DOI: 10.17605/OSF.IO/5MEZ7

Emergence, expectation and causal schemas

Alexander Powell

ap.cybercraft@googlemail.com
ORCiD: https://orcid.org/0000-0001-6911-1890

November 2017

Abstract

The concept of emergence, which came to renewed prominence with the rise of the complexity sciences, has proved surprisingly resistant to analysis. After summarizing some of the characteristics associated with phenomena typically described as emergent I contrast the apparent ease with which we make sense of human behaviour with the difficulties we encounter in relation to emergent phenomena. What distinguishes and unites such phenomena, I suggest, is not primarily a matter of their objective ontological nature. Rather, we regard such phenomena as emergent because we lack cognitive schemas capable of modelling their causal structure.

Keywords: emergence, causal schemas, causal understanding, cognition, complexity, downward causation, imagination, mental simulation, reduction

Author's note: This article is a lightly revised version of Powell (2009), Chapter 6. [Version 2: addresses heading level issues; Version 3: minor adjustments to wording, additional reference.]

Introduction

The concept of emergence, which frequently surfaces in discussions about complexity (Darley 1994), has proved – like the concept of complexity itself – remarkably resistant to analysis. A considerable body of literature surrounds attempts to go beyond the well-known notion of a whole being ‘greater than the sum of its parts’ (e.g. Holland 1998; Clayton and Davies 2006; Ryan 2007). A number of taxonomies of emergence, sometimes quite intricate, have been proposed (e.g. Stephan 1999; Fromm 2005; Deguet, Demazeau and Magnin 2006), with a common feature of several accounts being a distinction between strong and weak variants. Emergence of the former kind is commonly mentioned in connection with consciousness and some of the puzzling aspects of quantum mechanics, and is typically held to be ontologically problematic (Kim 1999; Silberstein and McGeever 1999). Some contend that it may be seen too in the reaction kinetics of compartmented biochemical systems (Boogerd et al

DOI: 10.17605/OSF.IO/5MEZ7

2005).[1] More usually associated with complex systems – to which category we might reasonably suppose such biochemical systems belong – is weak emergence, which is generally considered to be more benign. This is often seemingly on the grounds that it is considered to be as much a problem of epistemology as ontology, and a number of more or less technical explications have been proposed (e.g. Bedau 1997; Holland 1998). Nonetheless some recent authors are at pains to stress the ontological reality of emergence (Bedau 2008; Huneman 2008), as if any epistemic aspect were of secondary importance to our understanding the phenomenon. Here my investigation of the concept of emergence leads me to propose that it be seen jointly in ontological and epistemological terms. Failure to do so, I claim, renders the basis on which emergence attributions are made appear needlessly mysterious.

As a starting point we might note that the word emergence, like reduction, can be used in a variety of senses ranging from the trivial to the technically sophisticated. A rabbit appearing at the mouth of its burrow exemplifies emergence of the former, everyday, kind. But even this example is not quite as banal as at first it appears, and I will return to it later. John Holland begins his 1998 book with the example of the tale of Jack and the beanstalk. Holland suggests that the transformation of bean into beanstalk is a demonstration of emergence in a sense that relates to his subsequently elaborated, highly technical, account of the concept. It seems preferable to say, however, that the example combines both everyday and more technical senses: there is the mere fact that something appears from the ground, and then there is the more mysterious transformative process by which bean becomes plant. Some of the intuitive force of terms like reduction and emergence, when used in technically specialist ways, no doubt trades on ambiguities arising from the possibility of conflation with everyday construals.

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

[1] But see discussion in Powell (2009), Chapter 4.

DOI: 10.17605/OSF.IO/5MEZ7

Facets of emergence

Reducibility, deducibility and predictability

Philosophical accounts of emergence, such as that of Broad (1925), have often been framed in terms of irreducibility or non-deducibility. Broad was concerned especially with synchronic emergence, or with the dependencies that hold simultaneously between the properties of wholes and parts of systems. This contrasts with diachronic emergence, which deals with how phenomena unfold over time. A property of a system is emergent, he says, if it cannot be deduced from the properties of its parts as they occur in it or simpler systems, and with this non-deducibility comes unpredictability.[2] Closely related to this, especially if reduction is seen in terms of formal derivability, is the idea that a system property is emergent if it is not reducible to the properties of its parts. For this to be illuminating requires that we already have a satisfactory account of reduction, but reducibility can mean many things (Powell 2009, Chapter 1). The formal derivability interpretation has the merit of clarity but in general lacks applicability to biological phenomena, even though we presumably would not say that those phenomena are in general emergent. On the other hand if we construe reducibility as explicability in terms of some factor X then we obtain a concept which has greater potential salience to biology, but which now seems to be almost too weak to be worthy of anti-reductionist ire.[3] (It is possible to see a stronger motivation for that as arising from a combination of factors. One is the fact that an organism's phenotype is not fully accounted for by its genotype, even though genes are undoubted difference-makers (Waters 2007). Another factor, more spatial or mereological in character, is the recognition that causal influence does not simply radiate outwards from the micro to the macro, but rather acts bi-directionally.)

One problem with linking predictability and deducibility is that doing so seems to make predictability too dependent on the possibility of formal, analytic, derivability. The possibilities presented by in silico simulation methods, such as those applied to the study of protein folding and dynamics, substantially decouple prediction from the constraint of mathematical analyticity. Some of Broad's examples, to do with the derivability of the physico-chemical properties of compounds from those of their constituents, are rendered suspect if not actually mistaken by scientific developments.[4] Nonetheless, the relationship between predictability and emergence is interesting, and not straightforward. Many things happen in the world that we do not, and could not,

---

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

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

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

DOI: 10.17605/OSF.IO/5MEZ7

foresee, just because of interactions between multiple contingencies in space and time – contingencies of a kind we often expect. In addition, many phenomena we call emergent occur reliably given certain conditions, and hence we expect them to occur.[5] Thus often the issue is not that phenomena are unforeseeable as such, for frequently we do expect them to occur, so much as that they threaten us with unintelligibility, in that we are unable to connect initial or underlying conditions with the resulting phenomena.

Downward causation and levels

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

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

Regarding the first point it could be argued that selection is causally rather distinctive when compared with more proximate biological processes (Mayr 1961). Many of the material phenomena we wish to explain can be accounted for in terms of the continuous or direct interactions and transformations of particles and larger aggregations of matter – it presumably being the frequent realization of this kind of possibility that establishes and reinforces the microphysicalist's intuitions. Evolution by natural selection operates by applying a filter to a population of genome-bearing organisms, against a backdrop of slowly accumulating genetic change. The generation of individual organisms does seem to be resolvable in principle, from one perspective, into continuously connected micro-level material events and processes. But the filter of natural selection typically cuts across, i.e. applies simultaneously and without regard to, the levels we standardly recognise, from the whole organism down to its constituent molecules. It effectively deletes subsets of a population so as to change the extent to which a particular class of genome sequences is represented in that population. This change in relative representation level biases the chances that a particular sequence will be propagated to subsequent generations.

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

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

The sense of downward causation I take Campbell to have intended to emphasize relates primarily to the diachronic determination of the microphysical (DNA sequence) by population-level and macroscopic events and processes. This connects the term with the causally distinctive feature of evolutionary selection noted above: its capacity to 'cause by deletion' (of a subset of a population). Another, more exclusively spatial, sense of downward causation is pervasive, however. It enters into Campbell's account in terms, for example, of the way in which particular selection events may be realized by the consumption of one organism by another (think of an amoeba engulfing a bacterium). But it can also be illustrated rather well by thinking about a protein's conformation as represented by way of Ramachandran plots (Ramachandran and Sasisekharan 1968). When the crystal structure of a protein is analysed to produce a Ramachandran plot, most of its residues are found to fall within the allowed areas – but there is considerable variation amongst the precise phi/psi values, even amongst residues of the same type. Whilst some of the variation can be attributed to imperfections of crystallographic structure refinement (Kleywegt and Jones, 1996), for most residues local steric factors of course leave considerable latitude as to the exact conformation adopted. So what determines the actual phi/psi values that obtain in the context of a specific protein? One surely has to say: downward causation. To a first approximation, in the context of the folded structure any single residue is subject to the sum of the forces exerted on it by the other residues, and these will compel it to adopt that configuration which is energetically accessible and in which the overall energy of the molecule is minimized – even if as a result the individual residue is distorted into a conformation it would not, 'in isolation', frequently adopt. Proteins thus illustrate in microcosm how wholes can determine the properties of their parts.

Generalizing from this example it is possible to express one kind of downward causation – which could be referred to as *inward causation* – in terms of the way in which the influences, or causal potentials, that act at a particular location or region can come from multiple directions and distances. The way the influences sum and act, together with the nature of whatever exists at that location, determines what happens there. [6] (Influences here often means forces arising from fields, whether acting over

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

long or short ranges. At different scales, different fields predominate.) Certain sorts of effect depend for their realisation on the attainment of a minimum local concentration of causal potential, and sometimes this can only be brought about by a summation of the individual causal potentials of multiple parts. (Think of light being focused onto a small region so as to heat it to its ignition point.) Conversely, parts are compatible – in the sense that they can retain their integrity or identity – only with particular contexts, within bounds beyond which they are changed. The conditions under which, and the manner in which, they are changed depend in complex ways on energetic and other factors. Highly non-linear behaviours can thus be generated: if a causal threshold is met, X happens; if it is not met, X does not happen.

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

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

DOI: 10.17605/OSF.IO/5MEZ7

Agents, environments and parallel processes

The concept of emergence frequently arises in the context of debates about systems consisting of multiple agents and the collective behaviours to which their interactions give rise. Termites, for example, are individually rather simple, with important aspects of their behaviour being interpretable in terms of a relatively small number of rules that key their actions to contexts. Despite this individual simplicity, termites are collectively capable of sophisticated adaptive behaviours, such as the construction of complex mound structures (Clark 1996). Cellular processes – including the fundamental capacities of metabolism, genome replication and cell division – may be thought emergent in a sense that is rather like this, for they result from the combined operation of numerous apparently independent activities. However, it has become increasingly apparent that many processes are in fact highly integrated. (See, for discussion of this, Powell (2009), Chapter 5.) Those activities are implemented by molecular mechanisms – ‘protein machines’ – of sometimes surprising sophistication (Alberts 1998; Frank 2012). A common feature is that the actions of individual agents or mechanisms appear not to be directed in any readily apparent manner towards the realization of the ultimate ends to which they do in fact contribute. An analogy is with a musical performance in which multiple instrumental parts are played simultaneously (Noble 2006). The overall melodic, harmonic and rhythmic effects for which the composer has striven arise – sometimes, and to an extent that varies according to the composer and the composition – out of the simultaneous performance of the different parts, but cannot necessarily be identified with any part in isolation.

Patterns and surprise

Musical analogies suggest another aspect of the phenomena we describe as emergent: the appearance of patterns or order of some kind. Sometimes the order is straightforwardly visual or spatial, as in the Belousov-Zhabotinsky reaction, in which the existence of alternative reaction pathways gives rise to spatial patterns that are analogous to the wave-like growth patterns exhibited in certain circumstances by the slime mould *Dictyostelium discoideum* (Peacocke 1989, pp.40-41). Sometimes, however, it takes the form of mathematical regularity or lawfulness that is apparent only under certain forms of analysis. The fact that the frequencies of words in a natural language corpus are inversely proportional to their rank order in a table of word frequencies (a relationship also referred to as Zipf’s Law[7]) is an example of this kind of ‘emergence under analysis’.

Why the appearance of order, symmetry, repetition or lawfulness in phenomena should strike us as significant is potentially fertile territory for speculation. Perhaps there is an entropic explanation: entropy favours disorder, and a lack of disorder implies the operation of some organizing force or causal principle that if understood or mastered might confer powers to predict or control. However, some spontaneous processes tend to result in homogeneity and isotropy (the dissolution of a substance in a solvent is an obvious example), and this could be characterized as a high degree of

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

symmetry. In many cases little significance is attached to the symmetry associated with such uniformity; it is as if the significance we attach to patterns and symmetries reflects what we know or feel justified in assuming about their etiology. When and why significance is attached to order and pattern are thus somewhat problematic issues and it is perhaps unwise to generalize from specific cases. Nonetheless, it seems mistaken to say that emergence attributions necessarily depend on the appearance of patterns. The oscillation of a single pendulum exhibits a simple kind of order that we would not think of as demonstrating emergence. However, when a second pendulum is coupled to the first the regular periodicity is disrupted, to be replaced by chaotic oscillatory behaviour – and that behaviour probably *would* be described as emergent.

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

Divergent and convergent processes

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

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

DOI: 10.17605/OSF.IO/5MEZ7

There may be value, so far as understanding underlying causal mechanisms goes, in distinguishing here between 'divergent' and 'convergent' processes. The outcomes of processes of the former sort are sensitive to the particularity of events, and evolution appears to demonstrate this property. If a population is reduced to a single potential breeding pair then the fate of the species and the possibility that it will give rise to successor species turns on the fates of just two individuals, which may in turn depend on a host of contingent environmental factors. Gould has asked what would happen if the 'tape' of evolution were 'played' again (Gould 1989), and if nothing else the question tests our intuitions about determinism.[8] However, it seems (to me at least) easier to entertain the idea that the outcome of each 'run' will be significantly different than the possibility that we will repeatedly observe the same outcomes, unless conditions are identical in all respects for all runs or unless the total system in which evolution takes place is extremely small and simple. The idea of there being identical conditions between runs might make no sense, if for example there were a possibility of random fluctuations at the quantum level 'leaking upwards' to give rise to macroscopic variations. The second condition – that the system is simple – effectively appeals (again) to an entropic argument: if the state space that evolutionary processes explore is sufficiently small then the probability that the same configuration will be generated within a certain number of runs may be non-negligible. Obviously, the chances are increased as the number of runs increases. It is in this sense – of extreme sensitivity to initial conditions – that the emergence to which evolution gives rise may perhaps be considered divergent.[9] Convergent processes, on the other hand, are those that are robustly indifferent to minor perturbations, and they include many of those that figure in the life of the cell and the individual organism. Development of the individuals of a species, for example, typically results in the attainment of highly characteristic forms despite taking place in sometimes seemingly quite different environments.

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

[8] An extensive debate has grown up around this question and the issue of whether there are laws in biology (see e.g. Fontana and Buss 1994 and Beatty 2006), but as its twists and turns are not central to my present concerns I sidestep them here.
[9] The intuitions I have expressed here are presumably susceptible to testing by in silico modelling and experiments in artificial life.

given. Although this account has a strong epistemological component, my starting point is the same as that from which such a more exclusively ontological project would probably proceed: consideration of the causal underpinnings of emergence.

Causal comprehension and imaginative simulation

It is presumably an uncontroversial claim about emergence to say that it relates in some way to our capacity to account for phenomena in terms of antecedent, underlying or surrounding conditions. As Mark Bedau has observed, emergent phenomena often exhibit a paradoxical duality, being both dependent on and yet apparently autonomous from their causal foundations (Bedau 1997). This can make levels talk, for all its difficulties (already noted), hard to avoid. The paradox, frequently, is the existence of macro-level phenomena that appear to conform to different ordering principles from those obtaining at the micro-level (one might say the levels are defined by the distinct sets of ordering principles). Shifting attention from one level to the other involves making something like a gestalt switch, since the connection between the macro-level phenomena and their micro-level underpinnings is not apparent.[10]

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

Debates of the 1980s and early 1990s concerning the nature of folk psychological explanation (see e.g. Davies and Stone 1995) have evolved into debates about what has been referred to as 'mindreading' (Nichols and Stich 2003). There are substantial elements of continuity to the debates, although the emphasis has shifted. Where once there was a preoccupation with whether and in what sense folk psychology could be considered a theory (Churchland 1988) there is now much greater

---

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

DOI: 10.17605/OSF.IO/5MEZ7

interest in the role and status of mental simulation, imagination and pretence (Goldman 2006; and see also Nichols 2006). One of the probable reasons for this shift is that in the intervening time philosophy of mind has been reshaped by the development of a new interest in consciousness (Searle 1992; Chalmers 1996) and in issues of embodiment and emotion in relation to cognition (Damasio 1994, 1999; Noë 2004). Additional impetus, especially on the simulation side, has come from the discovery of 'mirror neurons', first in monkeys and then in humans, that fire both when a subject performs a motor activity and when the subject views another individual performing the same activity (Gallese and Goldman 1998). [11]

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

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

Imagination and scientific thought

Let us suppose then that abilities to imagine and simulate play a part in comprehending not just the behaviour of our fellow humans – and the behaviour of other species too, to the extent that it conforms to schemas that stand in some reasonably direct relation to folk psychology – but phenomena more generally. A

[11] The existence of mirror neurons in humans is still a contested area, however (Lingnau, Gesierich and Caramazza 2009; Kilner et al. 2009). So too is the issue of how such 'resonance systems' might adjudicate between simulation- versus theory-based theories of mind (Gallagher 2007). [Update (2017): If anything the status of mirror neurons is more uncertain now than ever. However, the account presented here is I think sufficiently loose as to be able to accommodate without significant modification a wide range of possible developments on that particular front.]

plausible conjecture then is that scientific understanding is sometimes a matter of being able to imagine the behaviour of material structures and systems through processes that are underpinned by cognitive schemas of various kinds. Scientific communication can be seen as reflecting this. Much scientific writing clearly aims at evoking images in the mind of the reader or conveying patterns of causality. One way of viewing mathematical expressions in physics is as communicable data structures associated with the operation of schemas for encoding and manipulating the quantitative relationships that obtain within actual or counterfactual scenarios.

Further evidence of a role for imagination in scientific thought comes from a number of directions. First there is plain introspection: when I think about how a protein might bind to DNA, say, or about the migration of endoplasmic reticulum to the cell surface, verbal processes seem very much secondary to visuospatial ones. This introspective experience resonates with the first-person accounts reported by Jacques Hadamard (1945) that accord a central role to a visuospatial cognitive mode in mathematics, and with recent work in the epistemology of mathematics (Giaquinto 2007). In the course of his discussion of the role of visuospatial and mechanical thinking in engineering Eugene Ferguson presents evidence gathered from a variety of sources across a range of sciences (Ferguson 1992). He specifically mentions Galton, Boltzmann, Einstein, Bohr, and Heisenberg in connection with visual thinking, and notes the tendency of Faraday, Kelvin, and Maxwell to think in terms of 'models and mechanical analogies' (pp.44-45). Regarding the latter tendency he discusses Pierre Duhem's suggestion that British and French physicists of the nineteenth century adopted different styles of thought, reflecting wider cultural differences of cognitive approach.

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

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

Duhem goes on to note that in Oliver Lodge's treatise on the theory of electricity 'there are nothing but strings which move around pulleys, which roll around drums, which go through pearl beads, which carry weights...' Despairingly he says that 'We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory' (p.71). It may be that Duhem's distinction can itself be thought of as a manifestation of the very tendency he articulates so sharply, but if nothing else it highlights the possibility that scientific activity is compatible with different styles of thought. That the scientists Ferguson mentions in connection with a visual-mechanical style of thought are (Galton apart) mostly physicists and physical chemists if anything lends additional significance to the reports. For if visual thinking plays a part even in physics, where thought is often assumed primarily to be mathematical, its role in biology – in which the visualization of phenomena is often the principle aim of research – must (it could be argued) be considerable. Of Einstein, Ferguson notes that he said how he ‘rarely thought in words at all; his visual and “muscular” images had to be translated “laboriously” into conventional verbal and mathematical terms’ (p. 45).

How are first-person reports of a visuospatial component to scientific thought to be understood? A variety of findings in cognitive psychology point towards possible neuropsychological underpinnings. Roger Shepard famously noted that the time it takes for an individual to judge whether the objects shown in two pictures have the same shape is proportional to the angular difference in depicted orientation of the objects (Shepard and Metzler 1971). The two depicted objects have roughly the same shape, and it is as if the subject answers the question by mentally attempting to rotate one of the objects in order to bring its major morphological features into the same orientation as those of the other so that they can be superimposed and compared. From this kind of work has arisen a major strand of research in cognitive science and psychology concerning the part played in cognition by imagery and visual thought (e.g. Kosslyn 1994; Cornoldi et al. 1996; de Vega et al. 1996; Stenning 2002). Such work can be seen in part as working alongside fundamental research in neuropsychology concerning the participation of different brain regions in particular mental processes. The evidence that the rear portion of the neocortex is involved in visual processing is overwhelming, of course, and the neuronal basis of visual perception is (I take it) well understood (see e.g. Zeki 1993).[12] Against this neuropsychological background, cognitive psychological models of visual thought can be regarded as a layer of conjecture capable of – and necessary for – guiding research by prompting particular questions which can then be tested by experiment.

Given the rather speculative character of these ideas about simulation and imagination it is necessary to preface any philosophical theory of emergence that might draw upon or make contact with them with some words of qualification. In the first place there is little doubt that individuals differ – irrespective of their nationality! – in terms of the reliance they place on different cognitive modes in different situations (Galton 1883; Hadamard 1945; Reisberg et al. 2003). Secondly, in performing different tasks we presumably switch between, blend or integrate imagistic, mechanistic, linguistic and other cognitive styles as necessary; and sometimes we appear to make use of mental models that combine aspects of logical and abstracted visuospatial thought (Johnson-Laird 1983). And thirdly I do not mean to imply that we can necessarily take the findings and ideas from cognitive psychology in these areas for granted. However, proceeding on the basis that a certain view may be roughly right at least exposes that view to further testing; and success in clarifying specific problems might, on an abductive basis, count in the view’s favour. We should not suspend trying to make sense of the issues that interest us until, for example, the question of the nature of the mental representations that underlie phenomenal experience has been resolved. The

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

resolution of such questions may actually be facilitated by work that incorporates assumptions about the likely answers in attempting to make sense of other problems.

Cognition and material systems

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

Amongst the more epistemically tractable of material systems are the machine-type artefacts discussed at some length in Chapter 2 of Powell (2009). These artefacts consist on the whole of solid-state components in well-defined spatial relationships, capable of moving relative to one another in a limited number of ways defined by points, lines and planes of articulation established by stable part shapes. The number of degrees of freedom is extremely small relative to the number existing in the equivalent amount of matter in gaseous or liquid form. Psychological experiments suggest that we are able to simulate mentally the operation of simple mechanisms such as cogs and pulleys in order to trace how causal influences propagate through them (Hegarty 2004). It seems reasonable to speculate that the simulation process in such cases draws upon the same kinds of mental processes as those putatively involved in Shepard's work on the mental rotation of objects. Simple localized transformations of parts – principally rotations and translations – can be imagined and we can be confident, because of the properties of the solid state, that these do not have non-local consequences. Thus (it can be conjectured) the capacity of working memory is not exceeded.

Simulation and causal schemas

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

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

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

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

DOI: 10.17605/OSF.IO/5MEZ7

intelligible if we can keep in step with it, using the schemas at our disposal in order to coordinate the states of the system at different times.

I said earlier that I would return to the rabbit at its burrow entrance. What makes that case non-banal is that we cannot actually claim to know that the rabbit at the entrance was a rabbit when it was in the depths of the burrow. It may have been a dragon, and the burrow may not be what we suppose either; these things are matters of belief. But we get through life by making hosts of assumptions, on the basis of what seems most likely and most consistent with our perceptions, memories and various interpretative schemas. We imagine – as a matter of cognitive parsimony – that before we saw it at the entrance the rabbit was in the burrow, being a rabbit. Hence we do not interpret its appearance as an instance of any technically interesting sense of emergence.

Emergence and epistemology

These ideas about schemas, simulation and expectation provide a basis for thinking about emergence. In a nutshell, the concept of emergence can be seen as pertaining to phenomena that we have difficulty connecting with initial or underlying conditions. The difficulty arises when we are unable to find or construct a causal schema that fits the case, and when we describe something as emergent one of the things we are doing, in effect, is providing a report on our own cognitive state: we are acknowledging that we lack the cognitive resources to trace through, see connections within, or simulate the causal structure of some aspect of the world.[13] It may be that the requisite representations are too complex, for example because they contain more adjustable parameters than we are capable of manipulating simultaneously (because working memory becomes overloaded, perhaps). Or, as with the simple chaotic oscillator, behaviour evolves in ways that we are unable to parallel adequately in folk physical or other imaginable terms because we cannot represent or manipulate the physically important quantities in a sufficiently precise manner. An analogy is with the generation of an exception by a computer program, and we can of course speculate about the nature of the exception-raising mechanism(s). Perhaps it is simply not possible to derive a viable causal schema on the basis of what is known about the phenomena – there is nothing for a putative simulator to 'run'. On the other hand maybe a schema can be found, but when run it 'crashes' the simulator because of inconsistencies of some kind.

Whatever the psychological specifics, we can say that to the extent that we can see how they arise, we tend not to think of phenomena as emergent. We can see how phenomena arise if we can access causal schemas that are compatible with our cognitive constraints. And such schemas can be pretty rough-and-ready yet still be conducive to a sense of understanding if by entertaining them we can avoid leaving

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

DOI: 10.17605/OSF.IO/5MEZ7

glaring explanatory gaps. (See Keil (2003) for discussion of our ability to ‘paper over the cracks’ in our understanding.) This kind of perspective does not provide an objective, observer-independent method of classifying phenomena as emergent or not. Instead it suggests that if different individuals categorize in the same way then it is because they share the same perceptual and cognitive propensities and capacities – and because they share similar sensitivities to linguistic and conceptual cultural conventions.

Ontological emergence

An objection to the perspective just presented is that it seems to cast emergence in an entirely epistemic light, inasmuch as emergence attributions are seen to be made in accordance with a psychological criterion. But this is not to say that emergent phenomena have no ontological grounding, if we grant that ontological weight is conferred by the possession of irreducible causal capacities (Silberstein and McGeever 1999, p.182). Here we should think of the molecular dynamics (MD) simulations performed by protein scientists (see Powell (2009), Chapter 3). Imagine that one could perform a colossal MD simulation in which there were polypeptide chains, DNA, water molecules, and so on and so forth. Imagine too that we could run the simulation for as long as we liked, with time steps as small as we liked, and employing completely accurate interatomic potential functions, realistic control of thermodynamic parameters, etc. Given these provisos we would presumably expect that, ceteris paribus, the polypeptide chains would fold into native protein structures, and maybe we would actually see them binding to the DNA.

This example may be far-fetched, but it is not absurd. The point is that such a simulation – which I take it amounts to a virtual re-enactment of what we suppose happens in the physical world – would presumably give rise to evidence for equivalents of many of the causal capacities we would expect to see manifested by this kind of molecular system. An important qualification is that the results would be subject to the limitations imposed by the simulation’s boundedness, to acknowledge the possibility of downward/inward causation. Sometimes, for some phenomena to be generated, the spatial bounds will have to be set so large as to make simulation over anything exceeding extremely small times scales impracticable. And an MD simulation would only go so far in mimicking reality, since it would not without suitable modification model the occurrence of chemical reactions involving covalent bond breaking and making. To do that would, I take it, require the incorporation of quantum mechanical theory into the simulation, and even a very small simulated system might exhaust currently available computational resource. (It might even be that the only practicable system for modelling a particular physical system would be the physical system itself.) If we imagine that we had all the computational resource we needed, we generally reckon (I suggest) that scientific theory would enable the simulation of the molecular events that occur in biological systems. Thus we have confidence in our theoretical understanding, even though we cannot foresee what phenomena the über-simulation we are envisaging would generate (just as it rapidly becomes hard – and then impossible – to foresee future cell configurations in a Game of Life from knowledge of the current state and Conway’s rules as the number of cells increases).

DOI: 10.17605/OSF.IO/5MEZ7

Emergence and the mind

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

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

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

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

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

Bedau's position and mine overlap substantially, although it should be obvious that he is much more disinclined than I am to say that emergence reflects our cognitive limitations. And the position against which he musters his argument, which is the view that emergence might be considered to be 'merely in the mind', has something of the character of a straw man. Surely no one ever thought that emergent phenomena were *just* in the mind? We apply the term to particular phenomena, and the problem is to understand the basis on which we pick out some phenomena and not others. I have already noted the extreme ontological diversity of emergent phenomena, which ranges variously across properties, structures, dynamics, behaviours, patterns, and laws. This is potentially a problem for Bedau's account: as he notes, his interpretation of the phrase 'weak emergence' does not apply to some of the phenomena that have been described as weakly emergent by others (p.445, n6). He also grants that complexity-related weak emergence is a matter of degree (p.447). But in this case one has to ask what determines the point along the ontological scale – if indeed there is some unitary scale – at which we switch from not seeing emergence to seeing emergence.

DOI: 10.17605/OSF.IO/5MEZ7

My proposal is that the principal basis on which we pick out emergent phenomena is a negative one: an inability cognitively to model a pattern of entailment within a system. When the complexity of a phenomenon in some respect exceeds a particular level, our cognitive schemas become incapable of tracking or paralleling its causal structure (or as Bedau might put it, they become insufficient for 'crawling the micro-causal web' (Bedau 2008, p.446)). This is when we see emergence. A virtue of my account is that it provides the means to account for the attribution of emergence to radically different kinds of phenomena, in respect of which no ontological unity need exist. This will be so if making sense of the causal structure of phenomena involves multiple sorts of cognitive resource, working either in tandem or independently. Imagine that causal comprehension of a phenomenon of ontological kind P1 necessarily draws on cognitive resource C1, while comprehension of phenomena of ontological kind P2 draws on resource C2. Then P1 will look emergent if the capacity of C1 is exceeded, and P2 will look emergent if C2 is likewise over-stretched. But note that P1 and P2 need have nothing in common, ontologically speaking. Such an account has both epistemic and ontic aspects, in that we classify certain phenomena as emergent because of how their objective physical natures interact with our psychological capacities. The emergence attributions of others are intelligible to us when we make sense of the causal structure of the world using the same or similar cognitive schemas.

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

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

Bibliography

Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**(3): 291-294.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

DOI: 10.17605/OSF.IO/5MEZ7

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

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

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

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

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

Frank, J. (ed.) (2012) *Molecular Machines in Biology – Workshop of the Cell*. (Cambridge: Cambridge University Press).

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

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

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

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

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

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

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

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

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

Hadamard, J. (1945) *The Psychology of Invention in the Mathematical Field* (Princeton, NJ: Princeton University Press).

Haynes, J.-D. and Rees, G. (2005) Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Current Biology* **15**: 1301-1307.

Hegarty, M. (2004) Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences.* **8**(6): 280-285.

Holland, J.H. (1998) *Emergence* (Oxford: Oxford University Press).

Huneman, P. (2008) Emergence Made Ontological? Computational versus Combinatorial Approaches. *Philosophy of Science* **75**: 595-607.

Johnson-Laird, P.N. (1983) *Mental Models* (Cambridge: Cambridge University Press).

Kay, K.N., Naselaris, T., Prenger, R.J. and Gallant, J.L. (2008) Identifying natural images from human brain activity. *Nature* **452**: 352-355.

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

Kilner, J.M., et al. (2009) Evidence of Mirror Neurons in Human Inferior Frontal Gyrus. *The Journal of Neuroscience* **29**(32): 10153–1015.

Kim, J. (1999) Making Sense of Emergence. *Philosophical Studies* **95**: 3-36.

Kleywegt and Jones (1996) Phi/Psi-chology: Ramachandran revisited. *Structure* **4**(12): 1395-1400.

Kosslyn, S.M. (1994) *Image and Brain: The Resolution of the Imagery Debate*. (Cambridge, MA: MIT Press).

Kveraga, K., Ghuman, A.S. and Bar, M. (2007) Top-down predictions in the cognitive brain. *Brain and Cognition* **65**: 145-168.

Lingnau, A., Gesierich, B. and Caramazza, A. (2009) Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences USA* **106**(24): 9925-9930.

Lombrozo, T. and Carey, S. (2006) Functional explanation and the function of explanation. *Cognition* **99**: 167-204.

Miller, G.A. (1956) The Magical Number Seven, Plus or Minus Two. *Psychological Review* **63**: 81-97.

Mayr, E. (1961) Cause and Effect in Biology. *Science* **134**(3489): 1501-1506.

Morowitz, H.J. (2002) *The Emergence of Everything – How the World Became Complex* (Oxford: Oxford University Press).

Morton, A. (1980) *Frames of Mind* (Oxford: Oxford University Press).

Nagel, E. (1961/1979) *The Structure of Science – Problems in the the Logic of Scientific Explanation* (Indianapolis, IN: Hackett).

Noë, A. (2004) *Action in Perception* (Cambridge, MA: MIT Press).

Nozick, R. (1981) *Philosophical Explanations* (Oxford: Oxford University Press).

Nichols, S. (ed.) (2006) *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction* (Oxford: Oxford University Press).

Nichols, S. and Stich, S.P. (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford: Oxford University Press).

Noble, D. (2006) *The Music of Life: Biology beyond the Genome* (Oxford: Oxford University Press).

Norris, V., Cabin, A. and Zemirline, A. (2005) Hypercomplexity. *Acta Biotheoretica* **53**: 313-330.

Peacocke, A.R. (1989) *An Introduction to the Physical Chemistry of Biological Organization* (paperback edition) (Oxford: Oxford University Press).

Powell, A. (2009) *Molecules, Cells and Minds: Aspects of Bioscientific Explanation*. PhD thesis, University of Exeter. (Available at https://ore.exeter.ac.uk/repository/handle/10036/95416.)

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Advances in Protein Chemistry* **23**: 283–438.

Reisberg, D., Pearson, D.G. and Kosslyn, S.M. (2003) Intuitions and Introspections about Imagery: The Role of Imagery Experience in Shaping an Investigator's Theoretical Views. *Applied Cognitive Psychology* **17**: 147-160.

Resnick, M. (1996) Beyond the Centralized Mindset. *The Journal of the Learning Sciences* **5**(1): 1-22.

Ronald, E.M.A. and Sipper, M. (2001) Surprise versus unsurprise: Implications of emergence in robotics. *Robotics and Autonomous Systems* **37**: 19-24.

Ryan, A. (2007) Emergence is coupled to scope, not level. *Complexity* **13**(2): 67-77.

Schubotz, R.I. (2007) Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences* **11**(5): 211-218.

Searle, J. (1992) *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).

Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* **171**: 701-3.

Silberstein, M. and McGeever, J. (1999) The Search for Ontological Emergence. *The Philosophical Quarterly* **49**: 182-200.

Sloman, S. (2005) *Causal Models: How People Think About the World and Its Alternatives* (New York: Oxford University Press).

Smith, M.L., et al. (2007) Force-Induced Unfolding of Fibronectin in the Extracellular Matrix of Living Cells. *PLoS Biology* **5**(10): 2243-2254.

Stenning, K. (2002) *Seeing Reason* (Oxford: Oxford University Press).

Stephan, A. (1999) Varieties of Emergentism. *Evolution and Cognition* **5**(1): 49-59.

Van Gulick, R. (2007) Reduction, Emergence, and the Mind/Body Problem: A Philosophic Overview. In P. Clayton and P. Davies (eds.) *The Re-Emergence of Emergence* (Oxford: Oxford University Press).

de Vega, M., et al. (1996) *Models of Visuospatial Cognition* (Oxford: Oxford University Press).

Waters, C.K. (2007) Causes that make a difference. *Journal of Philosphy* **104**: 551–579.

White, P.A. (1998) The dissipation effect: A general tendency in causal judgments about complex physical systems. *American Journal of Psychology* **111**(3): 379-410.

Zeki, S. (1993) *A Vision of the Brain* (Oxford: Blackwell Scientific Publications).