

Emergentisms, Ancient and Modern

JONARDON GANERI

University of Sussex

j.ganeri@sussex.ac.uk

Jaegwon Kim has argued (Kim 2006a) that the two key issues for emergentism are to give a positive characterization of the emergence relation and to explain the possibility of downward causation. This paper proposes an account of emergence which provides new answers to these two key issues. It is argued that an appropriate emergence relation is characterized by a notion of 'transformation', and that the real key issue for emergentism is located elsewhere than the places Kim identifies. The paper builds on Victor Caston's important work on ancient philosophy of mind (Caston 1997, 2001), but appeals to sources he has not considered.

1. Emergence: the core issues

As a position in contemporary philosophy of mind, emergentism developed out of the early work of a number of British philosophers.¹ According to British emergentism, each special science (chemistry, biology, psychology, and so on) describes a range of causal powers that emerge from but are irreducible to the causal powers of physical particles:

British Emergentism maintains that some special science kinds from each special science can be wholly composed of types of structures of material particles that endow the kinds in question with fundamental causal powers. Chemical elements, in virtue of their minute internal structures, have the power to bond with certain others. Certain biological organisms, in virtue of their minute internal structures, have the powers to breathe, to digest food, and to reproduce. *And certain kinds of organisms, in virtue of the minute internal structures of their nervous systems, have 'the power of cognizing, the power of being affected by past experiences, the power of association, and so on'* (Broad 1925, p. 436). The property of having a certain type of structure will thus endow a special science kind with emergent causal powers. Such a structure will have an emergent causal power as a matter of law, but the law will not be 'reducible to' or 'derivative from' laws governing lower levels of complexity and any boundary conditions involving the arrangements of particles. (McLaughlin 1992, pp. 50–1; my italics)

¹ J. S. Mill (1854), and then Lewes (1875), Morgan (1923), Alexander (1920), and Broad (1925).

Generalising, the satisfaction of two conditions is typically regarded as necessary for any theory to be emergentist:

- (1) Mental properties *supervene* on physical properties.²
- (2) Mental properties confer on their instances causal powers *irreducible* to the causal powers conferred by the properties supervened on.³

One idea, indeed the leading one in discussions about emergence, is that systems of appropriate organizational complexity have causal powers which the components in the system, whether individually or together, do not. Emergentism hopes to give sense to the idea that mental properties are metaphysically *dependent* on physical properties but yet possess causal *autonomy* with respect to them.⁴

Jaegwon Kim (Kim 2006a; cf. also Kim 2006b) agrees that any version of emergentism is committed to a supervenience thesis and an irreducibility thesis, and specifically that the core emergentist idea that emergent properties contribute new causal powers neither explicable by nor predictable from the basal properties is a denial of functional reducibility. The two key issues for the development of emergentism as a viable theory, he argues, are (i) to give a *positive characterization* of the relation of emergence, beyond the mere denial of reducibility; and (ii) to solve *the problem of downward causation*, otherwise known as the exclusion problem or the supervenience problem. This is the problem that an instantiation of the supervenience base is apparently a sufficient cause for any effect attributed to an instantiation of the supervening properties. One seems forced to choose between reductionism (mental properties are ‘nothing but’ physical properties) and epiphenomenalism (mental properties are distinct from physical properties, but the residue is inefficacious): genuinely novel emergent causal power is excluded.

² See for example Van Cleve 1990; O’Connor 1994; McLaughlin 1997; Kim 1999, 2006a, 2006b; Chalmers 1996, 2006; Crane 1999; Shoemaker 2007; Macdonald and Macdonald 2010a. Humphreys (1997), O’Connor (2000), and Lowe (1996), however, defend non-supervenience-based accounts.

³ I will discuss an alleged conception of emergence which denies this, so-called ‘weak emergence’, below.

⁴ As Mark Bedau observes, ‘emergent phenomena are Janus-faced; they depend on more basic phenomena and yet are autonomous from that base’ (Introduction to Bedau and Humphreys 2008, p. 6); he elsewhere identifies dependence and autonomy as the two hallmarks of emergence (2003, p. 155).

As I will demonstrate here, the ancient philosopher of mind *Bṛhaspati*⁵ and at least some of his successors are emergentists. Responsive to the key issues Kim has identified, their work contains materials for the articulation of a conception of the mind's dependence on, and autonomy from, the physical, one that will be of considerable interest to contemporary philosophers of mind.

2. Indian physicalism

There are references in the Indian texts to a group of renegade free-thinkers whose views about human life are radically at odds with then-prevailing belief. These worldly intellectuals deny the existence of anything that smacks of the supernatural — such as transcendental beings, immaterial souls, or heavenly other-worlds. Life, they say, is for living here on earth. And they have a most interesting account of what human life itself consists in. A human person is a material body, made, like everything else, out of the four elements — but one in which thought, reason, intelligence, and consciousness arise as the physical elements are transformed, in a way similar to the way that the process of fermentation leads to the emergence of the power to intoxicate in a mixture of appropriate ingredients. The views of these philosophers, who were known as *Lokāyata* ('worldly'), or more commonly *Cārvāka*, and whose central figure is the enigmatic *Bṛhaspati*, have been deeply unfashionable, their treatises destroyed or left to rot, their ideas subject to fierce and hostile criticism. That they were nevertheless still known in the sixteenth century is evinced by the report of Abu al-Faḍl, who describes their theory for the benefit of the Mughal Emperor Akbar, saying that 'They regard paradise as a state in which man lives as he chooses, free from the control of another, and hell the state in which he lives subject to another's rule ... They admit only of such sciences as tend to the promotion of what is external, that is, a knowledge of just administration and benevolent government' (al-Faḍl 1873–1907: vol. 3, p. 217). No truck is given here to religion and other 'inner' spiritual disciplines.

The contemporary inquiry into the foundations of naturalism gives us new reasons to examine the views of these thinkers. Their most important contribution, I will claim, is a distinctive interpretation of the doctrine that psychological states are emergent physical states. A separable claim is that the self is identical to the physical body. This

⁵ Pronounced Bri-hus-puti; date unknown (but see Appendix).

second claim, which I am not going to consider here, has to do with the material constitution of the self and its identity over time, and the view is a version of Animalism, the claim being that a person is identical with the human animal and not with either an immaterial soul or a psychological continuum.

According to Bṛhaspati, thinking is due to the four constitutive principles of matter, just as the power to intoxicate is due to the ingredients in the wine. What we call a human body, or a sense organ, or a physical object, is just a combination (*samudāya*; an assemblage) of earth, fire, air, and water; indeed, these four kinds of matter are all there is. A person is a human body endowed with thinking, and individual lives differ one from another as bubbles differ in water. Recent work on the reconstruction of Bṛhaspati's text allows us to conjecture that it begins as follows:

- 1.1 Next then we will examine the nature of the reals.
- 1.2 Earth, fire, air, and water are the reals.
- 1.3 Their combination is called the 'body', 'senses', and 'objects'.
- 1.4 Consciousness (*caitanya*) [is formed] out of these [elementary reals].
- 1.5 As the power to intoxicate [is formed] out of fermenting ingredients.
- 1.6 A human being (*puruṣa*) is a body qualified by consciousness.
- 1.7 [Thinking is] from the body alone.
- 1.8 Because of its presence when there is a body.⁶

Two initial observations: First, Bṛhaspati's commitment to physicalism seems to be unambiguous. He says that earth, water, fire, and air are real, nothing else, and that what we call an object, a body, or a sense-organ is just an aggregation. The science of the four elemental reals just is (or rather, was)⁷ the science of physics, and everything which exists, it is asserted, is identical to the elements or to some combination of them. Second, Bṛhaspati's commitment to the completeness of physics is evident in his further claim that all variation in the world is due to variation in 'origin' (*janma*). The varied patterns

⁶ athātas tattvaṃ vyākhyāsyāmaḥ | pṛthivyāpastejovāyur iti tattvāni | tatsamudāye śārīrendriyaviśayasamjñāḥ | tebhyaś caitanyam | kiṇvādibhyo madaśaktivat | caitanyaviśiṣṭaḥ kāyaḥ puruṣaḥ | śārīrād eva | śārīre bhāvāt | (Bhattacharya 2002, pp. 603–4).

⁷ I take it that emergence as a philosophical thesis about the nature of metaphysical dependence on the physical is independent of the truth of any particular physical theory. For this reason I reject the suggestion that ancient philosophies of mind are no longer credible because ancient physics is not (Burnyeat 1992). Burnyeat's argument, in any case, is based on specific details of Aristotle's account of the physiology of vision, which, he alleged, prevent us from finding in him a proto-functionalist analysis of perception.

which are seen in the eye of a peacock's tail feathers come about as a result of details in their provenance, and the same explanation works for all other worldly variation:

2.1 The world is varied due to variations in origin.

2.2 As the eye in the peacock's tail.⁸

I take this to mean that there is a complete physical causal history for every change or difference, that is, as a version of the thesis that every physical effect (every 'variation') is determined by antecedent physical causes (its 'origin').

Bṛhaspati's philosophy of mind can be resolved into a pair of theses:

(Thesis 1) A human being consists in a living body made out of the four elements which, in that combination, instantiates mental properties.

It is striking that the term 'self' is not used here at all, but only the term 'human being' (*puruṣa*).⁹ The difficulty is to extract further resolution from a second thesis:

(Thesis 2) It is *due to* the combination of the elements in the body that mental properties are instantiated.

The trouble is with the ablative, which I have translated, as neutrally as possible, as 'due to'. Is the claim that thinking consists in the elements combined in a certain way, in other words, that it is *made from* them (an ablative of composition); or that it is the claim that *because of* the elements there is thinking (an ablative of explanation); or is it the claim that thinking is *produced out of* them (an ablative of causation)? Later sources will disambiguate this ablative in two different ways, as well as offering a distinct three-way disambiguation of the statements taken as a group. These disambiguations generate a range of philosophical positions about the mind–body problem. I will argue that from among the ensuing positions one can retrieve materials for a distinctive variety of emergentism.

⁸ janmavaicitryabhedāḥ jagad api vicitram | mayūracandrakavat (Bhattacharya 2002, p. 604).

⁹ Bṛhaspati reserves for the term *puruṣa* 'human being' the same sense that P. F. Strawson does the term 'person', that is with reference to specimens of a type of entity 'such that to each individual of that type there must be ascribed, or ascribable, both states of consciousness and corporeal characteristics' (Strawson 1963, p. 104).

3. Epiphenomenalism: Bṛhaspati and Dicaearchus

Among these various possibilities one suggestion is that Cārvāka philosophy of mind is a form of epiphenomenalism. According to this possibility the mind is a by-product of the material body, lacking in causal powers of its own. The question has a long history, and indeed goes back to the first presentation of Bṛhaspati's thought to a European audience. In a famous and widely circulated lecture given to a public meeting of the Royal Asiatic Society in London, February 3, 1827, Henry T. Colebrooke conjectured of Bṛhaspati that '[a]mong the Greeks, Dicaearchus of Messene held the same tenet' (Colebrooke 1837, p. 429). Dicaearchus (c. 350–285 BCE) was a member of the Lyceum and a defender of the *harmonia* theory, put forward by Simmias in the *Phaedo*, that the soul is a 'tuning' (*harmonia*) or 'tempering/blend' (*krasis*) of the body, that a blend of hot, cold, fluid, and dry material is to the soul what the tuning is to the lyre (*Phaedo* 86b7–c2). Dicaearchus wrote a dialogue, now lost, about the soul, which is mentioned by several later authors. One important source is Cicero, who reports:

In the remaining two books, [Dicaearchus] introduces a certain Pherecrates, an old man from Phthia, said to be a descendant of Deucalion, who maintains the following. The soul is nothing at all and this name completely vacuous — animals and animate things are so-called in vain [*anima* meaning 'soul'], for there is neither soul nor spirit in either man or beast. That whole power by which we act or are aware extends evenly through all living bodies and is not separable from the body. In fact, [that power] is nothing, nor is there anything else, apart from the body just alone by itself, so configured that it lives and is aware by the tempering of its nature. (*Tusc. disp.* 1.10.21 (Cicero 1927))

This is certainly the source relied on by Colebrooke, for he describes the tenet in question, the one which he finds also in Bṛhaspati, as being 'that there is no such thing as soul in man; that the principle, by which he perceives and acts, is diffused through the body, is inseparable from it, and terminates with it' (Colebrooke 1837, p. 430). The view which Sextus Empiricus attributes to Dicaearchus is that thinking is 'nothing apart from the body disposed in a certain way' (*Adv. math.* 7.349 (Sextus Empiricus 1935)). Plutarch introduces a very similar view, without attributing it: 'Or is this the case? Namely, that the substance of the soul isn't anything at all; rather, it is the tempered body which possesses the power of thinking and living' (*Against Colotes* 119ab (Plutarch 1967)).

The analogy with the tuning of a musical instrument is helpful because it reminds us that there are three different things we must keep apart: (i) the blend itself; (ii) the dispositional properties and causal powers that the body has, and for which the blend is the categorical base; and (iii) the effects of the blend either on the body or on other things. One might think of a block of ice, with crystalline structure, brittleness, and a capacity to cool other things. Reviewing ideas about the *harmonia* theory, Victor Caston observes that the idea was understood by Dicaearchus as a claim that the soul just *is* the blend (Caston 1997, 2001). Dicaearchus is described as holding that the soul is an attunement of the elements of matter that comprise the body, rather than as a power ascribable to the body in virtue of the attunement.¹⁰ Denying that the attunement has any causal powers of its own, Dicaearchus is an epiphenomenalist.¹¹

Certain classical Indian thinkers likewise interpret Bṛhaspati as identifying the mind with a combination of elements in the body (so that it has no causal power over and above that of the body), making him an epiphenomenalist like Dicaearchus, while others claim that his view is that the mind is a distinct power which emerges from the combination but is not identical to it. There is textual evidence of this disagreement among the Indian materialist philosophers of mind. Referring to the basic thesis, that it is due to the physical elements that there is thinking, we are informed that:

Here, some commentators explain that thinking arises from (*utpadyate*) the elements, while others say that it is made manifest (*abhivyajyate*) [by them]. (Kamalaśīla 1968, pp. 633,15–634,1)

¹⁰ Aristoxenus, another supporter of the *harmonia* theory, and someone who went with Alexander the Great to India, took the tuning to be of the organs and limbs rather than of the four elements within the body.

¹¹ Caston 2001, p. 185: 'He accepts Aristotle's claim that a *harmonia* cannot have causal powers. But he does not think that this is a reason to reject the *harmonia* theory; if anything, it is a reason to change our views about the soul. He thinks that while there are mental events, they are completely inefficacious—their alleged effects are to be accounted for solely in terms of the powers of the body. Dicaearchus' position is that of the modern epiphenomenalist'. Caston mentions a passage from Plutarch's *On Desire and Grief* which attributes the view, apparently to Dicaearchus, that 'Some straightforwardly extend belief and calculation into the body, saying that the soul is not a cause at all, but that it is rather by the difference, quality, and power of the body that such things come about' (Caston 2001, p. 185; cf. Caston 1997, p. 345).

Again:

Some people restore the connecting verb [in ‘due to the elements, thinking’] with ‘is manifested’ (*abhivyajyate*), but others with ‘comes into being’ (*prādhurbhavati*). (Prabhācandra 1991, p. 342,2–3)

This second source goes into much greater detail than any of the others in explaining the concept ‘manifestation’ in play here. He tells us that a manifestation is something which ‘puts together well’ or ‘refines’ and ‘perfects’ (*saṃskāraka*) what is already there, rather than bringing into being something that was not there before (Prabhācandra 1990, p. 226,12–13). As such, the manifestation is not a separate thing, over and above the four elements, even though it does have a distinctive characteristic of its own (1990, p. 225,25); it is not a ‘distinct reality’ (*tattvāntara*) (1990, p. 115,13). In this, there is certainly an affinity with the Greek word *harmonia*, which ‘derives from a verb for “fitting together”, for joining things so as to adapt or accommodate them to each other’, such that tempering ‘is the balancing of one against another so as to produce a dynamic whole’ (Caston 1997, pp. 321–2). So, the manifestation account, in thinking of states of mind as refinements of the body, is a close cousin of the *harmonia* theory as that theory was understood by Dicaearchus.

Other Aristotelians, notably Galen and Alexander, are drawn to a different reading of the *harmonia* theory, that the *harmonia* gives rise to new causal powers—and our source too distinguishes such a view among the possible interpretations of Bṛhaspati’s claim that consciousness arises from the physical elements in the right combination. This same source indeed provides us with a helpful three-fold classification of materialist solutions to the mind–body problem. A materialist must claim that the lack of distinction between mind and body consists in either (i) their necessary identity (*svabhāva*), or (ii) mind being a quality or state (*guṇa*) of body, or (iii) mind being an effect (*kārya*) of the body (Prabhācandra 1990, p. 120,22–23). This division might be brought into correspondence with the one we have seen in connection with the *harmonia* theory: the mind is either identical to the tempered body, or to the tempering itself, or to a power caused by that tempering. It might also be said to correspond with the modern distinction between reductionism, epiphenomenalism, and emergentism. From here I am interested in the first and third possibilities.

4. From covariance to material causation

The notion of supervenience is explicitly formulated in the Indian discussion of materialism, particularly in critics' descriptions of what materialism is committed to. The supervenience claim is that fixing the body's physical state fixes its mental state: two bodies cannot be distinguishable in terms of their mental properties and yet be indistinguishable in terms of their physical properties. The Latin term *super-venire* is a rendering of the Greek *epi-ginesthai* and *ginesthai epi*, terms which are used in a sense close to the modern one by Alexander and Philoponus. Philoponus in particular uses the notion in contrast with the idea that psychological characteristics simply result from (*apotelesma*) and follow (*hepesthai*) the blend of chemical ingredients, and in such a way as to allow mental states to react back on the body.¹² Donald Davidson was the first contemporary philosopher to promote the use of the notion. He did so as follows:

Mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events exactly alike in all physical respects but differing in some mental respects, *or that an object cannot alter in some mental respects without altering in some physical respects.* (Davidson 1980, p. 214; my italics)

We can see supervenience as having two components: *dependence* ('nothing can have mental-properties unless it also has physical-properties'), and *determination* ('nothing can be just like a given thing as regards its physical-properties without also being just like it as regards its mental-properties'). In short, 'every mental-property, some physical-property', and 'same physical-properties, same mental-properties' (Van Cleve 1990, p. 221).

Supervenience, I have claimed, is explicit in the formulations we have of Indian physicalist philosophy of mind. It is not to be found in what one might think of as the obvious place, however. The obvious place is a certain standard argument for physicalism, one at which Bṛhaspati hints in his laconic formulae 1.7–1.8. Other sources present the argument in similar terms:

Thinking is a quality of the body, because it is present when there is a body and absent when there is none.¹³

¹² I owe this information to Richard Sorabji. See further his 2003; 2005, vol. 1, pp. 194, 201, 202; 2010, pp. 33–4.

¹³ Gautama 1997, p. 203,3 (*Nyāya-sūtra* 3.2.46). Compare Śaṅkara 1917 (*Brahma-sūtra-bhāṣya* 3.3.53).

I will call this the *covariance argument*. Its premise is that there is a relation of ‘presence and absence’ between states of the body and states of the mind, and its conclusion is that mental states are states of the body. This relation has two components, ‘covariance in presence’ (*anvaya*) and ‘covariance in absence’ (*vyatireka*). The basic pattern, as Cardona (1967–68; 1981) has shown, is:

Covariance in presence: When B occurs (*tadbhāve*), A occurs (*tadbhāvāt*).

Covariance in absence: When B is absent (*tadabhāve*), A is absent (*tadabhāvāt*).

Here, A = mind or mental property, and B = body or physical property. On scrutiny, it is clear that we do not yet have a supervenience relation. The two halves of the rule of presence and absence resolve themselves as follows:

[Presence:] (Necessarily,) anything which has certain physical properties is thinking.

[Absence:] (Necessarily,) anything which is thinking is a physical thing (i.e. if *x* does not have physical properties, then *x* does not have any mental properties).

It is clear that this does not specify a relation of supervenience. If this is all there is to covariance, then the covariance of the physical with the mental and the mental with the physical lacks the determination component of the supervenience definition. It does not have ‘same B-properties, same A-properties’ feature, but only the weaker ‘every A-property, some B-property’. More particularly, while supervenience is an asymmetric relation between A and B, covariance is entirely symmetric. The covariance argument is at best enthymematic, but it is opaque what suppressed premiss is in the background of appeals to this argument.

Some of our sources, however, introduce the new thought that the body is the ‘material cause’ of thinking (*upādāna-kāraṇa*), and it turns out not only that the operative notion of ‘material cause’ does imply supervenience, but that this is made explicit. The idea is that, just as a sculptor could not change the features of a statue without making changes to the material out of which it is made, so too one cannot alter mental states without there being some alteration in their physical basis: we would now call this a relation of constitution.

Our sources tell us that that it is part of the notion of material cause that alterations in the material cause are implied by alterations in that which it is the material cause of. In other words, the idea of a material cause carries with it the idea of a supervenience base. Having established this principle, the critics of materialism go on to argue by *reductio* that thinking does not have the physical body as its material cause: if it did, then the mental would supervene on the physical, but it does not. Here is one important text:

Is it possible that the elements of matter be the ‘cause’ of thinking, either as the material cause or as a co-operating cause? Certainly not as the material cause, because even when they alter [thinking] does not. If one thing does not alter when another does, that other is not its material cause; the relation between a horse and a cow illustrates this. Thinking is not altered when the material elements that have been transformed into a body alter. This is not [merely] an undemonstrated assertion, for it is well known that thinking which is otherwise engaged is unaltered even by the stab of a knife, which feels [to the preoccupied thinker] no different from a rub of sandal-paste. *In exactly the same way, there can be alterations in thinking without alterations in the [elements comprising the body].* This too is not an unfounded claim, since the joyful emotion one feels when near to a lovely woman alters without one’s body changing state. (Prabhācandra 1991, p. 344,9–15; my italics)

Another source is if anything even clearer:

Nor is the ‘material cause’ view correct. For it is well known that the particular cause regarded as the material cause is one such that an alteration in the effect is impossible unless one brings about an alteration in it ... That is why someone who wants to alter something alters it only by altering its material cause, and in no other way. For when the material cause is present and its power is unimpeded, nobody can prevent the occurrence of its subsequent effect. (Kamalaśīla 1968, pp. 642,23–43,5)

What these passages show is that when someone claims that the relation between mind and body is one of ‘material causation’, better described as a relation of constitution, that is indeed to make a supervenience claim.

A modal operator is used explicitly in these two formulations, and we are in a position to consider whether the supervenience involved is strong or weak.¹⁴ Inverting the conditional, the claim here is that if *x* is

¹⁴ In the modal rather than the possible worlds formulation, given two families of properties *A* and *B*, *A* *weakly* supervenes on *B* if and only if necessarily, if anything *x* has some property *F* in *A*, then there is at least one property *G* in *B* such that *x* has *G*, and everything that has *G* has *F*; and *A* *strongly* supervenes on *B* if and only if necessarily, if anything *x* has

the material cause of y then it is not possible to bring about an alteration in y without an alteration in x . What has been said is that if x undergoes an alteration, then it is *impossible* to prevent the alteration in y .¹⁵ So the force of the statement is that there are no circumstances in which the intended alteration in y does not occur and yet the alteration in x does; that is, necessarily, if x undergoes alteration G, then y undergoes alteration F. Presuming that the entire claim has the status of a 'rule' or 'law' (*niyama*), and so that there is a second, wide scope, necessity, we can conclude that what is being attributed to the materialist who sees the relation between mind and body as one of 'material causation' is a strong supervenience thesis.¹⁶ Our source, not himself a materialist, points out that to deny strong supervenience of mind on body is not to commit oneself to denying that there are any circumstances in which physical changes necessitate alterations of the mind.

What we have so far established is that the classical theory endorses supervenience, the first requirement for a theory of emergence. With regard to the second requirement, Mark Bedau argues that we should distinguish between *strong* and *weak* conceptions of emergence in the following way.¹⁷ Strong emergence involves 'a requirement that emergent properties are supervenient properties with irreducible causal powers' (Bedau 2003, p. 158; cf. O'Connor 1994). Weak emergence involves a less demanding requirement, which in Bedau's account is the requirement that 'emergent properties can be derived from

some property F in A, then there is at least one property G in B such that x has G, and necessarily everything that has G has F.

¹⁵ Compare with Kim's formulation of what he calls the Principle of Downward Causation: 'To cause any property (except those at the very bottom level) to be instantiated, you *must* cause the basal conditions from which it arises' (1999, p. 24; my italics).

¹⁶ The locative absolute in Sanskrit, like the genitive absolute of Greek, can have a conditional, causal, temporal, or circumstantial force. Caston has observed that the genitive absolute is used, in ancient formulations of supervenience, with conditional force, expressing an antecedent (1997, p. 335). Here the locative absolute is being used in the same way. Caston has also pointed out that 'Aristotle ... might have made his claim with the outermost necessity operator left implicit; philosophers often overlook this operator when speaking more loosely' (1992). Something very similar occurs here. An agreement in presence and absence is elsewhere described as a 'rule' or a 'principle' (*niyama*) (Uddyotakara says, for instance, that 'material things possessing weight fall because of it: this is [a case of] a rule'). It follows that the supervenience relation in emergence is here affirmed to carry nomologically rather than modally strong necessity (agreeing thereby with Noordhof 2010, pp. 71–2 that supervenience in emergence is *nomologically* strong).

¹⁷ See Bedau 1997, 2003; cf. Rueger 2000a, 2000b.

micro-level information but only in a certain complex way'. The complexity requirement is what distinguishes weak emergent causal powers from the resultant properties of the system: one cannot *deduce* weakly emergent phenomena from one's knowledge of the basal conditions, but only *simulate* them.¹⁸ Weak emergence uses dynamical systems theory to demonstrate how systems can come to present emergent properties without the strong requirement. The worry is that if mental properties are only weakly emergent, then they will be epiphenomenal. In the next section, I will consider two ways in which the classical theory of Bṛhaspati is modified in later Cārvāka, precisely in response to this worry. It is a worry which was present in the minds of the classical thinkers themselves.

Let me bring this phase in the discussion to a close by returning to Colebrooke and the lecture on Indian materialism he gave in London in 1827. It is striking now how many of the ideas that were to find a place in British Emergentism are already there. The first of the British Emergentists, J. S. Mill, used the example of chemical change to illustrate his idea of a 'heteropathic law' in *A System of Logic* (Mill 1843). Mill goes on to say that 'All organised bodies are composed of parts, similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but *the phenomena of life, which result from the juxtaposition of those parts in a certain manner, bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents*' (1843, Bk. III, Ch.6, §1; my italics). It seems likely that Mill, a person whose duties as a senior official of the East India Company included correspondence with Colebrooke, and who belonged with him to a circle of London literati based around the Royal Society, would have heard Colebrooke's lecture or read it when it was published in 1837, the very period he was working on *A System of Logic*. Colebrooke's work enjoyed an extremely wide circulation — even Hegel had some of his writings, and his translations of Sanskrit mathematical treatises

¹⁸ Chalmers (2006, pp. 252–3) considers a definition of weak emergence based on complexity ('Weak emergence is the phenomenon wherein complex, interesting high-level function is produced as a result of combining simple low-level mechanisms in simple ways') but prefers a more overtly epistemological definition, resting on notions of *interest* and *unexpectedness* ('A weakly emergent property of a system is an interesting property that is unexpected, given the underlying principles governing the system'). He recommends that strong emergence is best characterized as 'non-deducibility even in principle', and suggests that consciousness is the only strongly emergent phenomenon, all other examples of emergence being weak. The conception of emergence I am developing will be stronger than Chalmers's 'weak emergence', but weaker than his 'strong emergence'.

were very well known to De Morgan and Boole. I cannot help but wonder if Bṛhaspati did not have a hand in the emergence of British emergentism.¹⁹

5. Transformation

A central element in mature Indian emergentism is the notion of a ‘transformation’ (*pariṇāma*). In many later sources the materialist is represented as holding the view that there is a transformation of those elements which are in the combination making up the body. Not mentioned in the earliest statements, it is appropriate to regard it as a development; and we can now begin to see what the motivations for such a development are. It is here that we should look for a positive characterization of the emergence relation.

Emergentism begins with the idea that systems which achieve appropriate levels of organizational complexity instantiate causal properties which are not exhibited by the components, whether as individuals or in aggregate. It is expressly stipulated that no familiar compositional model will render intelligible the emergence of these new properties. They are not scalar sums, as the mass of a whole is the scalar sum of the masses of its component parts, nor are they not vector sums, as the sum of a collection of forces results in a single new force. Nor are they mixtures, as the mixture of the colours of the parts results in the colour of the whole. The capacity to think is different in kind from any of the capacities or properties of the four elements, no matter how they are combined and synthesized. This is why Mill speaks of *heteropathic* laws and Broad of *trans-ordinal* laws; of which Bṛhaspati’s ‘thinking is from the elements’ is alleged to be an instance. The emergentists’ much-favoured example is chemical synthesis: for example, the emergence of salt and water from a reaction involving two quite different compounds, or, as our author said, the emergence of alcohol’s powers from a process whose ingredients are sugar, yeast, barley, and water.

Searle argues that micro-neuronal features are *causally* sufficient for the instantiation of macro-mental features, and that this is what it is

¹⁹ Colebrooke’s primary source for Cārvāka, Rāmātīrtha’s commentary on the *Vedānta-sāra*, was first published in 1828. It was translated into English by Ram Mohun Roy in 1832 and into German by Othmar Frank in 1835. Two influential British Indologists, J. R. Ballantyne and A. E. Gough, published translations in subsequent decades. Thus, classical Indian emergentism was readily available to English-speaking audiences in the early nineteenth century.

for the mental to ‘causally supervene’ on the physical. Mental properties are ‘system features [which] cannot be figured out just from the composition of the elements and environmental relations; they have to be explained in terms of the causal interactions among the elements’ (Searle 1992, p. 126). Searle claims that this is enough to describe them as emergent properties of the system, but distinguishes *his* concept of emergence (which he calls ‘causal emergence’) from what he describes as a ‘more adventurous’ conception, according to which an emergent feature such as consciousness could cause things that could not be explained by the causal behaviour or the neurons: ‘The naive idea is that consciousness gets squirted out by the behaviour of the neurons in the brain, but once it has been squirted out, it then has a life of its own.’ The difficulty with Searle’s account is that *neither* of his two conceptions of emergence is adequate: ‘causal emergence’ is too weak a notion, failing to sustain a robust explanation of the autonomy of emergent features, while the ‘more adventurous’ conception fails to do justice to the requirement that emergent features are dependent on the micro structures from which they emerge.²⁰ Relatedly, the idea that the relation between mental and physical properties is one of ‘material causation’ (*upādāna-kāraṇa*) is not held, by later Indian materialists, to suffice for a satisfactory characterization of emergence. They recommend a conception of emergence distinct from either of the two distinguished by Searle.

It is a deeply held intuition that nothing completely new can come into existence—nothing can come into existence which cannot be understood in terms of the nature of fundamental components and the ways they can be combined. What had formerly seemed mysterious about chemical reaction no longer surprises us, with our much better understanding of the nature of chemical bonds and the structure of atomic matter. The Cārvāka hypothesis about transformation can be seen as a way to reconcile this attachment to homopathic law with the key features of emergentism. Without a transformation in the micro-base, a homopathic theory of the emergence of psychological capacities is driven inevitably in the direction of panpsychism, for (so the thought goes) a complex could not think if the elements do not, any more than a whole could have a mass if all its parts were massless. The panpsychist alternative to emergentism has indeed

²⁰ Causal emergence has also been recommended in O’Connor and Wong 2005; Wong 2006.

been taken seriously by a number of philosophers in recent times,²¹ but our sources provide two strong counter-arguments. One is that any object at all should then have psychological capacities, and we are lacking a clear criterion why only some do and others do not. The other is that, within a single body, there will be many sites of awareness, but no ‘governing principle’ orchestrating them:

Even as the power to intoxicate resides to a small measure in each part of the intoxicating liquor, so too [the materialist must claim that] thinking is to a small measure in the parts of the body. And then many things will be thinking in one body. But it is impossible for the respective aims of many thinking entities to act in conformity, any more than many flying birds, bound by a single cord but disposed to move in conflicting directions, are able to cross even the distance of a span, even though the capacity is there for them to do so. So too the body would be unable to do anything. (Vācaspati 1980, p. 767,21–24; cf. 1996, p. 531,13–19)

It is in order to provide a non-panpsychist but not epiphenomenalist explanation of mental causation that the transformation theory is introduced. Let us suppose that the blending or combining of the elements ‘transforms’ them in such a way that in their transformed state their combination, according to homopathic principles, instantiates psychological properties. Then it will be true to say that mental properties are reducible to the properties of the transformed physical base but equally true that they are irreducible to the properties of the untransformed base.

One of our sources says that the view is that ‘matter, although insentient in its inert state, will be bestowed with consciousness when in a body transformed’ (Jayanta 1982, pp. 201,26–202,1). Another says that it is the view that thinking, although not observed in the material earth out there, is present in the elements as transformed in the form of a body (Śaṅkara 1917, p. 765,7-8). As so expressed, the idea seems to be that the elements themselves acquire new causal powers when they are in a certain state, namely the state of jointly composing a body, powers that they did not have beforehand when they were in other combinations with other elements. This is different from the view that the body as a whole has powers which

²¹ Nagel 1979, pp. 181–95; G. Strawson 2006; Van Cleve 1990. Nagel’s argument for panpsychism goes as follows: Human beings are complex systems composed entirely of matter [Materialism, Anti-Dualism]. Mental properties are not logically implied by any physical properties [Anti-Reductivism]. Human beings do have mental properties [Anti-Eliminativism]. There are no emergent properties [Anti-Emergence]. *Therefore*, the basic physical constituents of the universe have mental properties [Panpsychism].

none of its parts have individually. It is instead the view that the parts themselves have new powers *conditionally* upon their membership of the whole.

There is a resonance of this idea in the way Galen distinguishes between resultant and emergent properties. He says:

Consider the first elements. Even though these substrata are unable to perceive, a body capable of perceiving can at some point come into being, because they are able to act on each other and be affected in various ways in many successive alterations. For anything constituted out of many things will be the same sort of thing the constituents happen to be, should they continue to be such throughout; it will not acquire any novel characteristic from outside, one that did not also belong to the constituents. *But if the constituents were altered, transformed, and changed in manifold ways, something of a different type could belong to the composite that did not belong to the elements ...* Consequently, something heterogenous cannot come from elements that do not change their qualities. But it is possible from ones that do ... Therefore, it is necessary that that which is going to sense be constituted either (i) from first elements capable of sensation or (ii) from ones incapable of sensation, but naturally such as to change and alter. (Galen, *On the Elements according to Hippocrates*, 1.3, 70.15–25, 72.19–21, 74.14–17, trans. Caston, 1997, pp. 355–7; my italics)²²

Here Galen distinguishes two possibilities. One possibility is that the power to sense is an additive, resultant property, a possibility which leads directly to panpsychism. The other possibility is that the power to sense is an emergent property, and Galen's commitment to the principle that 'something heterogenous cannot come from elements that do not change their qualities' leads to the conclusion that the elements must be transformed.

The early British emergentists also use the word 'transformation', but seem to mean something rather different by it. Thus Samuel Alexander:

physiological complexes of a sufficient complexity carry mind or consciousness. They may be said to be 'transformed' in the consciousness they

²² Although rightly seeing in the passage an early anticipation of the distinction between emergent and resultant properties, Caston curiously does not remark on the role to which Galen accords the concept of transformation. Neither does Kim, who quotes the passage in his 2006b, but glosses it in such a way that the idea of transformation entirely disappears: 'Galen is saying that a composite object made up of simpler constituents, when these constituents enter into special complex relationships ("act on each other and be affected in various ways"), can come to exhibit a novel property ("something of a different type") not possessed by its constituents' (2006b, p. 189). It seems to me that Caston and Kim are too keen to read Galen as anticipating *modern* understandings of emergence, and in doing so fail to notice an idea which an examination of the Indian theory makes vivid.

carry ... the parts are used up to produce something different from them and transcending them, but, used up as they are, they are not altered or superseded but subserve. In this special sense there is a 'transformation' of the parts in building up a higher existence, *but the parts remain what they were*. (Alexander 1920, p. 370; my italics)

Alexander clearly asserts that in his view the 'parts remain what they were'. Carl Gillett (2006) has proposed that one reads this as the claim that an emergent property partly determines which causal powers are *contributed* by the base properties, that the base properties contribute causal powers in a way that is conditional upon the fact that they realize an emergent property. What distinguishes the Indian transformation theory from Samuel Alexander's is its claim that the emergent property determines not only what causal powers the base properties 'contribute', but what causal powers they actually possess. The idea is that the parts have new powers in virtue of being parts of the whole and therefore intelligible only in reference to the whole to which they have come to belong. What powers an element has is conditional on what combination it is in. Emergence by transformation is the idea that the elements have cognitive powers only when in the frame of a living body, powers they do not have in other sorts of combination or in no combination at all.

What this brings into view is the availability of a conception of emergence distinct from either of the two conceptions distinguished by Searle. The proposal motivating the transformation theory is that, when micro-entities come together in appropriately complex systems of organization, the micro-properties they instantiate are transformed so as to give rise to novel causal powers in the macro-entity they constitute. The emergence of conscious states is not merely a fact about our inability to predict the behaviour of very complex systems, nor is consciousness something which is just 'squirted out'. It is a fact about the powers of micro-entities when they belong to macrophysical structures.

6. Downward causation as assistive causation: Udbhāṭa

Downward causation is causal influence going down from the higher level of the mental to the lower level of the physical. The exclusion problem presents the very notion with seemingly insuperable difficulties. Kim's preferred way to formulate the problem is to begin with mental-mental causation (Kim 1998, pp. 41–3). Suppose that M is a mental property, that it has causal powers, and that one of its

instantiations is the cause of the instantiation of a second mental property M^* . M^* supervenes on, but is not reducible to, a physical base P^* , a set of physical properties. The instantiation of M can cause an instantiation of M^* only by causing an instantiation of its physical base P^* . This is the downward causation. But M too has a physical base P , whose instantiation is sufficient for the instantiation of M . If an instantiation M is causally sufficient for an instantiation of P^* and hence of M^* , then so too is an instantiation of P . There is an overdetermination in the causation of instantiations of M^* . This is a *reductio* of the supposition that M has additional causal powers.

Anxieties about downward causation are evidently at work in the remarkable proposal of a ninth century Indian materialist Bhaṭṭa Udbhaṭa, who acquired a reputation as a very ‘cunning’ interpreter of Bṛhaspati.²³ He observes that in the sentence ‘Due to the elements (*bhūtebhyah*) there is thinking’, the grammatical case can be construed as ablative or as dative, and he proposes that the force of the dative has to be acknowledged. The source of our knowledge about his proposal is this brief passage:

The ancient materialists like Bhāvivikta and others interpreted [sūtra 1.4] as asserting ‘Thinking is from the elements’, because the ablative has been employed in the expression ‘(from) the elements’. But Udbhaṭa interprets the expression as being in the dative, meaning thinking is to or for the elements; [he says that] thinking is autonomous (*svatantra*) and is an assistant (*upakāraka*) to the material elements which constitute the body. (Cakradhara 1982–4, vol. 2, p. 257; cf. Bhattacharya 2002, p. 606, Bhā.9)

Thinking is now not simply an effect of the combinations of material elements, transformed or otherwise, but also functions as a cause with respect to them. Thinking is ‘autonomous’ (*svatantra*), and it ‘assists’ the elements. The important notion of ‘assistive’ (*upakāraka*) causation supplements that of material causation (*upādāna*). Udbhaṭa’s distinction between two concepts of causation offers him the hope to be able to explain how the mental can display an appropriate autonomy and yet be emergent.²⁴

One of our sources offers an analogy by way of explanation of the new idea. He says that a traveller will start a fire from sparks generated

²³ Udbhaṭa seems to have attempted to adapt Cārvāka emergentism so as to accommodate Nyāya insights about the unity and autonomy of the self (his commentary on the *Cārvāka-sūtra* being a sort of metaphrasis).

²⁴ The proposal has contemporary supporters. Dretske 1988, for instance, has proposed a distinction between what he calls ‘structuring’ and ‘triggering’ causation. I will refer to an idea in Lowe 1996 below.

by rubbing sticks together, but will then use the flames to keep new material burning. Similarly, mental properties emerge through transformation from matter, and are thereafter jointly responsible along with matter for future mental states.²⁵

Another of our sources discusses the idea in the course of a careful examination of the materialist account of the causation of one mental event by another, which is also, as I noted above, the context in which Kim prefers to formulate the problem of downward causation. A mental event at one time causes a mental event at a later time. How? Two possibilities are considered. The first is this:

Now suppose one says that [for the mental event] at the later time, the body acts as an assisting cause and not as a material cause. How so? The body is an assisting cause in that it helps bring about the later effects of the thinking which has it as its material cause in the present. So that is how thinking is causally effective but not independent of the body. (Kamalaśīla 1968, p. 646,8–11)

The fundamental issue, that of reconciling autonomy with dependence, is very clearly identified here. The proposal is that mental events are jointly produced by earlier mental events in tandem with the physical bases of the earlier events. In terms of Kim's formulation of the problem of downward causation, the proposal is that M and P are individually necessary and jointly sufficient for M*, just as the fire and fuel at one time are jointly productive of fire at the next. What is difficult to understand, in this proposal, is why P is not itself sufficient, if M supervenes on P. The argument seems already to assume that M has a causal power not shared with P, but this is exactly what we are trying to explain.

A revised proposal follows:

The body, assisted by the earlier mental event, is the material cause of the later mental event. (Kamalaśīla 1968, p. 646,20–21)

²⁵ Prabhācandra 1990, pp. 118,11–118,15. E. J. Lowe (1996, p. 82) develops an account of 'autonomous mental causation' based on the idea that agents engage in 'enabling' or 'facilitating' causation rather than 'initiating' causation. More recently he has sought to defend a volitionist account of agency within an emergentist framework with the help of a distinction between event causation and what he calls 'agent' or 'substance' causation (Lowe 2008). He says that his view 'may also fairly be described as an *emergentist* position, in that it regards the causal powers of persons as *complementing* and *supplementing*—rather than either being reducible to or existing entirely independent of—those of their bodies' (2008, p. 92), and that it is 'a form of emergentism in the philosophy of mind, according to which non-physical mental events and states are causally autonomous and yet are themselves ultimately the products of prior physical evolution' (2008, p. 41). That description seems exactly what the traveller analogy captures. In spite of this similarity, an essential point of difference exists: there is in Indian materialism no suggestion that *subjects* qua individual substances are agent causes.

The new idea is that the physical base P^* of the newly produced mental event M^* has M as an assistive cause. I suggest that what this new proposal does is to combine the idea of assistive causation with the earlier idea about transformation. The point of the proposal, then, is that, as is suggested by the analogy of lighting and maintaining a fire, the earlier emergent mental event contributes to the transformation of the physical base of later emergent mental events. The transformation of the micro-elements is, at later stages in a mental process, partly subject to causal input from mental properties at earlier stages.

This idea is clearly to be distinguished from an idea which Kim attributes to Sperry, that there is what he terms 'synchronic reflexive downward causation' (1999, p. 26), in which a macro feature such as a pain exerts a causal influence on *its own* micro-constituents, the basal neural process such as a C-fibre firing. In the view being articulated here, the only downward causation is diachronic, while the synchronic relation (the 'material causation' relation) is constitutive. The diachronic downward causation introduced is similar to that for which Paul Davies coins the phrase *level-entanglement*, and illustrates as follows:

Consider a computer that controls a microprocessor connected to a robot arm. The arm is free to move in any direction according to the program of the computer. Now imagine a program that instructs the arm to reach inside the computer's own circuitry and rearrange it, for example, by throwing a switch or removing a circuit board. This is software-hardware feedback, where software brings about a change in the very hardware that supports it. (Davies 2006, p. 43)

One way to unpack the idea further is as follows. It is granted that both P and M are sufficient causes of P^* , but the proposal is that P^* does not represent a physical state which is specifiable as basal to M^* in any way *other than as* the base of M^* . Prior mental states assist in the production of later ones by delimiting regions of the physical world along lines and boundaries that the physical world, by itself, does not acknowledge. A sort of context principle is in play: only in the context of the presence of a mental property does it make sense to ask for the region of the physical which that mental property supervenes upon. A transformation account of emergence explains why this should be so, for the physical base P is itself the product of a transformation of the constituent elements consequent upon their constituting the emergence base for preceding mental properties, rather than merely an assemblage of micro-physical entities in a complex pattern.

The proposal insists that mind-mind causation has an explanatory priority over bottom-up physical causation in the specific and restricted sense that the physical causes of mental states can be identified only because of mind-mind interactions. Broad says that the only peculiarity of a trans-ordinal law is ‘that we must wait till we meet with an actual instance of an object of the higher order before we can discover such a law; and that we cannot possibly deduce it beforehand from any combination of laws which we have discovered by observing aggregates of a lower order’ (Broad 1925, p. 79). Its irreducibility consists in our inability to specify the physical base or ‘material cause’ of the higher level property instantiation other than as the basis of the instantiation of that higher level property. The reason for this is that patterns of diachronic assistive causation are responsible for patterns of transformation in the basal conditions.

7. Mind: flame or tornado?

Rephrased in a more contemporary vocabulary, my interpretation of the transformation theory is as follows. The emergence of mental states occurs only within a dynamical system, one whose physical states are in a constant process of flux produced by the fusion of elements and the fusion of their microcausal powers. The microdynamic in this dynamical system jointly specifies the total physical and mental state of the system at any given time, and does so only with reference to both the physical and the mental state of the system at earlier times.²⁶ This must happen in such a way that the initial conditions mention only physical states — that was the point of the analogy with the traveller and the fire. At each moment in time a mental state has a physical realizer on which it supervenes, but this is a purely formal covariance, since there is no way to identify the subvening physical state other than through the description ‘the physical state upon which M supervenes’. Supervenience functions simply as a constraining condition, serving to provide one of the equations that define the system. Because the domain is itself in a state of flux, the microdynamic will have variables taking mental property instantiations, physical property instantiations, *and* micro-entities as values.

²⁶ This enables a role for *systemic causation*: ‘Mental properties emerge because one of the capacities of emergent systems is to help generate new emergent systems. That is, systemic causation involves the creation of stable diachronic patterns (systems distributed over space and time) in which the stability and integrity of such patterns is maintained across constant changes in the micro-base of such systems’ (Silberstein 2006, p. 205).

The model postulates that there are privileged dynamical systems which display *strong* emergence, systems in which the emergent properties are not merely complexity effects. Perhaps, as Chalmers (2006, p. 253) suggests, the weak emergence one finds in normal dynamical systems is sufficient for all natural phenomena except mind. Perhaps there is only one sort of dynamical system with the requisite feature, and that is the embodied mind. Bedau (2003, p. 158) claims that strong emergence is irrelevant to natural science, and that might be exactly why the mind is resistant to natural scientific intelligibility, now seen as an attempt to understand the mind by way of a false analogy with a natural phenomenon, in this case the sort of complex non-linear dynamical system one regularly finds in biology, ecology, and cosmology. The claim, in other words, is that the mathematical apparatus needed to model embodied conscious minds need not be mathematics that finds application anywhere *else* in the natural world. And after all why should it? — there *is* something special about minds. It is the standpoint of a *liberal* naturalism, but one in which the departure from scientific naturalism is the most minimal departure possible.

This is, therefore, the point at which our view parts company with the ‘enactive’ model of Francisco Varela. Varela recommends that we see the relation as one of emergence through self-organization: ‘The aggregates would arise as one moment of emergence, as in a resonating network where strictly speaking there is no all-or-nothing separation between simultaneous (since the emergent pattern itself arises as a whole) and sequential (since for the pattern to arise there must be a back-and-forth activity between participating components)’ (Varela, Thompson, and Rosch 1991, p. 98), adding that ‘in a culture that did not have access to scientific notions of circular causality, feedback/feedforward, and emergent properties, nor to logical formulations for handling self-reference, the only recourse for expressing an emergent may have been to say that a process is both cause and effect’ (1991, p. 119). This is a dynamical systems model with weak emergence and, seemingly, synchronic downward causation.²⁷ I have, of course, been at pains to show that the culture had full access to a good range of relevant concepts. Varela’s neurophenomenological project aims for the naturalization of all aspects of mind using a type of dynamical systems model common to the description of the behaviour of

²⁷ See also Thompson 2007, Appendix B, for a spirited attempt to side-step the exclusion problem within a framework of weak emergence.

physical systems. The consequence of what I have said is that such a project is untenable.

The account I am recommending also differs from an analysis of emergence recently suggested by Sidney Shoemaker (2002, 2007). Shoemaker develops a line of thought, which indeed he claims to find in C. D. Broad, in terms of a notion of ‘micro-latent’ causal powers. A micro-latent power is defined as follows:

The component entities have powers that, collectively, determine the instantiation of the emergent property when they are combined in an emergence-engendering way. But these being cases of emergence, these cannot all be powers that manifest themselves when the components are not combined in emergence-engendering ways. Some of them must be ‘latent’ powers. Or, since these powers do not remain latent when their possessors are combined in emergence-engendering ways, let us speak of them as ‘micro-latent’ powers. We can contrast these with the ‘micro-manifest’ powers which these same entities manifest when they are not combined with other entities at all, or are configured in ways that are not emergence-engendering. (Shoemaker 2007, p. 73)

Shoemaker claims that emergence should be understood in terms of the existence of micro-latent causal powers which manifest themselves when the elements are combined in what he calls ‘emergence engendering’ ways, thereby giving rise to a ‘micro-emergent state of affairs’. He argues that the distinction between micro-latent and micro-manifest powers is sufficient to solve the exclusion problem for downward causation:

Supposing that micro-entities have micro-latent powers, when a group of micro-entities that are among the constituents of a macro-entity are configured in an emergence-engendering way there will be one micro-physical state of affairs consisting of these particles being configured as they are and having the micro-manifest powers they have, and another micro state of affairs consisting in all of this plus their having the micro-latent powers they have. The first micro state of affairs, which can be called a micro-physical state of affairs, will be a part of the second, which can be called a micro-emergent state of affairs. It will be the latter that has the causal clout required for downward causation. And it will be the latter that is the instantiation of the one micro-structural property, an emergent one, that the macro-entity has in virtue of certain of its micro-constituents being propertied and related as these micro-entities are. (Shoemaker 2002, p. 63)

Such a proposal does indeed give sense to a notion of transformation: we might say that the transformation of the elements consists in

the *activation* of their micro-latent powers. Shoemaker, I believe, thereby does articulate the notion of transformation which the British emergentists Broad and Alexander had in mind in their use of the term. However, the two accounts of transformation differ with respect to the assumption that micro-physical states of affairs persist and co-exist in emergence-engendering circumstances with emergent micro-properties. Shoemaker's is a description of emergence for which an appropriate analogy would be the cyclone or tornado.²⁸ The alternative view is characterized by a rather different thought. The idea is instead that the whole dynamical system, including the micro-elements themselves, is in a process of mutation, and that emergent macrostates are consequent upon this process of mutation. An analogy better than that of the tornado, therefore, is the flame, pictured as something emergent from a process in which the constituent material is itself continuously changing, and in ways causally determined by emergent macrostates. It would be a mistake to describe this view only in terms of particles shifting their patterns of aggregation or their 'emergence-engendering' distribution.²⁹

In this respect transformation theory approaches the 'fusion emergence' account of Paul Humphreys. Humphreys defines a fusion operator for property instances, where 'By a fusion operation, I mean a real physical operation' (1997a, p. 10). In his 2000, he extends the treatment to fused objects, such as the overlapping wave-functions of two electrons in close proximity, and offers a partial analogy: 'Consider poker chips in a casino. The basic units are red chips, and as soon as you have accumulated two red chips you can trade them in for a blue chip that is worth two units. The blue chip is not composed of two red chips and you cannot count its two components because it does not have any, but it behaves exactly as if there were two such units present' (2000, p. 28–9). Yet the transformation theory diverges from that account too, in that its ambition is to preserve supervenience within a generative model of emergence. Humphreys is explicit

²⁸ Thus Bedau 1997, p. 375:

Some [examples of emergence] involve inanimate matter; e.g. a tornado is a self-organizing entity caught up in a global pattern of behaviour that seems to be autonomous with respect to the massive aggregation of air and water molecules which constitute it. Another source of examples is the mind: our mental life consists of an autonomous, coherent flow of mental states (beliefs, desires, etc.) that presumably somehow ultimately arise out of the swarm of biochemical activity among our brains's neurons.

²⁹ Shoemaker's attempt to argue that a causal-powers metaphysics is consistent with non-reductive psycho-physical realization is criticized in O'Connor and Churchill 2010.

that the technical notion of ‘fusion’ he introduces cannot sustain supervenience (Humphreys 1996; 1997b). His fusion operation applies to property instances, and is characterized by what Wong (2006) aptly describes as ‘basal loss’:

What is most distinctive in fusion emergentism is Humphreys’s property fusion operation, which takes property instances (at the i th level) and generates an emergent property instance (at the $i+1$ st level) with novel causal powers. When property instances at the generating i th level are fused, the individual property instances are destroyed and are nonindividuable within the emergent fusion existing at the $i+1$ st level. Call this the *basal loss* feature of fusion emergentism. (Wong 2006, p. 345)

This feature of the technical operation Humphreys calls ‘fusion’ is essential to his strategy for negotiating the exclusion argument, for it is the *necessitated* destruction of the base in fusion which for him blocks overdetermination. Transformation emergence, however, is distinguished from fusion emergence in its technical guise in wishing to endorse what is the majority view, that emergence is a supervenience-based relation, and in therefore considering that this formal notion of ‘fusion’ cannot correctly describe the relationship of metaphysical dependence involved in emergence. More precisely, there is no commitment to the *necessity* of ‘basal loss’ in the transformation theory, although it will sometimes, perhaps often, be present as a contingent feature of emergence-engendering dynamical systems. Fusion, in a transformation theory, is not a function that can be defined in advance, but is rather something that is ‘solved for’ in the dynamical system, when that system is subject to the general constraint afforded by supervenience.³⁰

Materialist theories of mind, ancient as well as modern, seek to do justice to two compelling but apparently incompatible scruples. One is that ours is a physical world, everything happening within it open to physical explanation. The other is that mindedness is a matter of causal significance, that it makes a causal difference that there are minds. The more we feel the pull of one of these scruples, the more mysterious becomes the other. A robust commitment to physicalism leaves the mind looking like an epiphenomenal by-product of natural processes, a causally inert shadow. But a view of the mind as possessing aetiological autonomy threatens to re-enchant the physical world with supernatural causes and effects. The attraction of emergentism is

³⁰ For an extended critical discussion of Humphreys, particularly his understanding of the demands imposed by causal overdetermination, see Wong 2006.

that it offers a way to escape the dilemma. An emergentist tries to prise free the soundly motivated scruples about the dependence and autonomy of the mental from too-rigid theory, to see the problems as symptoms of the fact that an insight has been poorly encoded in doctrine. Such is precisely, I have argued, how the philosophy of Bṛhaspati came to be seen. Later philosophers feel the tension that is created between the demands of dependence and autonomy, and seek proposals that help one to see a way for these demands to be compatible. What I have sought to bring to view is the *general form* that such a proposal will take.

Without wishing to diminish the points of contact that certainly exist between transformation theory and the accounts of Varela, Shoemaker, and Humphreys, I nevertheless want to affirm that it is a distinct theory with distinctive and important virtues. In particular, I believe that it makes clear what the key issue for emergence really is. The metaphor of the traveller and the fire leads to a conception of body as a dynamical system, flame-like in its mode of persistence, and fully able to sustain an intertwined mental life, as long as there is an initial 'spark' of mindedness. As soon as we have located this maximally material and minimally mental 'spark', the model of assistive causation shows how to move to an account of an embodied stream of consciousness. Then, by a second application of the same account, we are led to a conception of a stream of thought as a dynamical system, flame-like in *its* mode of persistence, and fully able to sustain an intertwined first-person life, as long as there is an initial 'spark' of self-awareness. Once we have located this instigator of minimal self-awareness, the model of assistive causation shows how to move to an account of an embodied and enstreamed self. Further recursion on the model generates higher-order first-person thought, and arguably at some point must settle on moments of reflexive self-awareness, which serve as attractors in the dynamical system. The entire description, I must stress, is a purely formal model of self-consciousness and not an account of the actual psycho-biological genesis of mind. The question is: does anything correspond in fact to the initial 'sparks' of mindedness and self-awareness which the formal model posits? Rather than either of Kim's, I suggest that *this* is the key issue upon which the prospects for emergentism rest.³¹

³¹ Many thanks are due to Christian Coseru, Richard Sorabji, Jonathan Lowe, Amber Carpenter, the audience of the Pranab Kumar Sen Memorial Lecture (Kolkata July 2009), an audience at a 2009 meeting of the Durham Philosophy Society, the editor of *Mind*, and an anonymous referee.

References

- Alexander, Samuel 1920: *Space, Time, and Deity: the Gifford Lectures at Glasgow, 1916–1918*. London: Macmillan, 2 vols.
- al-Faḍl, Abu 1873–1907: *Aīn-i Akbarī*. In H. Blochmann (ed.) H. *The Aīn-i Akbarī*. Calcutta: Asiatic Society of Bengal, 3 vols.
- Beckermann, Ansgar, Hans Flohr, and Jaegwon Kim (eds) 1992: *Emergence or Reduction?* Berlin and New York: de Gruyter.
- Bedau, Mark 1997: 'Weak Emergence'. *Philosophical Perspectives*, vol. 11: 'Mind, Causation and World', pp. 375–99.
- 2003: 'Downward Causation and Autonomy in Weak Emergence'. *Principia Revista Internacional de Epistemologica*, 6; rptd in Bedau and Humphreys (eds) 2008, pp. 155–88.
- Bedau, Mark and Paul Humphreys (eds) 2008: *Emergence: Contemporary Readings in Philosophy and Science*. Cambridge, MA: MIT Press.
- Bhattacharya, Ramkrishna 2002: 'Cārvāka Fragments: A New Collection'. *Journal of Indian Philosophy*, 30.6, pp. 597–640.
- Broad, C. D. 1925: *The Mind and its Place in Nature*. London: Routledge and Kegan Paul.
- Burnyeat, Myles 1992: 'Is an Aristotelian Philosophy of Mind Still Credible?' In Nussbaum and Rorty 1992, pp. 15–26.
- Cakradhara. 1982: *Granthibhaṅga*. In Gaurinath Shastri (ed.): *Nyāyamañjarī with the Commentary 'Granthibhaṅga' by Cakradhara*. Varanasi: Sampurnananda Sanskrit University, 3 vols.
- Cardona, George 1967–68: 'Anvaya and vyatireka in Indian Grammar'. *The Adyar Library Bulletin*, 31–2, pp. 313–52.
- 1981: 'On Reasoning from *anvaya* and *vyatireka* in Early Advaita'. In Malvania and Shah 1981, pp. 79–104.
- Carrier, Martin, Gerald J. Massey, and Laura Ruetsche (eds) 2000: *Science at Century's End*. Pittsburgh: University of Pittsburgh Press.
- Caston, Victor 1992: 'Aristotle and Supervenience'. *Southern Journal of Philosophy*, 31, Supplement, pp. 107–35.
- 1997: 'Epiphenomenalisms, Ancient and Modern'. *Philosophical Review*, 106.3, pp. 309–63.
- 2001: 'Dicaearchus' Philosophy of Mind'. In Fortenbaugh and Schütrumpf 2001, pp. 175–93.
- Chalmers, David 1996: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

- 2006: 'Strong and Weak Emergence'. In Clayton and Davies 2006, pp. 244–54.
- Cicero, Marcus Tullius 1927: *Tusculan Disputations*, trans. J. E. King, Loeb Classical Library, 141. Cambridge, MA: Harvard University Press.
- Clayton, Philip and Paul Davies (eds) 2006: *The Re-Emergence of Emergence*. Oxford: Oxford University Press.
- Colebrooke, H. T. 1837: *Miscellaneous Essays*. London: W. H. Allen, 2 vols.
- Cowell, E. B. 1862: 'The Cārvāka System of Philosophy'. *Journal of the Asiatic Society of Bengal*, 31, pp. 317–90.
- Crane, Timothy 1999: 'The Significance of Emergence'. In Gillett and Loewer 1999, pp. 207–24.
- Davidson, Donald 1980: *Essays on Actions and Events*. Oxford: Clarendon Press.
- Davies, Paul 2006: 'The Physics of Downward Causation'. In Clayton and Davies 2006, pp. 35–51.
- Dretske, Fred 1988: *Explaining Behaviour: Reasons in a World of Causes*. Cambridge, MA: MIT Press.
- Fortenbaugh, William W. and Echart Schütrumpf (eds) 2001: *Dicaearchus of Messana: Text, Translation, and Discussion*. New Brunswick, NJ: Transaction Publishers.
- Galen 1996: *On the Elements According to Hippocrates*, trans. Phillip De Lacy. Berlin: Akademie Verlag.
- Gautama 1997: *Nyāya-sūtra*. In Anantalal Thakur (ed.): *Gautamīya-nyāya-darśana with Bhāṣya of Vātsyāyana*. Delhi: Indian Council of Philosophical Research.
- Gillett, Carl 2006: 'Samuel Alexander's Emergentism: Or, Higher Causation for Physicalists'. *Synthese*, 153.2, pp. 261–96.
- Gillett, Carl and Barry Loewer (eds) 1999: *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.
- Humphreys, Paul 1996: 'Aspects of Emergence'. *Philosophical Topics*, 24.1, pp. 53–70.
- 1997a: 'How Properties Emerge'. *Philosophy of Science*, 64, pp. 1–17.
- 1997b: 'Emergence, Not Supervenience'. *Philosophy of Science*, 64, pp. 337–45.
- 2000: 'Extending Ourselves'. In Carrier, Massey, and Ruetsche 2000, pp. 13–32.
- Jayanta 1982: *Nyāyamañjarī*. In Gaurinath Shastri (ed.): *Nyāyamañjarī with the Commentary 'Granthibhaṅga' by*

- Cakradhara*. Varanasi: Sampurnananda Sanskrit University, 3 vols.
- Kamalaśīla 1968: *Tattva-saṅgraha-panjikā*. Swami Dwarikadas Shastri (ed.). Varanasi: Bauddhabharati.
- Kim, Jaegwon 1998: *Mind in a Physical World: An Essay on the Mind-body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- 1999: ‘Making Sense of Emergence’. *Philosophical Studies*, 95, pp. 3–36; rptd in Bedau and Humphreys 2008, pp. 127–54.
- 2006a: ‘Emergence: Core Ideas and Issues’. *Synthese*, 151.3, pp. 547–59.
- 2006b: ‘Being Realistic about Emergence’. In Clayton and Davies 2006, pp. 189–202.
- Lowe, E. J. 1996: *Subjects of Experience*. Cambridge: Cambridge University Press.
- 2008: *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.
- Macdonald, Cynthia and Graham Macdonald 2010a: ‘Emergence and Downward Causation’. In Macdonald and Macdonald 2010b, pp. 139–68.
- Macdonald, Cynthia and Graham Macdonald (eds) 2010b: *Emergence in Mind*. Oxford: Oxford University Press.
- McLaughlin, Brian 1992: ‘The Rise and Fall of British Emergentism’. In Bedau and Humphreys 2008, pp. 19–60. Originally published in Beckerman, Flohr, and Kim 1992.
- 1997: ‘Emergence and Supervenience’. In Bedau and Humphreys 2008, pp. 81–98. Originally published in *Intellectica*, 25, 1997.
- Malvania, D. and N. J. Shah (eds) 1981: *Studies in Indian Philosophy*. Ahmedabad: L. D. Institute of Indology.
- Mill, John Stuart 1843: *A System of Logic, Ratiocinative and Inductive*. London: John W. Parker, 2 vols.
- Morgan, C. Lloyd 1923: *Emergent Evolution*. London: Williams and Norgate.
- Muir, John 1862: ‘Verses from the *Sarva-darśana-saṅgraha*, the *Vishnu Purāna*, and the *Rāmāyana*, illustrating the tenets of the Chārvaṅkas, or Indian Materialists, with some Remarks on Freedom of Speculation in Ancient India’. *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, 19, pp. 299–314. (Read to the Society on Saturday, 14th December, 1861.).
- Nagel, Thomas 1979: *Mortal Questions*. Cambridge: Cambridge University Press.

- Noordhof, Paul 2010: 'Emergent Causation and Property Causation'. In Macdonald and Macdonald 2010b, pp. 69–99.
- Nussbaum, Martha and Amélie Rorty (eds) 1992: *Essays on Aristotle's De Anima*. Oxford: Clarendon Press.
- O'Connor, Timothy 1994: 'Emergent Properties'. *American Philosophical Quarterly*, 31, pp. 91–104.
- 2000: *Persons and Causes*. Oxford: Oxford University Press.
- O'Connor, Timothy and John Ross Churchill 2010: 'Is Non-reductive Physicalism Viable within a Causal Powers Metaphysic?' In Macdonald and Macdonald 2010b, pp. 43–60.
- O'Connor, Timothy and Hong Yu Wong 2005: 'The Metaphysics of Emergence'. *Nouûs*, 39, pp. 658–78.
- Plutarch 1967: 'Against Colotes', in *Moralia*, vol. xiv, trans. B. Einarson, Loeb Classical Library, 428. Cambridge, MA: Harvard University Press.
- Prabhācandra 1990: *Prameya-kamala-mārtaṇḍa*. Nyaya Shastri Mahendrakumar (ed.). Delhi: Sri Satguru Publications. 3rd edn.
- 1991: *Nyāya-kumuda-candra*. Nyaya Shastri Mahendrakumar (ed.). Delhi: Sri Satguru Publications, 2 vols. 2nd edn.
- Rāmatīrtha 1911: *Vidvamanorañjanī*. In G. A. Jacob (ed.), *The Vedānta-sāra of Sadānanda together with the commentaries of Nṛsiṃhā-sarasvatī and Rāmatīrtha*. Bombay: Tukaram Jayaji.
- Reydams-Schils, Gretchen (ed.) 2003: *Plato's Timaeus as Cultural Icon*. Indiana: Notre Dame Press.
- Rueger, Alexander 2000a: 'Physical Emergence, Diachronic and Synchronic'. *Synthese*, 124, pp. 297–322.
- 2000b: 'Robust Supervenience and Emergence'. *Philosophy of Science*, 67, pp. 466–89.
- Śaṅkara 1917: *Brahma-sūtra-bhāṣya*. In N. A. K. Sastri and V. L. Sastri Pansikar (eds) 1917: *The Brahma-sūtra-bhāṣya with the Commentaries Bhāmatī, Kalpatarū and Parīmāla*. Bombay: Nirmaya Sagar Press.
- Searle, John 1992: *The Rediscovery of the Mind*. Cambridge, MA: The MIT Press.
- Sextus Empiricus 1935: 'Adversus Mathematicos VII', in *Sextus Empiricus II*, trans. R. G. Bury, Loeb Classical Library, 291. Cambridge, MA: Harvard University Press.
- Shoemaker, Sydney 2002: 'Kim on Emergence'. *Philosophical Studies*, 108, pp. 53–63.
- 2007: *Physical Realization*. Oxford: Oxford University Press.

- Silberstein, Michael 1998: 'Emergence and the Mind-body Problem'. *Journal of Consciousness Studies*, 5.4, pp. 464–82.
- 2006: 'In Defence of Ontological Emergence and Mental Causation'. In Clayton and Davies 2006, pp. 203–26.
- Silberstein, Michael and McGeever John 1999: 'The Search for Ontological Emergence'. *The Philosophical Quarterly*, 49, pp. 182–200.
- Sorabji, Richard 2003: 'The Mind-body Relation in the Wake of Plato's *Timaeus*'. In Reydam-Schils 2003, pp. 152–62.
- 2005: *The Philosophy of the Commentators, 200–600 AD*. London: Duckworth, 3 vols.
- 2010: *Philoponus and the Rejection of Aristotelian Science*. London: Institute of Classical Studies. Revised 2nd edn.
- Strawson, Galen 2006: 'Being Realistic: Why Physicalism Entails Panpsychism'. *Journal of Consciousness Studies*, 13.10–11, pp. 3–31.
- Strawson, Peter F. 1963: *Individuals*. New York: Anchor Books.
- Thompson, Evan 2007: *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Belknap Press.
- Uddyotakara 1997: *Nyāya-vārttika*. Anantalal Thakur (ed.). Delhi: Indian Council of Philosophical Research.
- Vācaspati, Mīśra 1980: *Bhāmatī*. In Jagdish Lal Shastri (ed.), *Brahma-sūtra-śāṅkara-bhāṣyam*. Delhi: Motilal Banarsidass.
- 1996: *Nyāya-vārttika-tātparyatikā*. Anantalal Thakur (ed.). Delhi: Indian Council of Philosophical Research.
- Van Cleve, James 1990: 'Mind-Dust or Magic? Panpsychism Versus Emergence'. *Philosophical Perspectives*, vol. 4: 'Action Theory and Philosophy of Mind', pp. 215–26.
- Varela, Francisco, Evan Thompson, and Eleanor Rosch 1991: *Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- Wilson, Jessica 2005: 'Supervenience-based Formulations of Physicalism'. *Nous*, 39.3, pp. 426–59.
- Wong, Hong Yu 2006: 'Emergents from Fusion'. *Philosophy of Science*, 73, pp. 345–67.

Appendix:

Authors and their dates

- Dicaearchus (350–285 BCE) Aristotelian epiphenomenalist
 Aristoxenus (fl. 335 BCE) Aristotelian epiphenomenalist

Galen (129–216 CE) Physician/Philosopher
Alexander (c. 200 CE) Aristotelian commentator
Bṛhaspati (date unknown³²) Indian Materialist

Śaṅkara (c. 710) Vedāntin
Kamalaśīla (c. 740–795) Buddhist
Bhaṭṭa Udbhata (c. 800) Materialist/Nyāya
Jayanta (c. 870) Nyāya
Vācaspati (c. 960) Polymath/Nyāya
Prabhācandra (980–1065) Jaina

H. T. Colebrooke (1765–1837); J. S. Mill (1806–1873); G. H. Lewes
(1817–1878); C. Lloyd Morgan (1852–1936); Samuel Alexander (1859–
1938); C. D. Broad (1887–1971)

³² The first known reference to Bṛhaspati is from the sixth century. He composed his text in the sūtra style, and the usual period of production for texts in that style is thought to be the interval between 200 BCE and 200 CE. Such texts were often distillations of ideas already in circulation. There are formal resemblances between Bṛhaspati's text and other sūtra texts whose dates have been more precisely ascertained; in the *Nyāyasūtra*, for example, one finds the sentence 'The senses are from the material elements (*bhūtebhyah*)'. It is reasonable to speculate, therefore, that Bṛhaspati is no later than 200 CE. For further speculation, see Muir 1862, Cowell 1862.