventions, their properties must remain relatively stable. Generally these parts can be spatially localized. (2002, p.S344)

In relation to the structural stability of parts, Glennan's definition of mechanism tilts, if only in one respect, towards the commonsense conception of a machine. This is one of the principal senses of mechanism recognized by the Concise Oxford English Dictionary:

(1) 'a piece of machinery', and
(2) 'a process by which something takes place or is brought about'.[6]

Machines – which typically are largely solid-state functional objects such as, say, lathes, dishwashers, and earth-moving equipment – consist of parts assembled hierarchically into specific configurations. Part shapes and spatial relationships, which are preserved by the structural robustness conferred by the solid state, serve to define particular points, lines and planes of articulation, in accordance with which the parts are able to move relative to each other (Gregory 1981). The number of degrees of freedom (or modes of configurational change) in a machine is extremely limited relative to the number that exists in the same amount of matter in fluid form. In general, changes in the configurations of a machine's parts that are constitutive of its normal operation are relatively unaffected by the specifics of the ambient environment, at least within the normal range of operating conditions.[7] A machine's parts perform particular functions, and when correctly assembled they enable the machine to discharge its overall function(s). The hierarchy of parts maps onto a parallel functional hierarchy, and another way of expressing the point is to say that a machine's structural and functional decompositions are aligned with each other.

Molecular machines & the diversity of biological phenomena

The machine conception comes strikingly to the fore in biology in the guise of 'molecular machines' (Morange 2006).[8] In the editorial overview of a 1998 special issue of *Cell* Bruce Alberts proposes that 'the entire cell can be viewed as a factory that contains an elaborate network of

---

[5] If Glennan's parts can be equated, at least approximately, with MDC's entities then perhaps the latter should be regarded as being more problematic than I suggested earlier – perhaps as problematic as activities.
[6] A third sense is given, relating to the mechanical philosophy that came to prominence in the 17th Century, but I shall not consider it here (Concise Oxford English Dictionary, 10th Edition).
[7] Obviously I am overlooking devices designed to sense aspects of their environment and respond accordingly, such as thermostats or pressure sensors.
[8] This is an idea that goes back at least nearly half a century, Monod writing in 1971 of how '[w]ith the globular protein we already have, at the molecular level, a veritable machine' (Monod 1971, p.98).

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).[9] 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'. These highly constrained causal relationships amongst localized parts do indeed sound highly machine-like. Moreover, the parts appear to be identifiable with particular functions, supporting the idea that the use of strongly mechanistic terminology is often underpinned or motivated by the possibility of identifying stable structure—function relationships. It appears to be considerations such as these that lead researchers to speak of certain macromolecular complexes in strongly mechanistic terms. For example, ClpX, an ATPase involved in protein degradation and disaggregation and a member of the so-called AAA+ superfamily of ATPases, is described as working 'like a machine with a two-speed transmission' (Zolkiewski 2006, p.1096).

Rather puzzlingly MDC do not mention protein machines, although sometimes what they say is suggestive of the mereological characteristics we associate with machines:

> 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)

Bechtel is even clearer:

> 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)

These passages suggest (as some of what Glennan says suggests) that to some degree biological mechanisms share with machines the characteristics of compositional stability and hierarchical mereology. Certainly where molecular machines are concerned the salience of such an association can readily be granted. But it would be a mistake to attribute such characteristics to cell biological phenomena in an unrestricted way. Here neo-mechanist accounts are insufficiently nuanced. The mechanism examples they discuss, such as DNA replication and neuronal long term potentiation, are not situated within an overall framework that comprehends the sheer

[9] Protein machines are a class of molecular machine.

diversity of the phenomena described by molecular and cell biology. The MDC account sometimes seems so loose as to subsume almost everything that exhibits order of almost any kind, yet at other times a more constrained notion involving only somewhat stable structures appears to be indicated. Without greater clarity it is difficult to know what fraction of phenomena neo-mechanist accounts are intended to cover. Which specific phenomena, if any, should we deem to be non-mechanistic? (And should any non-mechanisticity, so to speak, be regarded as an objective characteristic of the phenomena or as the product of aspects of human psychology, or perhaps as some compound of the two?)

The nature of the structure—function relations that pertain to phenomena, mentioned above in relation to the machine conception of mechanism and protein machines, provides one basis for differentiation and classification. While often we are able to identify particular biological structures with specific functions, where other phenomena are concerned it is much harder to make such associations. A structure may be associated with several functions (which sometimes depend on the cellular context), or a function may be associated with several structures – or it may be difficult to associate functions and structures at all. Unproblematic structure—function relations might lead us to favour mechanistic descriptions and in these cases we can recognise the existence of a parallel with machine-type artefacts, in which there exists just this kind of alignment of structure with function. But many of the cell processes that biologists speak of in mechanistic terms, including some which exemplify this alignment of structure and function, are also strikingly different from machines in particular respects.

One notable difference is the fact that cell processes frequently take place in a more or less fluid phase of matter. (Ganti describes the cell as a 'fluid machine' of a kind he terms a chemoton (Ganti 2003), and it is for good reasons that a book on biologically inspired nanotechnology is entitled 'Soft Machines' (Jones 2004).) Fluid milieux enable molecules to move more or less freely and to interact via recognition processes and events of varying degrees of specificity. These in turn constitute a basis for the establishment of specific causal networks that are to some extent functionally isolated and independent of other networks. Possibilities for highly specific molecular recognition mean that everything need not interact with everything else just as chance encounters dictate. In functional terms interaction networks may be thought 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. Some of these points Darden concedes in later work, for example when she notes that:

> in the mechanism of protein synthesis (as in any synthesis reaction), some of the entities of the mechanism are not intact and in place prior to the initiation of the start conditions. Some are made on the fly and rapidly degraded after they play their role (e.g., some messenger RNAs). Thus the analogy to a system with stable parts that either operate or do not fails for some changing components in this mechanism. (Darden 2006, p.281)

The reference to the analogy with a system of stable parts relates to discussion of whether a stopped clock is a mechanism on the MDC construal of the concept. Darden says that although it is a machine it is not a mechanism, for the reason that activities are lacking – 'the entities are in place but not operating. ... When appropriate set-up conditions obtain (e.g., winding a spring, installing a battery), then the clock mechanism may operate' (ibid.).

What is surprising here is that more explicit consideration is not given to the non-equilibrium nature of biological systems. Once we start worrying about structural dynamics then we must face up to the fact that while over certain timescales some biological systems such as cells and organisms look structurally somewhat invariant, this appearance of continuity of form and functional identity is achieved through processes that involve the continual turnover and throughput of matter at a variety of rates.[10] When those processes stop, the structure of the host system (unlike that of a mechanical clock) begins to decay and its revivification becomes impossible. In energy terms the organism must, figuratively speaking, run in order just to stay in the same organizational place. The relationship between structures and processes in biological systems is thus frequently by no means straightforward. The distinction between low-flux and high-flux processes is helpful as a way of expressing this important aspect of the diversity we see in biological phenomena (Powell 2009, pp.118-122). Moreover, it provides a basis on which to distinguish between certain biological mechanisms and machines, as well as an additional dimension through which to express the diversity of biological phenomena. We can see a functional protein complex as machine-like inasmuch as it involves the coordinated operation of hierarchically organized structures that are compositionally stable over the timescales over which the complex performs its operational cycles. In relation to this discussion, however, the interesting point is that functional processes grounded in relatively stable structures are quite different from those based on interactions between highly mobile, labile or transient structures, yet we readily apply mechanistic terminology to both.

Regularity

An aspect of the machine conception is regularity, and where a macroscopic mechanical contraption is concerned this typically equates to repetitive or consistent patterns of configurational change associated with the machine's operation. Protein machines often undergo comparable cycles of configurational change, and usually these are accompanied by (and reflect the occurrence of) sequences of chemical operations such as covalent bond making and breaking. MDC too talk about the importance of regularity, their basic mechanism definition ('entities and activities organized such that they are productive of regular changes') asserting that what are regular are the changes that mechanisms bring about. But what sorts of change are we talking about here, and in what respects are the changes regular? Presumably something like similarity of outcome is often meant – but does that mean structural outcome or outcome in terms of, say, the flux rate of a particular reaction or the level of expression of a specific gene, or

[10] Schrödinger discussed how living systems build order by feeding off what he termed negative entropy (Schrödinger 1944), and more recently authors such as Dupré have emphasized the distinctiveness of these and related properties of living systems (Dupré 2008).

result in terms of the instantiation or maintenance of some specific functional capacity? It can, I suggest, mean all these things and more.

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' (MDC, p.3). Again a parallel with the regularity displayed by machines just mentioned is implicit in this assertion, and the use of 'way' is interesting. Despite its great vagueness it functions well at an intuitive level: real-life examples of scientific mechanism talk do often seem to be at least partly illuminated when we say that 'the mechanism of X' means roughly 'the way in which X happens'. (There is a close connection here with the second dictionary definition of mechanism, as 'a process by which something takes place or is brought about'.) In this context it is interesting to consider the case of protein folding. This is the biophysically fundamental phenomenon whereby under physiological conditions a synthesized polypeptide (chain of amino acid residues) in many, and probably most, cases spontaneously adopts a compact globular shape or conformation. Now while protein scientists routinely talk about the mechanism of protein folding or protein folding mechanisms, it is not immediately clear what sense of mechanism is invoked in such usages. What is it about how proteins are thought to fold that invites mechanism talk? I shall quickly summarize the basics of protein folding in order to provide some sort of basis for confronting the question. [11,12]

Protein folding

The folding of a polypeptide chain requires that rotations occur around those of its chemical bonds that are free to rotate. The overall conformation of the molecule is established by the relative rotations that occur around the two bonds either side of the alpha carbon atom of each of its constituent amino acid residues.[13] The angles of rotation around these two bonds are called phi and psi, and it is possible to represent the conformation of each amino acid residue as a point on a plot of phi against psi (a so-called Ramachandran plot) (Ramakrishnan and Ramachandran 1965). Now the number of conformations in principle open to a polypeptide is astronomically large. Even if we assume that phi and psi can each take only three values per amino acid residue then for a polypeptide chain made up of 100 residues the number of different possible conformations is around 9^{100}, which is greater than the estimated number of protons in the universe.[14] Yet protein molecules fold spontaneously in timescales on the order of seconds. How does it happen?

A population of unfolded protein molecules in solution consists of an ensemble of different conformations. Thermal energy manifests as Brownian motion which causes molecules

---

[11] In recent years a class of proteins has come to light that lack a unique three-dimensional structure but which nevertheless have functional activity, the so-called intrinsically unstructured proteins (Wright and Dyson 1999; Gsponer and Babu 2009). Their existence poses no threat to the points I make here, however, since it remains the case that under physiological conditions most proteins do fold, and my proximate concern is with how to explicate the mechanism talk that surrounds the protein folding phenomenon.

[12] For further information on protein folding see e.g. Dill and Chan 1997; Karplus 1997; Dobson 2003; Clark 2007; Chen et al. 2008; Service 2008.

[13] Rotations around the sometimes bulky amino acid sidechains are also involved.

[14] The point was famously made by Levinthal (1969).

to jostle and writhe about so that they sample the huge conformational space defined by the energetically accessible possibilities for bond rotations. For a variety of reasons some conformations are more stable than others. The native conformations of proteins (i.e. their functional folded shapes), as revealed for example by X-ray crystallography, are telling in this regard. Generally they are quite compact and globular (fibrous proteins being an exception), and often they contain regions of regular structure such as alpha helices and beta sheets. These are characterized by particular patterns of hydrogen bonding between different amino acids. In addition, in native protein conformations non-polar (hydrophobic) amino acid residues usually cluster towards the interior of the molecule whilst polar (hydrophilic) residues, capable of hydrogen bonding to water molecules, occur more often at the surface. These factors together mean that the native conformation is a low energy state, indeed probably the minimum energy state, of the molecule in solution.

The tendency of physical systems to adopt low-energy states means that in the absence of factors that militate against folding (such as the presence of denaturing compounds like urea, or excessively high temperature) polypeptide molecules in solution tend to adopt compact conformations (Figure 1). Hydrophobic residues cluster together, reducing the surface-to-volume ratio of the cage of water molecules surrounding the protein.[15] Secondary structures such as alpha helices reduce the polypeptide's energy by satisfying the hydrogen-bonding potential of specific pairs of amino acid residues. The early formation of stabilizing structures such as these increases the chances that particular residues, possibly quite distant in sequence terms, are brought into proximity in particular configurations that lead to the creation of further structures. Thus there is a trend towards reduction of entropy and a progressive build up of stable structure. Of particular note here is that the order in which particular events occur is likely to be highly variable across different folding instances – because of the randomness of Brownian motion, for example – even when the instances all relate to the same polypeptide sequence. There is a funnelling towards the native conformation, but from different starting points and by different routes – and with occasional reverses. Structures that play a particular role in the folding of a polypeptide sequence on one occasion may not play the same role on other occasions, and may play no such role at all in relation to the folding of a different polypeptide sequence.

The mechanism talk surrounding protein folding looks prima facie different in kind from that associated with the machine concept. Whereas that concept implies highly constrained and regular motion within a simple state space, protein folding involves irregular and complex pathways through highly complex state spaces. And while machines can generally be characterized in terms of straightforward structure—function relations, protein folding is more resistant to such a characterization. Yet there is regularity of a sort across different protein folding instances, something that is similar or which holds constant across them. This is just the fact that the same physico-chemical principles (e.g. the hydrogen-bonding propensities of certain amino acids, hydrophobic effects, electrostatic interactions, etc.) serve as a basis for explicating all instances of the phenomenon. Thus in spite of the differences of structural outcome that

[15] The intrinsically unstructured proteins mentioned previously tend to lack the bulky hydrophobic residues that promote this clustering (Gsponer and Babu 2009, p.95).

commonly obtain for different polypeptide sequences,[16] and irrespective of the fact that the folding of two different proteins may involve minimal structural similarity even where similar amino acid subsequences occur, there are substantial elements of commonality. Perhaps the most basic shared feature is that proteins generally do attain definite conformations under physiological conditions (i.e. the native conformations). Overlaid on this is the fact that similar elements of secondary structure – alpha helices, beta strands and sheets, loops, etc. – can occur in proteins that differ markedly in amino acid sequence.

Ontology, epistemology & intelligibility

The view we are led to is that a protein folding mechanism is the way in which a protein folds, and this can be described by reference to particular physico-chemical factors and the effects these typically bring about over time. This notion fits the second dictionary definition of a mechanism (what could be called the causal process sense), as 'a process by which something takes place or is brought about'. It is perhaps rather unclear how it relates to neo-mechanist accounts such as that of MDC, however – in part because the unusual nature of that account's ontological claims deflects attention from its epistemological content. One interpretation of the neo-mechanist project is that it seeks to combine some of the flexibility of the causal process conception of mechanism with scientific knowledge about material patterns of entailment within particular domains (molecular cell biology and neurobiology in the case of MDC). The causal process conception of mechanism has a strong epistemic dimension, with a mechanism in that sense being what lies behind our words, images, equations, graphs, and so on – or, probably most often, by way of some admixture of these different kinds of cognitive resource – when we say how something comes about. Now whether a description has explanatory force is, I take it, a psychologically contingent matter. Perspicuous scientific descriptions of the etiology of some phenomenon show – and specifically show *us*, with our psychological capacities and propensities – how particular structures and processes we take to exist and occur in the world (on the basis of empirical evidence) give rise to the phenomenon. It is in relation to this idea, I suggest, rather than as regards its ontological claims, that the MDC account is most attractive. MDC say 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, MDC 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). This partial disavowal of regularity sounds like a nod towards something more akin to the causal process conception of mechanism than to the machine sense, and may make some sense of neo-mechanist denials that biological mechanisms should be

[16] Contrarily, very different sequences sometimes give rise to similar overall folded structures.

thought of in machine terms. MDC are suggesting, I take it, that mechanistic explanation is a matter of making phenomena and their production conceivable or imaginable, even though they do not speak in exactly those terms. They do consider briefly the sensory basis of intelligibility, however, and argue that it is not just a visual affair:

> 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 explanation in order to amount to a convincing claim, but the involvement of kinaesthetic and proprioceptive senses in the comprehension of phenomena accords with psychological findings. In addition, 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.

Cognitive entailment and mechanistic ways of thought

A few paragraphs ago I alluded to psychological contingency in the descriptions we give of how things come about. Now I want to go further in the same direction by suggesting that we side-step MDC's allegedly ontic, and apparently problematic, dualism of entities and activities to consider another dualism. This other dualism opposes the material patterns of entailment that might reasonably be held to occur in the world on the one hand and parallel sets of cognitive entailments that unfold in our minds on the other (an orientation which Craik (1943) can be seen as foreshadowing). According to this somewhat Kantian position we don't so much describe the constituent phenomena of the world as describe our models and conceptions of them and their various inter-relationships.

I noted earlier that frequently it is possible to construe 'mechanism of X' as meaning 'the way in which X occurs'. But what is a way exactly? For present purposes it can be understood to mean something not so very distant from the everyday sense of a path; more specifically it is a cognitive path or a chain of thought. Being able cognitively to connect a phenomenon with its antecedent conditions requires us to be psychologically structured in a certain way. Our cognitive dispositions must be such that our thoughts readily lead us from causes to effects, from antecedent conditions to phenomena. This dispositional wayfaring might proceed via quite direct processes of mental association, or it might involve our thoughts proceeding via a series of steps, as when we reason our way through a problem. (We may need to generate and externalize representations of various kinds to help us get around, for example, limitations of working memory or constraints on our capacity to visualize.) What is important is that our thoughts do not obviously contradict the available empirical evidence or what we believe to be the case. If our thoughts are incapable, either subconsciously or consciously, of making the relevant connections then we tend to see gaps in our understanding (although we are also capable in certain circumstances of papering over such gaps (Keil 2003). If on the other hand we can imagine how something comes about or could come about, or more generally are able to think through how it comes or could come about (i.e. can fit it to a causal template of some sort), then we enter psychological territory in which the employment of mechanism talk becomes a possibility. Additional cognitive constraints presumably apply, with the result that we speak in mechanistic terms less frequently than these broad initial conditions suggest. Thought of in these terms mechanisms look more like things we 'read into' or project onto the world than things that have some ontically robust, objective, existence.

One additional constraint on the invocation of mechanism talk might be the possibility of identifying some goal, purpose or functional benefit which can be associated with the phenomenon in question. In other words, the attribution of mechanism is often a matter of being able to connect antecedent conditions with their actual or imagined effects, and overlaying this actual or counterfactual pattern of entailment with intentional character. (A mechanism is *for* something; it enables certain goals to be achieved.[18]) Another common constraint on mechanism attribution is the identification of a class of phenomena, as opposed to singular cases. Where something happens only once, and the chances of it happening again seem remote or otherwise uninteresting, I suspect that we tend to avoid mechanistic language. (We probably could employ it, but it would seem excessive or contrived.) But where the ways by which we group phenomena and individuate classes of events lead us to recognise a category of entailment patterns that admit of explanation as a group, my hunch is that we are more likely to speak of the mechanism by which they come about, or could be made to come about. This raises the issue of the often schematic nature of mechanistic explanation, to which MDC rightly draw attention (see MDC Section 5.3).

Mechanism schemas pick out classes of phenomena on the basis of explanatorily relevant similarities but leave pockets of vagueness and ambiguity that are capable of being 'filled in' in different ways by different phenomena. (We could say that a mechanism schema ties a variety of phenomena together at their shared points of causal and explanatory salience.) More and larger pockets make for more abstract – more schematic – schemas. But why is explanation so often a schematic business in any case? The answer probably has to do with data reduction: we live in a complex world but one that manifests a considerable degree of order. We need to make sense of things quickly by cutting through the complexity and latching onto the order, and what has been described as our psychological bias towards simplicity (Chater 1999) can be viewed in terms of an adaptive orientation towards the identification of common underlying principles and patterns of entailment.

Some of these points can be illustrated by reference to the examples listed earlier. Chemical reaction mechanisms, which I mentioned but did not discuss in any detail, describe very broad classes of molecular phenomena in which many of the chemical specifics of particular

[18] The goals can be those we attribute to a system in order to explain it or, when we seek to make instrumental use of parts of our environment, our own goals.

cases are left as unfilled variables. It is just the occurrence of a specific configuration of bond-making and bond-breaking events that makes a specific reaction a member of a class of reactions that can be described in terms of the same reaction mechanism. Understanding a reaction mechanism confers powers of control, for we can seek to create molecules bearing the relevant functional groups in the appropriate configurations in order to bring about particular molecular transformations. Thus mechanism descriptions are capable of conveying ideas about causal manipulability, in roughly the sense developed by Woodward (2003). The schematic nature of the concept of a reaction mechanism makes for epistemic potency. It enables us to see elements of uniformity within the diversity of the molecular world, and provides us with a resource for factoring into our cognitive operations larger groupings of chemical phenomena than would be possible were we to be blind to that uniformity.

Another route to schematicity and the attendant benefits of abstraction and epistemic economy lies in the capacities to identify and attribute functions. By function I mean something like Cummins' notion of causal role within a system (Cummins 1975). The close relationship that exists between explanation, mechanism, function and causation can perhaps be put like this: explanations are often essentially causal; the explanatory value of mechanism descriptions lies largely in their capacity to elucidate causality; and functional concepts provide a means for relating what a system's parts do to what the system does.[19] An important feature of functions is that in principle they can be attributed to processes as well as to structures, and thus they provide us with a resource for conceptualizing mechanisms that is capable of subsuming low-flux (structural) and high-flux (processual) systems. Functional attributions reflect an ability to black-box the details of how part of a system works to focus on the causal relationships between a structure or process and the rest of the host system. This is partly a counterfactual matter: when we know about the function of something then we generally have a set of expectations about what might happen to the system if it were altered or removed. And mechanism descriptions often seem to work in part by providing us with expectations of this sort.

The MDC account says much of value about mechanism schemas and psychological aspects of mechanistic explanation, but my sense is that it does not go far enough into psychological territory. In addition, the relationship between seeking to elucidate mechanism talk in science and aiming to develop an ontologically grounded, philosophically normative account of mechanism – which I take to be the two principal neo-mechanist aims – is not entirely clear. Can we be sure that accomplishing the second objective will help to achieve, or indeed is compatible with, the first? A significant difficulty as I see it is that the ways in which scientists deploy mechanistic terminology are continuous with highly diverse patterns of usage that extend well beyond the scientists' specialist domains. Much scientific mechanism talk draws on something like the causal process conception of mechanism, as we saw in the case of protein folding. Now that conception is very broad, potentially taking in singular causal processes and not necessarily being confined to material processes. (We can say that setting the interest rate constitutes the mechanism by which the Bank of England attempts to regulate inflation, for example.) If 'mechanism' often denotes the way in which something comes about, and if 'way'

[19] We have an impressive capacity to view what a system does from a teleological standpoint, often readily seeing it as what the system is 'for'.

can as I suspect be unpacked in terms of cognitive connectability amongst more or less schematic cognitive models of phenomena and derivative representations of various kinds, then the prospect of elucidating mechanism talk and mechanistic explanation through an ontologically spare, normative account of mechanism looks somewhat remote. This is because the underlying psychological dispositional structures – cognitive mechanisms if you will – that instantiate the cognitive models and schemas we develop and use to think about and interpret phenomena in the world are (I assume we can agree) richly interconnected and highly plastic. Words evoke images; images may be associated with words; numbers, equations, graphs and images are often readily interconvertible; etc. In short, we are protean in our abilities to connect, blend and integrate varied representations of different aspects of the world in ways that succeed in spanning particular phenomena and modeling their causal structure, and in our capacity to synthesize explanatory narratives that evoke the requisite associations in others.

No doubt often it will prove possible to explicate the use of mechanistic terminology within a narrow domain in terms of apparently objective characteristics of the phenomena studied in (and perhaps constitutive of) the domain. The resultant sense of ontological groundedness is likely to confer on such explications a philosophically satisfying clear-cut and normative character. But the relevant phenomenal characteristics are liable to differ across domains and between phenomena, and the price to be paid for ontologically grounded normativity may then be that we find that phenomenal space must be carved up into miniscule sub-domains, perhaps no bigger than a specific type of phenomenon, each of which aligns with a particular ontic construal of mechanism. In that case we will hardly be any further forward: we will still have to answer the question, if we are interested in mechanistic explanation per se, of what it is that unites the varied senses of mechanism and diverse kinds of mechanism talk invoked in relation to different phenomena. Making further progress in understanding what mechanisms are and how they explain will require us to distinguish more explicitly between what happens in the world and what happens in our minds, and to be clear about when we are talking about the former and when the latter.

Acknowledgements

I am grateful to the participants of a seminar held at Exeter University in early 2009 for useful feedback on some of my early ideas about mechanism. Thanks go to Jonathan Davies, John Dupré, Katie Kendig, Sabina Leonelli, Pierre-Olivier Méthot, Dan Nicholson and Maureen O'Malley for thought-provoking discussions. (I am especially indebted to Pierre-Olivier for valuable comments on an early draft of the chapter and to Jonathan for commissioning it and then patiently awaiting its delivery.) Finally I thank Michel Morange for the stimulating insights he provided in February 2010 at my doctoral viva voce examination.

References

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

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

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

Bechtel, W. and Craver, C.F. (2006) Top-down causation without top-down causes. *Biology and Philosophy*. DOI 10.1007/s10539-006-9028-8.

Bechtel, W. and Richardson, R.C. (1993) *Discovering Complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

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

Chen, Y., et al. (2008) Protein folding: Then and now. *Archives of Biochemistry and Biophysics* **469**(1): 4-19.

Clark, A.C. (2008) Protein folding: Are we there yet? *Archives of Biochemistry and Biophysics* **469**(1): 1-3.

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

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

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

Darden, L. (2006) *Reasoning in Biological Discoveries* (Cambridge: Cambridge University Press).

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

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

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

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

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

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

Glennan, S. (2008) Mechanisms. In S. Psillos and M. Curd (eds.) *The Routledge Companion to the Philosophy of Science* (London: Routledge).

Gregory, R. (1981) *Mind in Science* (Penguin edition 1993) (Harmondsworth: Penguin).

Gsponer, J. and Babu, M.M. (2009) The rules of disorder or why disorder rules. *Progress in Biophysics and Molecular Biology* **99**: 94-103.

Ho, B.K. and Dill, K.A. (2006) Folding Very Short Peptides Using Molecular Dynamics. *PLoS Computational Biology* **2**(4)(e27): 0228-0237.

Jones, R.A.L. (2004) *Soft Machines: nanotechnology and life* (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.

Keil, F.C. (2003) Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences* **7**(8): 368-373.

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.

McKay Illari, P. and Williamson, J. (2010) Function and organization: comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences* **41**: 279-291.

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

Mackie, J.L. (1974) *The Cement of the Universe* (paperback edition 1980) (Oxford: Oxford University Press).

Monod, J. (1971) *Chance and Necessity* (Harmondsworth: Penguin).

Morange, M. (2006) 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.

Olby, R. (2009) *Francis Crick: Hunter of Life's Secrets.* Cold Spring Harbor Laboratory Press.

Powell, A. (2009) *Molecules, Cells and Minds – Aspects of Bioscientific Explanation* (PhD thesis, University of Exeter). http://hdl.handle.net/10036/95416.

Psillos, S. (2004) A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals. *Perspectives on Science* **12**(3): 288-319.

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.

Reynolds, A. (2007) The cell's journey: from metaphorical to literal factory. *Endeavour* **31**(2): 65-70.

Schrödinger, E. (1944) *What is Life?* (Cambridge: Cambridge University Press).

Service, R.F. (2008) Problem Solved* (*sort of). *Science* **321**: 784-786.

Skipper, R.A. and Millstein, R.L. (2005) Thinking about Evolutionary Mechanisms: Natural Selection. *Studies in History and Philosophy of Biological and Biomedical Sciences* **36**(2): 327-347.

Torres, P.J. (2009) A Modified Conception of Mechanisms. *Erkenntnis* **71**: 233-251.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford University Press).

Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *Journal of Molecular Biology* **293**(2): 321–331.

Zolkiewski, M. (2006) A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Molecular Microbiology* **61**(5): 1094-1100.