Biological Mechanisms, Emergence, and Some Reflections on the Neuroscientific Account of Human Action

Jonathan Davies

1. *Introduction*

"In the case of all things which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something besides the parts, there is a cause" (Aristotle VIII.6, 1045a8-10). I would like to emphasize the last part of this quote (often omitted when introducing the concept of emergence). The point is that for Aristotle emergence did not mean the end of explanation. For anyone who wishes to account for the apparently mysterious nature of the relation between (for example) life and chemistry or mind and body, emergence has proved a central and controversial notion.

Often the different sciences are thought to reflect the layered nature of the world, with physics investigating the lowest level, and chemistry, biology, psychology, sociology, ecology and so on engaging with higher levels. Levels can be defined by reference to the entities populating them (Oppenheim and Putnam 1958) the forces characteristic of them (Simon 1996) or by reference to the sciences that investigate them (as above). This is expressed clearly in William Wimsatt's picture of multiple levels of organisation within the material realm, these levels being defined as "hierarchical divisions of stuff […] organised by part-whole relations, in which wholes at one level function as parts at the next (and at all higher) levels (Wimsatt 1994: 212)." This is the case for a reductionist metaphysics as much as for an emergentist one, the difference being that for the latter at least some of these levels are inexplicable on the basis of an understanding of the lower levels. For example, systems-level features of organisms, cells, etc. are thought to emerge from the lower-level components and their interactions in ways that are non-reducible to (not explicable in terms of) the properties of these components. However, these system features are thought to derive entirely from these lower-level processes, no mysterious vital forces are at work, and no substance dualism is invoked to account for them.

This sets up a tension that has been the main source of controversy and will be the focus of my discussion. The problem of emergent explanations is that of understanding how an emergent phenomenon can both be dependent on underlying processes and autonomous from them at the same time. Two immediate possibilities present themselves, the first being to deny that emergence has a legitimate place in our understanding of the world, the second being to explicate a notion of emergence that accounts for the apparently mysterious nature of emergent phenomena and that does so in a way that does not introduce insurmountable metaphysical problems. I shall attempt to do the latter.

At this point, some preparatory comments are in order. I will be working within the new mechanistic epistemology of science, developed by P. Machamer, L. Darden and C.F. Craver to account for the explanatory strategies pursued in molecular biology and neuroscience (Machamer *et al.* 2000). I will discuss some of the details of this analysis later and I will emphasize the importance of the notion of a complex system and a causal-role function. The aim of the paper is to illuminate the epistemic challenges that confront attempts to uncover the mechanism underlying the features and behaviours of complex systems. In so doing, I will attempt to account for a number of confusions that give rise to notions such as the causal redundancy and/or inexplicability of emergent phenomena. An understanding of the relationships that hold between system components and the properties of the emergent whole has a direct bearing on long-standing questions in the philosophy of the life sciences and mind. I will not be able to engage in detail with the latter debate but I will make a couple of points with regard to the implications of neuroscientific findings for philosophers concerned with the longstanding problem of free will.

The issue I will be dealing with is that of constructing causal (or more adequately mechanistic) explanations for high-level phenomena (e.g. organismic development, cognitive processes, mental states). This will primarily be a theoretical reflection on the relations between different levels of abstraction in scientific reasoning. I will not touch on fundamental physics (where the place of causal reasoning is problematic), rather my discussion is focused on sub-cellular level biology, however I am hopeful that a detailed look at the epistemology of multi-level explanations will be relevant at higher levels as well. Some of the points that I will be raising are:

1. The ubiquity of ontological novelty.
2. The failure of system decomposition and functional localisation (1.

and 2. combining to generate the notion of emergence).
3. Fallacies arising from confusions about the relationships between emergent levels.

2. *Emergence*

Ontological emergence refers to "features of systems or wholes that possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible)[1] relations between the parts (Silberstein and Mc Geever 1999: 182)." This failure of reduction is more than an artefact of our cognitive limitations or of theoretical autonomy. Ontological emergence gives rise to higher-level entities endowed with causally significant properties that cannot be accounted for by reference to the properties of their constituent parts. Jaegwon Kim, a strong critic of non-reductive materialism, develops this basic conception into "the central doctrines of emergentism" as follows:

1. Emergence of higher-level entities: Systems with a higher-level of complexity emerge from the coming together of lower-level entities in new structural configurations.
2. Emergence of higher-level properties: All properties of higher-level entities arise out of the properties and relations that characterise their constituent parts. Some of these properties are "emergent" some merely "resultant."
3. The unpredictability of emergent properties: Resultant properties of higher-level entities are predictable from information concerning their "basal conditions;" emergent properties are not.
4. The "unexplainability" of emergent properties: Resultant properties of higher-level entities are explicable (reducible) in terms of their basal conditions; emergent properties are not.
5. The causal efficacy of the emergents: Emergent properties have autonomous causal powers, irreducible to the causal powers of their constituents (Kim 1999: 20-22).

It is interesting to note that perhaps the clearest enumeration of the core principles and claims of emergentism is provided by its most potent critic, a situation probably not unique in philosophical disputes. So we have two aspects of strong emergence, the metaphysical (ontological

[1] The force of this use of the term "reducible" appears to point to intrinsic (context independent) properties of the parts; this notion will be explored later in this section.

novelty) and the epistemic (inexplicability on the basis of the properties of components). I will explore both these aspects.

3. *Ontological Novelty (Causal Autonomy)*

Central to the emergentist claim is that the emergent entity or system has properties, and accompanying causal powers, that are qualitatively different to and, in an important sense, autonomous from those of its constituent parts. Causal efficacy is often taken to be the indicator of the real and any defence of ontological emergence must assert the reality (and so causal efficacy) of the emergent. For many critics of emergentism it is entirely unclear how an emergent entity can be causally autonomous from and simultaneously dependent upon its constituents. Kim argues that it is an inherently unstable position, either dissolving into dualism or collapsing into reductionism (Kim 1993).

On the one hand, Mark Bedau argues that the causal autonomy of emergents means that an adequate (or complete) explanation of how these higher-level entities are related to their constituents is impossible (Bedau 2003). It also allows the emergent entities to interact with other entities in ways that cannot be understood in terms of these constituents: "They are [possessed of] primitive or 'brute' natural powers that arise inexplicably with the existence of certain macro-level entities. This contravenes causal fundamentalism – the idea that macro causal powers supervene on and are determined by micro-causal powers (Bedau 2003: 10)."

If, on the other hand, the higher-level causal powers are dependent on the properties of the constituent parts of the system, then this gives rise to what Kim called the exclusion problem (Kim 1999). Kim initially applied this criticism to mental causation, i.e. the idea that functionally instantiated mental states could cause physical (or indeed psychical) changes, but it applies generally to emergent entities. The problem is that if a higher-level entity is constituted by lower-level entities, then it is unclear how it can effect changes, independently of the changes effected by these constituents. The higher-level cause is "screened-off," or excluded from causal efficacy by its instantiation by causally efficacious components.

These criticisms would seem to suggest that higher-level emergent properties or entities are either mysterious or causally redundant. However, they both assume that the properties of the whole are to be ex-

plained by reference to the *intrinsic* properties of its parts, a view that appears to neglect the explanatory importance of relational and organisational factors. If the higher-level causal capacities[2] are merely aggregates of the intrinsic causal capacities of the components, then exclusion follows; if they are entirely independent of the causal capacities of the components, then they are mysterious. The kinds of emergent properties attributed to biological mechanisms depend on the relational properties of collections of components organised in a particular way.

I will illustrate what I understand as ontological novelty by making use of a recent formulation rooted in a mechanistic epistemology: William Wimsatt's definition of emergence as non-aggregativity (Wimsatt 2000). Aggregativity is defined in terms of the relation between system-level properties and component properties. It requires that four conditions be met, failure to meet any one of these conditions results in the emergence of a novel property. These are:

1. The invariance of the property under operations rearranging the parts of the system, or the substitution of parts with a corresponding number of parts from an equivalent class.
2. Qualitative similarity of system property under size scaling.
3. Invariance of systems property under operations involving decomposition and reaggregation of parts.
4. That there be no cooperative or inhibitory interactions between the parts in the realisation of the system property (Wimsatt 2000: 275-276).

Wimsatt illustrates the demands of aggregativity with a discussion of a multi-stage linear amplifier and suggests that only in the idealised situation of assuming that each sub-amplifier behaves entirely linearly across an almost infinite range of inputs are the conditions for aggregativity met. Wimsatt is then able to distinguish between a mechanical system, like an oscillator, which is non-aggregative, and the entirely linear amplifier system, which is aggregative. Only the latter fails to demonstrate emergence. These conditions mean that to be aggregative, a system property would have to depend entirely on the intrinsic (context independent) properties of the component parts. This is an extremely strong demand and one that is usually only met in cases described by the conservation laws of physics.

Alex Ryan argues that we are confronted with emergent properties

[2] I am taking it as uncontroversial that entities have causal capacities in virtue of their properties (whether these be intrinsic or relational).

when, upon increasing the scope of an observation (for example by including extra components), we notice a qualitative change in the properties, and so in the causal powers, of the system under investigation (Ryan 2007). This would be the case if we were to expand the scope of our observation from one component of the oscillator to encompass all of the components. There would be a qualitative change in the system (its capacity to oscillate) that would not be apparent if we were to do the same with the components in the series of amplifiers (where the change would be simply quantitative).

Craver and Bechtel argue that the claim that a mechanism has higher-level causal powers (that is causal powers not reducible to those of the components) amounts to the following set of uncontroversial propositions:

1. Mechanisms are organised collections of entities and activities.
2. Mechanisms are affected by (and have effects) upon things.
3. The parts of a mechanism taken singularly cannot be so affected or have such effects (Craver and Bechtel 2006).

Causal (and so ontological) novelty does not arise mysteriously and emergent objects/properties are not causally redundant. Rather, I suggest, new causal powers arise all the time. Whenever multiple entities come together into organised collectives and interact with their environment collectively, there is ontological novelty. It is incumbent, in my view, upon those who deny that ontological novelty is possible to explain away the apparent ability of, e.g., living cells or minds to do things that their component molecular, chemical or neural elements cannot do in other circumstances.

However, we do not usually think of all ontological novelty as emergent. There is not thought to be any mystery about the capacity of a watch to tell the time or the capacity of an automobile to move and (usually) no temptation to describe their properties as emergent. In the next section, I will lay out an approach to emergence that illuminates the epistemic features that lead us to describe some novel entities/properties as emergent while others are thought to be unproblematic. But I first have to address another problem (again identified by Kim).

4. *Ontological Novelty (Downward Causation)*

The contention that emergent objects have non-reducible causal powers also gives rise to the idea of downward causation. The things with

which an emergent entity interacts include the entities from which it emerges, so giving rise to the puzzling idea that an emergent whole can causally influence its own constituents. Emmeche and his colleagues define downward causation as a "downward effect which emanates from the emergentically defined higher level onto its constituents in the lower level (Emmeche *et al.* 2000: 14)," and for Kim "downward causation is much the point of the emergentist program (Kim 1993: 350)." Some examples of supposed downward causation are: the effect of mental events on physiological states (Robinson 2005; Campbell 1974), and the effect of selection on gene frequencies or protein production.

A number of problems with the idea of downward causation have been pointed out, some of which, it is thought, make the very notion of ontological emergence incoherent. While I have argued that criticisms of emergentism that focus on the implausibility of the autonomous causal powers of emergent objects are not decisive, the idea of downward causation does appear to be problematic. A number of commentators have pointed out that the notion of inter-level causation (in either a downward or an upward direction) is incoherent. The idea is that conceptions of emergence that depend on causal interaction between levels of organisation misunderstand the nature of their relationship and seems to run counter to many people's understanding of causation. I should note that I am addressing the dominant modern notion of causation here (what Aristotle referred to efficient causation). While there is some interesting and promising work being done to cash out emergence claims in Aristotelian terms, in much of the current philosophical discussion of emergence, and in almost all cases in which emergence is used as a category by scientists, the modern notion is the one assumed.

Firstly, the casual relation is usually thought to be asymmetrical in time and manipulability. For example, Hume made temporal priority of cause over effect one of the criteria for the attribution of a causal relation and this view remains the dominant one (Hume 1748). You cannot alter a cause by manipulating an effect and causal interactions must take place in time, whereas the relation between components and emergent wholes is synchronous and symmetrical with regard to manipulability (Woodward 2003).

It may, of course, be argued that the causal relations between ontological levels are in fact synchronous and that this is what distinguishes them from intra-level causes. However, the assumption of the synchronicity of inter-level causal relations raises a rather sticky problem of circularity. Craver and Bechtel formulate the problem as follows: "if an

object, X, has its causal powers in virtue of possessing a property, P, then if X is to exercise its powers at time t, X must possess P at t. And one might believe further that if something causes X to acquire P at t, then X does not already possess P at t until something has acted. If X's acquiring P at t is a cause of S's having _ at t, and S's having _ at t is a cause of X's having P at t then it appears the X's acquiring P at t cannot cause S to have _ until S's having _ causes X to acquire P (Craver and Bechtel 2006: 6-7)." The simultaneous dependence of the properties of components on the properties of the higher-level system and the properties of the systems on the properties of the components makes explanation apparently impossible. These mutually dependent properties appear to have arisen together out of nowhere, and for biologists this can hardly be a satisfactory situation.

Salmon's version of causation is the transmission of a mark or the transmission of a conserved quantity from one event, process or object to another (Salmon 1998). This and many other formulations of causation, as a relation between things or events, specifically rule out causal relations between parts and the wholes that they make up, as these are not distinct event, objects or processes. There is no transmission of quantities or marks as the whole already has the quantity or mark possessed by its components: "Cause and effect must be distinct events – and not only in the sense of non-identity but also in the sense of non-overlap and nonimplication (Lewis 2000: 78)." Similarly Emmeche and his colleagues argue that "[o]n the biochemical level we see nothing but individual biochemical reactions causing one another. There is simply no identifiable process through which the cell ('as such,' i.e., non-biochemically conceived) inflicts a cause on biochemistry. The cell consists of biochemical processes, we could say, but this is a non-temporal (mereological) relation and therefore non-causal in the efficient-causality use of the word (Emmeche *et al.* 2000: 20)."

Craver and Bechtel suggest that one of the reasons for the persuasiveness of the view that relations between levels of mechanism are causal is that the techniques for assessing whether a component is part of a mechanism are so similar to techniques for testing causal claims. A method that features heavily in investigations in cell biology is "to seek out correlations between the presence of some component X or the occurrence of one of its activities _ and the behaviour _ of the mechanism as a whole S (Craver and Bechtel 2006: 7)." This is done by perturbing the component and seeing if the behaviour of the mechanism S is affected, or by altering the general state of the mechanism and seeing if the behaviour of

the component X is affected. Manipulationist theories of causation (the most influential being that of Woodward [2003]) make the identification of these sorts of correlations (suitably counterfactually modified, of course) constitutive of causal explanations, so proponents of these views appear to be committed to the view that relations between levels of a mechanism are causal unless they rule out symmetrical manipulability.[3]

However, in claims that you can alter components of a system by manipulating the system as a whole, it remains entirely mysterious to me how such a manipulation could be achieved without manipulating the parts. Of course, a manipulation of one part of a system could affect another part, but this is a case of common-or-garden causal interaction. There is nothing downward about it. On the other hand, a system-wide manipulation that includes a direct alteration to the part under investigation cannot be then said to subsequently have caused this alteration in this part independently of the manipulation. It was changed by the manipulation itself (indeed Woodward specifies that in the identification of a causal relation between A and B the change in B caused by the manipulation of A cannot be induced directly by this manipulation).

Craver and Bechtel suggest an understanding of prima facie inter-level causal claims that, they argue, illuminates the practices of scientists and eliminates the metaphysical confusions that arise from taking them at face value (Craver and Bechtel 2006). For Craver and Bechtel there is a perfectly coherent interpretation of the phrase "top-down causation" which picks out hybrids of causal and constitutive relations. Causal relations are not constitutive relations and constitutive relations are not causal. Causal relations are strictly intra-level and the inter-level relations are strictly constitutive. The hybrid relation that is often referred to as inter-level causation is christened "mechanistically mediated effects (Craver and Bechtel 2006)," but detailed analysis of any particular example of a mechanistically mediated effect reveals its hybrid nature. To illuminate the notion of a mechanistically mediated effect, they cite a couple of examples.

Two deaths are described. In the first case the subject dies as the result of infection by a virus. This is a case of common-or-garden causation, despite the difference in scale. The second death is different; in this case the subject dies of a heart attack. The heart is a component of the subject's bodily mechanisms, so according to Craver and Bechtel this death is a mechanistically mediated effect. We can trace the causal path-

[3] Woodward does in fact explicitly rule this out.

ways that result in a variety of physiological mechanisms ceasing to function, but the causal interactions are all intra-level. This sequence of physiological failure does not cause the death; rather it amounts to (or constitutes) the death.

An example of supposed top-down causation is Hal the tennis player "causing" blood borne glucose to be taken up by muscle cells, phosphorylated and bound into molecules of hexosediphosphate, by increasing his activity levels: "Why did Hal's cells start using more glucose? […] Because Hal started to play tennis. Similar stories could be told about Hal's respiratory mechanisms, visual systems (Craver and Bechtel 2006: 11)." This might appear to be a fairly compelling example of top-down causation, and it is certainly reminiscent of cases that crop up in emergentist descriptions of biological processes. However, the case can be accounted for by a combination of intra-level causal interactions and inter-level constitutive relations (i.e. through an account in terms of mechanistically mediated effects). Changes in the state of the mechanism as a whole *just are* (i.e. are constituted by) changes in the components of the mechanism. The components causally interact, certainly, but there is nothing "downward," or indeed "upward" about these interactions. There is a multi-level explanation going on here, but it is in terms of mechanistically mediated effects. Locutions such as "Hal's playing tennis caused his heart-rate to increase" should be understood in these terms.

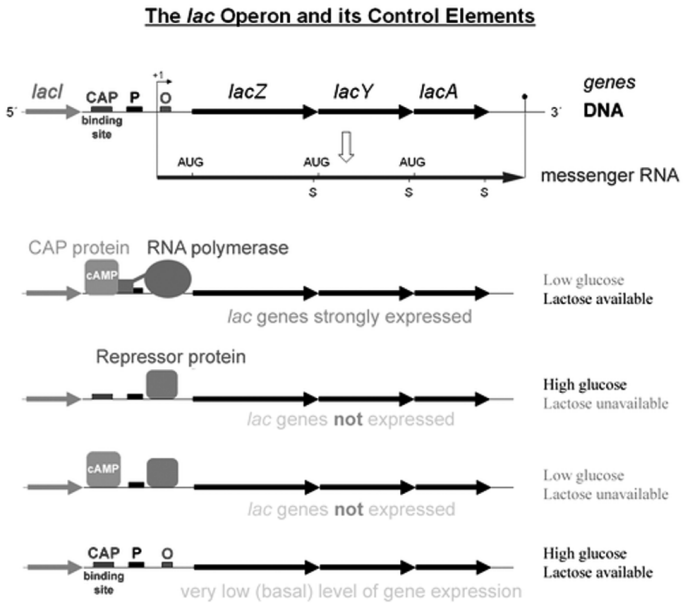
For Kim, the problems of downward causation and exclusion render emergence incoherent. However, the understanding of claims of downward causation outlined above and the previous argument for the ubiquity of ontological novelty undercut the relevance of these worries to the ontological aspect of emergence. There is no reason to assume the causal redundancy of organised collections of components and higher-level causation is not dependent on the problematic notion of downward causation. But how am I to do justice to the apparent mysterious nature of emergent phenomena as opposed to novel features thought to be unproblematic?

My approach is to supplement Wimsatt's notion of non-aggregativity with an analysis of the strategies of mechanistic explanation from Machamer, Darden, and Craver and from Bechtel and Richardson (Machamer *et al.* 2000; Bechtel and Richardson 2010). I will argue that it is the limitations of these strategies that lead us to label some system properties emergent.

5. *Decomposition and Functional Localisation*

For Bechtel and Richardson the development of a mechanistic explanation is the parsing of the behaviour and control of the system into causally (functionally) significant and intelligible segments. The explanation for the gross systemic behaviour is a process of system decomposition and the localisation of function in recognisable component structures. Bechtel and Richardson define decomposition as "the subdivision of the explanatory task so that the task becomes manageable and the system intelligible (Bechtel and Richardson 2010: 23)." Decomposition depends on the assumption that the behaviour of a system is a product of a set of subordinate functions, and that the interactions between the functional elements are minimal and can be handled additively. Functional localisation is "the identification of the different activities proposed in a task decomposition with the behaviour or capacities of specific components (Bechtel and Richardson 2010: 24)." Machamer, Darden, and Craver identify a number of different ways in which activities (the means by which a component fulfils a function within a mechanism) can be individuated, but a comprehensive mechanistic explanation is achieved only when "there are no gaps that leave specific steps unintelligible; the process as a whole [from set-up conditions to terminal conditions] is rendered intelligible in terms of entities and activities that are acceptable to the field at a time (Machamer *et al.* 2000: 12)." In molecular biology, mechanistic explanations are often multi-level in character, with lower level systems serving as components of higher-level systems. However, this hierarchical nesting bottoms out at entities and activities that are thought to be fundamental or unproblematic for the scientist's interests or field. Anomalies and problems can drive a scientist to seek explanations at lower levels, but this is the exception rather than the rule. The bottom-level entities that are dealt with by molecular biologists are macromolecules, smaller molecules, and ions, and bottom-level activities are classified, by Machamer and colleagues, into four groups: geometrical-mechanical, electro-chemical, energetic and electro-magnetic (characteristic activities of these entities [Machamer *et al.* 2000]).

To illustrate this strategy, I will make use of Jacob and Monod's operon model of gene regulation (Jacob and Monod 1961). Jacob and Monod were attempting to account for the ability of E. Coli to switch from its usual food, glucose to an alternative (lactose) when glucose is absent. (See diagram)



(Source: Wikipedia)

The *lac* operon consists of three structural genes, which code for the enzymes and transporter proteins required for lactose metabolization, a promoter and an operator, and a terminator. Nearby lies the constitutive (i.e. continually expressed) *lacI* gene that encodes a regulatory protein called the lactose repressor that, in the absence of lactose, binds to the operator. This results in the RNA polymerase being unable to recognise its binding site, the promoter, and thereby preventing expression of the structural genes. When lactose is present, a lactose metabolite, allolactose, binds to the repressor causing a conformational change which prevents it, in turn, from binding to the operator. However, the simple absence of lactose, in the presence of glucose, only results in a very low level of expression of the structural genes of the *lac* operon. For full expression, an additional regulatory protein, this time an activator, must bind to a site upstream.

We can describe this process in terms of a collection of bottom-level/primitive activities that acquire their intelligibility and explanatory adequacy by being performed by known entities, constituted in such a way that they have the required capacities for that role. For example, the binding of a repressor onto a particular DNA site is described in a popu-

lar biology text-book as follows: "the protein inserts into the major groove of the DNA helix and makes a series of molecular contacts with the base pairs. The protein forms hydrogen bonds, ionic bonds, and hydrophobic interactions with the edges of the bases […] many of the proteins responsible for gene regulation contain one of several particularly stable folding patterns. These fit into the major groove of the DNA double helix (Alberts *et al.* 2004: 271-272)." The shape and molecular constitution of the regulatory protein ensures that it is able to perform the appropriate bottom-level activities (in this case geometrical/mechanical and chemical bonding) that constitute an instance of gene regulation.

There are also, in this model, compound functional components that can in turn be decomposed into bottom-level activities performed by an identified bottom-level entity. The properties of these entities allow them to act in such a way so as to make their individual contribution to the mechanistic explanation for the behaviour of the system intelligible. The adequacy of the explanation for the regulation of the *lac* operon is a result of the gross systemic behaviour being the sum of the linear sequence of sub-tasks, which in turn are the sum of their component sub-tasks until we reach the lowest level of concern to the field. Although it is not necessary to assume that a single component (in the sense of an individual, spatially localised entity identified as a component of the system) is responsible for a sub-task, this assumption is often made, if only as a first approximation. Even in cases where components interact (a feature of what Herbert Simon called near-decomposable systems [1996]), "the behaviour of parts is intrinsically determined" and "it is feasible to determine component properties in isolation from other components. […] The organisation of the system […] provides only secondary constraints on the functioning of constituents (Bechtel and Richardson 2010: 23)." However, when the system is not near-decomposable (as is increasingly being thought the case in many biological systems), this strategy is not feasible.

6. *Failure of System Decomposition and Functional Localisation*

Molecular and cell biology is replete with cases in which a structural component, having been assigned a functional role in virtue of its inherent properties, has been subsequently incorporated into a much more complex, dynamic and integrated picture (the classic example is the role of DNA in heredity and development). Context dependence and the

non-linearity and sheer complexity of component interactions made the localisation of a component function onto an identifiable and stable structural component impossible. It is in these cases that emergence terminology is often introduced to describe the relationship between the component processes and structures and the features that they collectively give rise to. This can lead to the rejection of the possibility of mechanistic explanation (or indeed explanation of any kind). However, failure of functional localisation does not necessarily mean the impossibility of explanation. Bechtel and Richardson cite a number of examples in which explanations can be achieved in reference to component entities and activities that do not contribute in an intuitively satisfying way to the properties that they collectively constitute, one of the best known being Stuart Kauffman's network theoretic account of the stability of gene regulatory networks (Kauffman 1990; Id. 1993).

Because of the large number of components required for normal functioning, biologists' increasingly complex models of genetic regulation appear extremely fragile in the face of even quite low mutation rates. The proposed mechanism for the maintenance of functionality, selection pressure, appears in most circumstances to be insufficient to counter the apparent fragility of systems consisting of many components. Kauffman's model treats a gene regulatory network as consisting of simple interacting nodes – each node (gene) being in one of two states (on, off). The interactions between the nodes are simple activation or repression, so making transitions between successive states of the network Boolean operations. Networks of this sort are characterised by behaviour that encounters stable cycles (dynamic attractors). Thus, Kauffman maintains that, inasmuch as gene regulatory networks can be characterised in this way, they will have a natural tendency to evolve towards stability and, rather than being extremely fragile, a great deal of selection pressure is actually required to shift the system out of a stable state. By looking at the patterns of interaction between genes instead of inferring the functions of individual genes, Kauffman argues that there will be a generic order that does not require maintenance by the mechanism of natural selection.

The view that gene regulatory networks have the required connectivity to be inherently stable has recently been challenged (Sansom 2008). However, for my purposes the important point is that the purported stability of the network emerges out of the non-linear interaction of simple components. The network is organised such that its behaviour cannot be accounted for in term of the inherent properties of these components. A

hub, for example, is a hub in virtue of the number of connections it has to other nodes, and it is explanatorily significant (if at all) in virtue of this relational (extrinsic) property. The main explanatory work is done by properties that are distributed across the entire network, i.e. its structure (in this case a scale-free network characterised by a connectivity distribution that decays as a power law) and the average connectivity of the nodes. A key feature of the activities performed by the component nodes is, as Bechtel and Richardson point out, that they do not constitute functions that would appear in a mechanistic decomposition of the system. They appear to be different in kind from the activity of the system as a whole.

Kauffman's explanatory approach abandons the standard mechanistic strategy (characterised by decomposition and functional localisation) but is still, in important respects, recognisably mechanistic. Formal principles are used in an attempt to model the simultaneous activity of multiple components interacting in non-linear ways. The intuitively satisfying parsing of the behaviour into function modules is not possible, however the phenomena are still generated by material entities interacting in ways entirely acceptable to a mechanistic metaphysics. There are no "mysterious" forces at work here. The network is a complex system, producing a behaviour by the interaction of the parts, which can be characterised by direct, invariant, change relating generalisations. Or, to put it another way (echoing the definition of a mechanism by Machamer, Darden, and Craver) the network consists of parts and activities, organised so as to produce regular changes. The model describes this network. I would like to suggest that the formal models developed by Kauffman are description of a mechanism consisting of multiple interacting components (a mechanism sketch of a non-decomposable system perhaps) and as such constitutes an attempted mechanistic explanation of an emergent property. Importantly for the emergence/reduction debate, the mathematical treatment of the features of the network is a higher-level description that cannot be eliminated in favour of a lower-level story in terms of components activities (causal-role functions) inhering in discrete structural components.

7. *Emergence as Ontological Novelty Coupled with the Failure of System Decomposition and Functional Localisation*

I have laid out a notion of emergence that consists of two elements. Firstly, the coming together of multiple interacting components giving rise to ontological novelty (i.e. the emergence of new causal powers into

the world). This is revealed to an observer when the scope of their observation increases so confronting them with a qualitative change in the properties of the observed system. While there is much more that could be said about the notion of qualitative novelty, I have restricted myself to the crude observation that the emergence of novel causal capacities appears to be ubiquitous (e.g. cells can do things that their molecular components in other circumstances cannot) and the issuing of a challenge to those denying this to account for this appearance. The second, epistemological, element is that attempts to account for (explain) this ontological novelty, in terms of the intrinsic properties of the components and their linear interactions, fail. This failure renders the phenomenon unpredictable on the basis of knowledge of the properties of the components and may lead those committed to the mechanistic strategies of system decomposition and functional localisation to deny that explanation is possible or to argue that a higher-level account is required. It is the combination of these two factors that seems to underlie many of the instances of emergence talk, particularly in biology and among philosophers interested in recent developments in the life sciences.

8. *Some Consequences of This Understanding of Emergence for the Relationship Between Neuroscience and the Philosophy of Mind*

My focus has been on epistemology of biology at the level of the cell and below, however, I feel that the issues raised have relevance for other sciences as well. It seems that cognitive and neurological sciences are confronting some of the problems of the limitations of traditional mechanistic strategies that are a feature of post-genomic biology. The localisation of specific cognitive functions in areas of the brain and even the localisation of cognitive function in general in the brain as a whole are being challenged and new ways are being sought to account for mental properties. One philosophical treatment I would like to consider is Silberstein and Chemero's 2011 discussion of the dynamics of agency and intentional action (Silberstein and Chemero 2011). They treat cognitive systems as heterogeneous systems composed of brain, body and environment (by which they seem to mean natural, social and linguistic environment), non-linearly coupled into integrated complexes. Mental properties emerge out of these dynamic systems in a way strikingly (and plausibly) reminiscent of the way vital properties of organisms are thought to

emerge out of their chemical and molecular components. This understanding has consequences for how we think about some of the traditional problems in the philosophy of mind.

Firstly, I suggest that we must be attentive when discussing the relations between multiple levels of explanation in the sciences of the brain, the body and of behaviour. If mental properties are treated as emerging out of an extended brain-body-environment system, then certain assumptions about the capacities of the mind are put in a new light. Claims that, e.g., neural states cause mental states cannot be maintained. The reification of the mental, and the subsequent positing of a causal relation between the mental and the physical (or even a subsequent rejection of the possibility of such an interaction), echoes the reification of vital processes and its accompanying dualism. I would, rather, suggest that neural processes are a constitutive part of mental states. This is not a traditional identity, as other closely integrated components are involved and it is the collective processes, entities and activities that constitute the mind. It is the dynamics of brain-body-environment interaction that give rise through mechanistically mediated effects to the mental properties.

My second point is that arguments for incompatibilism, deriving from the causal chain intuition, often seem to deny the possibility of ontological novelty (the emergence of new causal powers through the interaction of multiple components in complex mechanisms). The assumption seems to be that "free will is either a force wielded by a homuncular agent or [...] free will and agency are illusions (Silberstein and Chemero 2011: 1)." Arguments for incompatibalism that rest on the assumption that the mind cannot do things that its constituent parts cannot do are not sufficient. Of course, it may be that there is something special about the power to make a (free) choice, but we cannot simply move from an identification of the intrinsic properties of the components to a conclusion that the emergent whole cannot acquire novel causal powers out of their interaction.

Finally, the elimination of higher-level description in non-decomposable complex systems does not seem to be possible. The rejection of dualism and an appreciation of the importance of neuroscience for understanding cognition and action do not commit one to the elimination of higher-level descriptions in terms of, e.g., beliefs/desires/reasons, etc. I will not comment on the adequacy of these higher-level descriptions beyond noting that efforts are underway to develop mathematical treatments of cognitive processes based on dynamical systems theory. Which approaches will ultimately turn out to be most useful is beyond my ken.

As with biological processes, it seems likely that a combination of strategies, reflecting the heterogeneity and complexity of the phenomena, will be called upon to account for cognitive process. In systems biology, computing power is being harnessed in an attempt to achieve such an integration. Still, the challenge is a huge one (even across quite small scales) and I am inclined to believe that the sheer number of factors that would have to be taken into account means that we should be modest in our ambitions with regard to simulating cognitive processes. However, it is a foolish philosopher who makes predictions about how far scientists can go.

References

Alberts, Bruce, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter, 2004, *Essential Cell Biology*, Garland Science, New York.

Aristotle, *Metaphysika*; Eng. trans. by Hugh Tredennick 1933, *Metaphysics*, Harvard University Press, Cambridge, MA.

Bechtel, William and Robert C. Richardson, 2010, *Discovering Complexity: Decomposition and Localisation as Strategies in Scientific Research*, MIT Press, Cambridge, MA.

Bedau, Mark, 2003, "Downward Causation and the Autonomy of Weak Emergence", in *Principia*, 6: 5-50.

Campbell, Donald T., 1974, "'Downward Causation' in Hierarchically Organised Biological Systems", in Francisco J. Ayala and Theodosius Dobzhansky (eds.), in *Studies in the Philosophy of Biology*, University of California Press, Berkeley: 179-186.

Craver, Carl F. and William Bechtel, 2006, "Top-down Causation without Top-down Causes", in *Biology and Philosophy*, 22: 547-563.

Emmeche, Claus, Simo Koppe and Frederik Stjernfelt, 2000, "Levels, Emergence, and Three Versions of Downward Causation," in Peter B. Andersen, Claus Emmeche, Niels O. Finnemann and Peder V. Christiansen (eds.), in *Downward Causation: Minds, Bodies, and Matter*, University of Aarhus Press, Aarhus: 13-34.

Hume, David, 1999, *An Enquiry Concerning Human Understanding*, Tom L. Beauchamp ed., Oxford University Press, Oxford; ed. or. 1748.

Jacob, François and Jacques Monod, 1961, "Genetic Regulatory Mechanisms in the Synthesis of Proteins", in *Journal of Molecular Biology*, 3: 318-356.

Kauffman, Stuart A., 1990, "The Sciences of Complexity and 'Origins of Order'", in *PSA: Proceedings of the Biennal Meeting of the Philosophy of Science Association*, vol. 2: 229-322.

Kauffman, Stuart A., 1993, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford.

Kim, Jaegwon, 1993, *Supervenience and Mind*, Cambridge University Press, Cambridge.

Kim, Jaegwon, 1999, "Making Sense of Emergence", in *Philosophical Studies*, 95: 3-36.

Lewis, David, 2000, "Causation as Influence", in J. Collins, N. Hall, and L.A. Paul (eds.), *Causation and Counterfactuals*, MIT Press, Cambridge, MA: 75-106.

Machamer, Peter K., Lindley Darden and Carl F. Craver, 2000, "Thinking About Mechanisms", in *Philosophy of Science*, 67: 1-25.

Oppenheim, Paul, and Hillary Putnam, 1958, "Unity of Science as a Working Hypothesis", in *Minnesota Studies in the Philosophy of Science*, 2: 3-36.

Robinson, William S., 2005, "Zooming in on Downward Causation", in *Biology and Philosophy*, 20: 117-136.

Ryan, Alex J., 2007, "Emergence is Coupled to Scope, not Level", in *Complexity*, 13: 67-77.

Salmon, Wesley C., 1998, *Causality and Explanation*, Oxford University Press, Oxford.

Sansom, Roger, 2008, "Countering Kauffman with Connectionism: Two Views of the Gene Regulation and the Fundamental Nature of Ontogeny", in *British Journal of Philosophy of Science*, 59: 169-200.

Silberstein, Michael and John McGeever, 1999, "The Search for Ontological Emergence", in *The Philosophical Quarterly*, 49: 182-200.

Silberstein, Michael and Anthony Chemero, 2011, "Dynamics, Agency and Intentional Action", in *Humana.Mente*, 15, 5: 1-19.

Simon, Herbert, 1996, *The Sciences of the Artificial*, 3rd ed., MIT Press, Cambridge, MA.

Wimsatt, William C., 1994, "The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets", in *Canadian Journal of Philosophy*, Supplement, 20: 207-274.

Wimsatt, William C., 2000, "Emergence as Non-Aggregativity and the Biases of Reductionisms", in *Foundations of Science*, 5: 269-297.

Woodward, James, 2003, *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, Oxford.