Emergence in Mind

Edited by
CYNTHIA MACDONALD AND
GRAHAM MACDONALD

OXFORD
UNIVERSITY PRESS

Contents

1

Introduction*

Cynthia Macdonald and Graham Macdonald

1. SOME HISTORICAL BACKGROUND

The issue of whether there are emergent properties, and if so, what their relation to other properties is, has been much debated over the past two centuries. One critical focus of these debates is on the very intelligibility of such 'emergence': critics of the notion are suspicious that emergent properties inhabit a kind of halfway house that is, by its nature, unstable, given that such properties are characterized by two features that are in tension with one another. On the one hand, an emergent property is said to be distinct from the properties from which it emerges; on the other hand, it is said to be dependent on those very properties. This combination of features creates the instability that threatens either to collapse emergent properties into their 'base' ones (the reductionist option) or to make them so different from their base properties that their relation to those properties is left mysterious or non-existent (the dualist option).

As a consequence of this tension, there are discernible fluctuations in the fortunes of emergence over different periods in the history of science. Emergence-style arguments go back at least as far as the time when psychophysical dualism was being promoted, but it was in the nineteenth and early twentieth centuries that doctrines bearing a resemblance to what we now call emergentism were most hotly disputed in various, mainly physiological and biological-medical, contexts. During this time it appeared to many that physics was not, and would not be able to explain chemical bonding, vital activity, or mental processes. Various 'additions' were thought necessary if the gap left by physical and chemical explanations was to be filled. One was the vitalist option mooted initially by Xavier Bichat, who thought that in order to explain life it was necessary to introduce 'sensibility' and 'contractility' as properties of tissues,

* We are indebted to an anonymous reader at Oxford University Press for comments and advice. Work on this volume and on the conference from which this volume originated was generously supported by a Mind Association Major Conference Grant, a Marsden Grant awarded by the Royal Society of New Zealand to Cynthia Macdonald and Graham Macdonald, and funding from the School of Politics, International Studies, and Philosophy, Queen's University Belfast.

these properties themselves not resulting from the combination of only physical or chemical properties. The processes of metabolism and fermentation were thought to require the activity of these 'vital' properties (for a brief history, see Bechtel and Richardson 1998). Another option was to invoke non-physical entities to do the explanatory work, such as Driesch's 'entelechies' which were introduced in order to explain the difference between organic and inorganic matter. It was questionable, however, whether the non-physical elements were genuinely explanatory, and what the nature of their relation to the physical was. Given their 'categorial' difference from the physical (Driesch conceived of 'entelechy' as essentially immaterial), it was difficult to see how they could be suitable candidates for interacting causally with the physical.

Another proposal did away with special entities altogether in favour of new forces: it was thought that at a certain level of complexity of physical organization, novel forces 'emerged'. The 'newness' of the forces was due to their being more than mere sums of the forces of the physical elements making up the complex structure. At this level of complexity, so it was claimed, some force that was not just the resultant of the antecedently given forces, the lower level forces, was discernible. As a consequence, not all of the forces operating in the world were physical; from the complexity of physical structures vital, mental, and social forces emerged, these novel forces enabling the explanation of the non-physical features of the world. The general doctrine was known as emergentism. With it there developed a hierarchical conception of the relation between the sciences, with physics being ontologically fundamental and the rest of the sciences stacking up in layers above it, each having its own laws describing the new forces arising at the relevant of complexity.[1]

This tendency to postulate supra-physical (and chemical) properties, entities, or forces did not go unopposed. In addition to generating some philosophical unease about the idea of independent causal forces operating 'downwards', threatening the pattern of purely physical causation of physical changes, the emergentist programme suffered blows from some major empirical discoveries, especially in the fast-developing field of biology. In the second half of the nineteenth century Darwinian accounts of design and speciation prompted many biologists to seek a unification of their theories with those of experimental physiologists. Particularly influential was the group of 'medical materialists' formed in Berlin in the 1840s under the leadership of Hermann von Helmholtz, who pleaded for the introduction of physico-chemical methods into biology.[2] The approach of the medical materialists was explicitly reductionist: they attempted to study the organism by studying isolated parts of it, using the experimental methods of physics and chemistry to reveal the underlying mechanisms of biological processes. Amongst those who were influenced by their work were Wilhelm Roux (via an

[1] Emergentism in Britain is illuminatingly described in Brian McLaughlin 1992.
[2] The relevant history here is derived from Garland Allen 1975.

embryologist Wilhelm Preyer, who had studied in Berlin), Jaques Loeb, and, later, Ivan Pavlov (who was introduced to experimental work on reflexes by a student of the Berlin school, Ivan Sechenov). Roux was an influential embryological experimentalist and also a persuasive propagandist for a programme he called *Entwicklungsmechanik*, roughly translated as 'developmental mechanics' (see Allen 1975: 33). According to this, the study of developing embryos is directed toward explaining how cell differentiation is caused by the internal physical and chemical constitution of the embryo.

Loeb, having imbibed much of the medical materialist philosophy at the University of Strasbourg, moved to the University of Würzburg where he was struck by the work of Julius Sachs on plant tropisms, automatic responses of organisms to specific circumstances, such as light. Loeb was convinced that the physico-chemical explanation applied to tropistic phenomenon could be applied more widely to life processes in general. He experimented on unfertilized sea urchin eggs, producing developmental changes in them by altering their chemical environment. His experiments on artificial parthenogenesis opened up the possibility of laboratory-created life, reinforcing Loeb's outlook that *all* living processes (including mental and social) could be explained as resulting from 'chemical mechanisms'. This determinist and mechanistic outlook was further strengthened by Pavlov's work on conditioned reflexes, which promised to show how learned behaviour could be explained as the result of prior conditioning, which in turn could be accounted for in neurophysiological terms.

It was not so much that any crucial experiments disproved the postulation of immaterial entities or emergent forces, but that a climate of reductionist optimism was fortified. The new developments earned respect because the theories produced were testable, at times mathematically expressible, and gave rise to detailed experimental work both on cells and supra-cellular tissues. Philosophers and scientists responded to these developments by retaining the hierarchical conception of science, with reduction replacing emergentism as the favoured relation between the levels. Arguably, this picture has been further supported by twentieth-century developments in molecular biology, with the discovery of the nature of genes, how they are transmitted, and how they are expressed in the course of the development of an individual. The use of chemical theory in all these developments has been crucial, suggesting that biology was reducible to chemistry and thereby to physics, given that the reducibility of chemistry to physics was thought to have been demonstrated by the physical explanation of chemical bonding.

The major trend in all of this scientific work was to explain processes at the macro-level by discovering more of the detail of microprocesses. Reductionism looked to be an eminently suitable research strategy. The appearance of emergent elements was to be explained away, and it was the reductionist's prediction that our increasing knowledge of physical processes would obviate the need for any special entities or forces. The difficulty of downward causation was avoided, since

reduction would place all such causation at the level of the physical. The problem of the causal powers of the higher-level properties was solved: the reductionist's picture would endorse their causal efficacy, but would do so at a cost, robbing them of any causal autonomy. For, according to the reductionist, any higher-level property that has causal powers has them because it is, really, a physical property. Physics is fundamental, where 'fundamental' means that the physical is causally, ontologically, and explanatorily all-encompassing.

However, other developments within the favoured science, physics, supported a somewhat different view. Quantum mechanics deals with the ultimate microparticles, but not everyone is convinced that all explanations in quantum mechanics fit the reductionist mould: not all quantum events, it appears, can be explained by citing properties of subatomic particles acting independently of one another. It has been suggested that entangled states in quantum mechanical systems are ontologically emergent, as 'the individual states before interaction do not determine the joint state after the interaction, whereas the joint state does determine the individual states. This feature makes compositional accounts of the joint system implausible.' (Humphreys 2008: 586). The example is important in that it shows that debates about emergence are not necessarily confined to interdisciplinary contexts (biology in relation to physics and chemistry, psychology in relation to biochemistry, and so on); issues to do with emergence can arise within a single domain, provided that the domain is partitioned into a 'substrate' and a level that putatively emerges from the substrate. As Humphreys puts it, emergence is essentially a relational phenomenon: a property is emergent only in relation to another set of properties from which it can be said to have emerged, and this relation can exist in both intra- and interdisciplinary contexts.

Apart from the controversies within quantum mechanics, there were other developments that supported an emergentist view, particularly ones connected with considerations about complex phenomena. In an important article, P. W. Anderson questioned what many assumed to be a corollary of the reductionist hypothesis that our minds and all animate and inaminate matter are controlled by the same set of fundamental laws. The putative corollary was that 'the only scientists who are studying anything really fundamental are those who are working on those laws' (Anderson 1972: 393). Against this, Anderson argued that understanding more complex phenomena may require new laws and concepts, so that psychology need not be mere applied biology, nor biology mere applied chemistry. And again, even in the physical domain, processes resulting in phase transitions, such as a solid becoming a liquid, require new properties to emerge, and so need to be understood at their own level. Anderson's message is conveyed by the title of his article: 'More is Different'. In some cases, he says, we can see 'how the whole becomes not only more than but very different from the sum of its parts' (Anderson 1972: 395).

But caution is needed: the type of emergence exemplified by phase transitions, or, to take another example of a supposedly 'emergent' phenomenon, by

termite organization (see Johnson 2001), does not seem to raise any significant philosophical problems. These are cases that have been 'tamed' by science itself, in its ability to explain how the transitions are effected, or how termites can follow an individual path *thereby* contributing to the organization of the larger colony. We are disinclined to enter into any argument about whether these cases are ones of 'real' emergence, or just apparent emergence; such an argument would presume that there is a widely understood, univocal sense of 'emergence', and that all that is in dispute is where and when it applies. This would be an oversimplification of complex debates. It is better to characterize the different types of relations that are involved in putative cases of emergence, noting where they differ and whether significant problems arise, given our present state of scientific knowledge. In what follows we shall set out in fairly broad terms different aspects of what could be called an emergence-relation, situating our discussion within recent developments in the philosophy of mind.

2. THE PHILOSOPHICAL CONTEXT

Anderson's 1972 paper was sandwiched between two seminal philosophical articles, those by Donald Davidson (Davidson 1970) and by Jerry Fodor (Fodor 1974), that set the agenda for debates concerning the reducibility of the mental and, more generally, the autonomy of the special sciences.[3] The context was one in which there was increasing dissatisfaction with what were thought to be only three possible views about the nature of mind: dispositionalism (e.g., forms of logical behaviourism), mind–body dualism, and mind–body type identity. Against the background of mounting pressure to 'naturalize' the mind, the identity option was favoured. However, its commitment to psychophysical type-type identities looked implausible; mental types, it was claimed, are 'multiply realized' by physical types, a single psychological property being realizable by different (sets of) physical properties, this prohibiting the type-type identities required for any reduction. This set the agenda: any solution to the mind–body problem must respect multiple realizability while satisfying the naturalist constraint that the mind be shown to be part of the natural (or more specifically, the physical) world. In different ways Davidson and Fodor proposed to satisfy the naturalist constraint by identifying the domain of the mental with that of the physical at the level of individual, or token, events while accommodating the possibility of multiple realization (or, more generally, non-reduction) at the level of mental properties or types. Fodor's argument was the more empirical argument. In the model he presented of the relation between higher-level special sciences

[3] Putnam's paper 'Psychological Predicates' (Putnam 1967) proved to be influential as well, but it did not have the immediate impact of those of Davidson and Fodor, partly due to being published in a less accessible volume.

and lower-level physics, the taxonomy imposed on physical events by physical theory was not required to align neatly with the taxonomies used by higher-level sciences; the same events could be subject to various taxonomies. The identity of the higher-level events with the lower-level (physical) events satisfied the physicalist requirement that all individual, or token, events be physical events, while the variation in taxonomies ensured non-reducibility. One result of this, as Fodor saw it, was that the higher-level explanations would require appeal to *ceteris paribus* laws, laws that allow for exceptions. The argument was 'empirical' because there was no way of knowing in advance of investigation whether a particular upper-level science would or would not be reducible. It could just happen that, given the results of such investigations, the most plausible conclusion to draw would be that no reduction was forthcoming: there could be just simple taxonomic divergence, this divergence corresponding to the differing interests of investigators. The proposed disunity of science was just a 'working hypothesis'.

Fodor's argument is clearly consistent with Anderson's view that different levels of reality may require new laws and concepts, and it was intended to apply quite generally to non-physical sciences (the 'special sciences'). One example used by Fodor is illustrative of higher-level multiple realizablity, that of the implausibility of supposing that descriptions of transactions involving money could ever be reduced to physical laws, given the diversity of (physical) objects that could serve as money and the variety of actions that could realize economic transactions. Davidson's argument for token identity was restricted to the psychophysical case, and so could make use of features of the mental not present in the subject matter of other special sciences. This restriction enabled Davidson to put forward an a priori argument, one making essential use of the fact of causal interaction between mental and physical events, the nomological character of causality (events causally related must be covered by a strict law), and the anomalous character of the mental (mental predicates being unfit to serve in strict laws). The argument for the anomalous nature of the mental also established non-reducibility, this being claimed to follow from the constraints that rationality imposes on the appropriateness of describing, for example, actions using mental descriptions, these constraints being inapplicable to the appropriateness of describing those same events in physical terms.

Despite these differences between the arguments of Fodor and Davidson, the resulting metaphysical picture was broadly the same: a non-reductive monism which embraced token identities between higher-level and lower-level (physical) events but which permitted re-descriptions of those events, and so explanatory autonomy, to higher-level disciplines. Explanatory autonomy could be achieved either by using essentially *ceteris paribus* laws (Fodor) or by employing a conceptual framework inimical to any psychophysical (or psychological) laws (Davidson). The autonomy was not absolute, however; naturalistic scruples would not permit free-floating higher-level descriptions. There had to be some connection to the physical level, however loose, and this was provided by

construing the higher-level/lower-level domains as standing in a relation of supervenience: there could be no change in a higher-level phenomenon without a change at the lower (physical) level. Supervenience allowed for multiple realizability while satisfying the naturalist thought that the physical was 'basic and general'. It looked as though non-reductive monism could allow the naturalistically minded philosopher to have his or her cake and eat it, to have explanatory freedom with ontological respectability.

3. THE DEBATES

Harmony, such as it was, did not last long. In what follows we provide an overview of some of the problems raised, with an eye to the issues that connect the papers in this volume.

A. Structured Events and Causation

The metaphysical picture offered by Davidson was one that is suspicious of properties; a picture of 'structureless' events bearing or satisfying different descriptions, physical and mental. These event-descriptions qualified as physical or mental if they participated in the vocabularies distinctive of their subject matters. So, in brief, Davidson claimed that to be a mental description is to participate in a vocabulary whose application conditions are answerable to norms of rationality; to be a physical description is to participate in a vocabulary governed by what may be termed the norms of nomologicality. Physical events are those events truly described using the physical vocabulary; mental events are those truly described using the mental vocabulary. Given that physical events unproblematically cause their effects, so, too, do those selfsame events cause their effects even if they are mentally described (and so are mental events). This much is guaranteed by the extensionality of the causal relation: if *A* causes *B* then no matter how *A* is described, whether the vocabulary is mental or physical, *A* 'under that description' still causes *B*.

Many were unhappy with what they saw as too quick a fix; it seemed to make mental events causally efficacious only '*qua* physical' (i.e., only insofar as they have a physical description), a suspicion reinforced by Davidson's insistence that only physical descriptions could be used in the formulation of causal laws. Foremost amongst those unhappy with this 'solution' was Jaegwon Kim, who provided a metaphysical 'structure' to the events: an event was taken to be the exemplification of a property in an object at a time. In the simple (monadic) case, Kim represented it schematically by an expression of the form '[x,P,t]', this being construed as a singular term referring to an event, where x is the object in which the property is exemplified, P the property whose exemplification in x is an event, and t the time of exemplification. Talk of relations between mental

and physical descriptions could now be replaced by talk of relations between mental and physical properties, and non-reducibility was expressed by the claim that mental properties are not identical to physical properties.[4] This non-identity claim can be stated with varying force, allowing for different types of dependence on physical properties, as we shall see.

Given this added structure, and given non-reductive monism, a mental event will have two properties, a mental one and a physical one. This being so, a question naturally arises: in virtue of which of its properties is it causally responsible for bringing about the effects it does? The question is complicated by a number of assumptions thought to be part of the physicalist's ideological baggage. The physicalist is thought to be committed to the 'basic' or fundamental character of the physical, and an expression of this is contained in the assumption that the physical domain is causally closed: any event that has a cause has a complete (sufficient) physical cause. The thought that a mental event (or a mental property) could cause an effect without relying on, or working through, physical events (or properties) was rightly deemed inimical to physicalism. The tension with the claim that mental properties do ineliminable causal work is palpable, especially since systematic overdetermination of effects by both mental and physical properties is highly implausible.

Many of the lead chapters in this volume grapple with finding a satisfactory answer to the questions arising from this tension. A number of alternative answers are available: (a) the hard-core physicalist one: only the physical property is causally efficacious, the mental property being either eliminable or epiphenomenal (Papineau); (b) both mental and physical properties are causally efficacious, this being ensured by the identity of their exemplifyings or instancings (Macdonald and Macdonald); (c) both mental and physical properties are causally efficacious, this requiring a departure either from physicalism and the causal closure of the physical (Crane, O'Connor and Churchill, Hendry in the case of chemistry), or from strict extensionality of the causal relation (Menzies and List). The latter option suggests that when a mental property brings about a physical effect it could be the case that no physical property is doing the same work, so again causal closure of the physical is rejected.

[4] An event is mental just in case P is a mental property, and physical just in case P is a physical property. Since the identity conditions on events require identity of their constitutive objects, properties, and times, and since Kim maintains both that each event has only one constitutive property and that mental properties are constitutive properties of mental events, his version of the property exemplification account has the consequence that psychophysical event identity entails the identity of mental and physical properties constitutive of events, effectively ruling out non-reductive monism. There are other versions of the property exemplification account, however, that do not have this consequence, and these are compatible with non-reductive monism. For more on this see Macdonald and Macdonald (2006).

B. The Distinction Between Causation and Causal Explanation

The extensionality of the causal relation (more precisely, the extensionality of sentences describing a causal relation) is predicated on causal realism, the assumption being that causes do their work 'in the world', and so the events implicated in the causal relation, however complicated they are, can be variously described without any change in the truth of the causal claim. This requires that causes are not description-dependent, and if they are not there may be many true causal statements which are irrelevant to an *explanation* of a given effect. Some think that certain recent theories of causality might cast doubt on this assumption. In particular, some of the work done in the formulation of the 'manipulation' account of causation (Woodward 2003; see also Hitchcock 2001 and other work) may lead to the idea that what is essential in causation is that interventions changing the value of a variable in the cause (e.g., increasing the pressure of a fixed volume of gas) be 'matched' by an appropriate change in a variable in the effect (an increase in the temperature of the gas). Applied to our topic, the thought is that if an intention to drink some beer causes an appropriate action (e.g., drinking some beer), then the *relevant* 'change in the value of the variable' will be a change in the *intention* leading to a different effect, and not a change in any underlying physical property or properties (see Menzies and List, this volume). This will make the intention the relevant cause of the action, rather than some underlying physical state or event.

The result is that the truth of causal statements is much more dependent on explanation than had previously been thought. A satisfactory resolution of these issues requires answers to a number of questions that are connected to the metaphysics of events. Chief among these are whether the causes here are events or properties, and what the relation is of mental causes to physical causes. One possible way of keeping the causal relation extensional is by resisting the identification of issues concerning explanation with ones concerning causation, insisting on the difference between property-instance-causation (causal efficacy) and property causation (relevant to explanation) (Macdonald and Macdonald 2006).

C. Multiple Realizability and Non-Reducibility

One way of making the mental properties do causal work without violating the causal closure of the physical is by identifying them with physical properties. Non-reductive physicalists are not attracted to this solution for obvious reasons: such identities are usually thought to be a consequence of the reduction they reject. At least part of what motivated this rejection was the thought that mental properties are 'multiply realized' by physical properties, one psychological property being

'realizable' by different (sets of) physical properties, this meaning that there could not be the one-to-one relation required for any identity between the two. But given the tension described above between the physicalist assumptions and the belief in the causal efficacy of the mental, reduction has been thought to be the best way out. Kim eventually opted for this route, avoiding the multiple realizability claim by 'slimming down' the mental properties to suitable size. Initially (Kim 1993), instead of taking properties such as pain to be the mental properties available for reduction, he took species-specific properties, such as human pain, or bat pain, to be the target ones, maintaining that 'within species' psychophysical identities escape the force of any argument from multiple realizability. More recently (1998, 2005), Kim has 'slimmed down' the mental properties still further, to individual types of mental events or states as experienced at specific times by particular subjects, such as Jones' pain at 3 p.m. on Monday, 29 June 2009, claiming that such 'individual types' are capable of being functionally reduced to particular physical types of states.

It is a moot point whether this ploy is successful or not. There is an important issue to be raised about what is required for different tokens to be of the same type (see Heil 2003 for discussion), and so about whether instances of, say, bat pain are *sufficiently similar to* instances of human pain to warrant their being co-typed as instances of pain, but Kim's strategy appears more radical than one proposing caution with respect to co-typing. The suggestion seems to be that if there is a sufficient difference at the physical, *realizing* level then we should count the realized (mental) tokens as type-different. One objection is that Kim's refined position is only terminologically distinct from a non-reductionist token-identity proposal: tokens of higher-level (including mental) properties are to be collected into similar types if, and only if, they are of the same realizing (physical) type; otherwise they are assigned to different types. But further, the ad hoc nature of this manoeuvre is not really congenial to non-reductionists, as it makes the typing of the special science properties work from bottom-up, the typing being dependent on subvenient similarities, rather than being controlled by the requirements of the special science in question. It would, for example, rule out what could be interesting economic generalizations just because monetary exchanges are effected by different 'materials' (electronic versus chapter exchanges, for instance). For this reason the position will be uncongenial not only to the Fodorian wing of the non-reductionist camp but also to the Davidsonian one, since Davidsonians are inclined to stress the 'categorial' difference between the supervening and subvening properties, with the consequence that control from below will be seen to conflict with what is essential to the supervening set of properties.

Despite the apparent artificiality of Kim's proposal, his insistence on the causal troubles facing the non-reductionist has been bracing. There has also been much discussion of what makes for multiple realization, a notion central to the non-reductionist argument, with different accounts being given. Aizawa and Gillett (2009, citing Endicott 2005) mark a distinction between a computational

or mathematical account, where X is said to realize Y if the elements of Y map onto (are isomorphic with) the elements of X, and a *causal-mechanist* account where, very roughly, a higher-level property has causal powers which are determined by the causal powers of lower-level properties. According to the latter,

> A property G is multiply realized if and only if (i) under condition $, an individual s has an instance of property G in virtue of the powers contributed by instances of properties/relations F1 -Fn to s, or s's constituents, but not vice versa; (ii) under condition $* (which may or may not be identical to $), an individual s* (which may or may not be identical to s) has an instance of a property G in virtue of the powers contributed by instances of properties/relations F*1 -F*m to s* or s*'s constituents, but not vice versa; (iii) F1 -Fn $\neq$ F*1 -F*m and (iv), under conditions $ and $*, F1 -Fn and F*1 -F*m are at the same scientific level of properties. (Aizawa and Gillett 2009: 188)

Several features of this account are worth noting. First, just as it is important to note the relational character of emergence, it is also important that multiple realization is a relation between distinct levels of properties. How one distinguishes the relevant levels is crucial. It may turn out that *every* higher-level property is multiply realized if at the quantum-mechanical level there are different properties and relations realizing the same higher-level property and relation. But this possibility would not have much relevance for a debate about the possibility of reducing psychological properties to, say, neurophysiological properties. That reduction is blocked only if the psychological properties are multiply realized *with respect to* the neurophysiological level.

Second, those wishing to defend the non-reducibility of, say, psychology via a claim of multiple realization have an obligation to specify what makes the different realizations relevantly multiple: not just *any* differences between different instantiations of a property will do. That a psychological property's instantiations are realized by neurophysiological processes that sometimes occur in the right, sometimes in the left, hemisphere of the brain may not be evidence supporting a claim of multiple realization; location just may not be an important difference from the perspective of the neurophysiologist (see Shapiro 2000 and 2008 for discussion of this issue). The differences between different realizations must be ones that are relevant to the science, or theory, imposing the taxonomy at that level.

Relatedly, there is a question as to what would constitute evidence for a claim of multiple realizability. Some have thought it sufficient to imagine creatures composed of different 'stuff' from humans but displaying, for example, the same pain behaviour, the idea being that we would not desist from attributing pain to such an alien merely on the grounds that the alien's pain was thus differently realized. Here it is the *possibility* of different realizations that is thought to suffice for the truth of multiple realizability. Others insist on actual multiple realizations, the evidence being provided by the relevant sciences, with some being more sceptical than others. (Sceptics include Bechtel and Mundale (1999);

Polger (2009); Shapiro (2008). Optimistic responses include Aizawa and Gillett (2009) and Aizawa (2009). For a nice example of multiple realization from animal behaviour, see Keeley 2000.)

Third, the emphasis on causal powers leads to a problem discussed in more detail below (sub-section 5): the same causal power (of property G) is comprised by the causal powers of different sets of lower-level properties (F1-Fn and F*1-F*m), the difference between these sets of lower-level properties being (at least partly) constituted by a difference in their causal powers.[5] How can different lower-level causal powers constitute or 'comprise' the same higher-level causal power? The emphasis on realization makes the question about the causal power of higher-level properties more urgent: if the higher-level property's causal power is constituted by the contribution from the lower-level realizing properties, then it is difficult to see how its causal power could fail to be exhausted by that contribution—it seems that it will contribute nothing of its own to the effects it is said to cause.

These worries reflect the previously noted inherent tension in an emergentist perspective. Bedau specifies two characteristics of emergent phenomena:

(1) Emergent phenomena are somehow constituted by, and generated from, underlying processes.
(2) Emergent phenomena are somehow autonomous from underlying processes. (Bedau 1997: 376)

The complications arising from reconciling (1) with (2) are grist to the mill for some of those who, being inclined to reject the reductionist perspective, opt for a more radical position, rejecting the monism entailed by 'orthodox' non-reductive physicalism (Crane, O'Connor, and Churchill, Hendry, Menzies and List, all this volume). In doing so they embrace emergentism, and for them the question becomes one of how their position differs from orthodox non-reduction, given that this is already a property-dualist view.

D. Property Dualism and Emergence?

Once structured events appeared on the scene it became apparent that non-reductive physicalism was committed to a dualism of properties (at least). One suspicion was that this was just old-fashioned dualism in a different guise, and that property dualism was just as 'mysterious'. Defenders took the view that non-reductive physicalism is really very different from substance dualism, since a respectable naturalism can still be defended by making the mental properties *dependent* on physical properties. Non-reductionists took this dependence to require supervenience, the general idea being that one set of properties supervenes

[5] Aizawa and Gillett (2009) assume that, for scientific properties, a difference of property is sufficient for a difference in the causal powers of those properties.

on another 'base' set of properties if there can be no change in an object with regard to a supervening property without a change in the base set of properties. A great deal of work has been done to make this intuitive thought more precise (see, for example, Kim 1993 and papers therein; Horgan 1993; McLaughlin 1995, 1997a, 1997b).

The relevance for our purposes is that some emergentists wish to remain naturalists, thus rejecting the complete autonomy of non-physical properties (Crane, Noordhof, O'Connor and Churchill, Stephan, all this volume). For many, even emergent properties must be suitably related to physical ones if the mysteries of substantial dualism are to be avoided and naturalism is to prevail. Given that both types of non-reductive naturalism (non-reductive monism and emergence) require distinctness of properties (a requirement for non-reduction) and dependence (a requirement of naturalism), two ways of marking the distinction between non-reductive monism and emergence recommend themselves. The first would be to vary the type of *distinctness* of properties that generates the difference between the base and emergent properties, one suggestion being that non-reductive monism insists only on numerical distinctness of the properties, with emergence requiring mereological distinctness (see Stoljar 2008: 276). The second would be to vary the strength of the dependence between the base and supervening properties, one suggestion being that emergent properties are only nomologically dependent on physical properties, whilst non-reductive physicalism is committed to a stronger, metaphysical, dependence (Noordhof, this volume). This second way of drawing the distinction seems more promising to us, relating as it does to (sets of) properties, whereas the first seems to relate more properly to property instances.

Matters are in fact more complicated than this, since there is in the literature a third way of marking a relevant distinction, by distinguishing between 'weak' emergence and 'strong' emergence, this distinction being drawn in terms of causation. Strong emergence requires 'direct' downward causation, a causal power irreducible to the causal powers of the base set of micro-properties (O'Connor 1994 done); weak emergence then requires that the higher-level property be a structural property, one 'constituted wholly out of its microstates' (Bedau 1997: 378). Consequently, there is no 'mystery' of irreducible downward causation, and weak emergence becomes ubiquitous. For the 'emergence' of weakly emergent properties Bedau requires (roughly) that the higher-level (macro)state be derivable from the environmental input and the dynamic governing the microstates, but only by simulation; this guarantees a certain autonomy for the higher-level states without any 'metaphysical illegitimacy' (Bedau 1997: 396). The question remains as to whether strong emergence, so characterized, can maintain metaphysical legitimacy. (For further discussion and characterizations of weak and strong emergence see Gillett 2002; Chalmers 2008; and Macdonald and Macdonald, Noordhof, and Stephan this volume.)

E. The Coherence of Laws

As long as there is some form of dependence on physical properties, the non-reductivist/emergentist is burdened with giving some account of the connection between base and supervening properties. This connection has two facets, a synchronic dependence already remarked upon (some form of supervenience), and a diachronic 'harmony'. The former dependence has been much discussed, the latter less so. Diachronic harmony is required by the structure of supervenience when this structure is 'put in motion', when one considers the structure over time. Suppose, for example, that an instantiation of a mental property, in a suitable context of other properties, causes an action. This effect will itself have both mental and physical properties, and the process of the causing of action will need to respect supervenience. That is, the resultant state of affairs, the effect, will have to accord with the general picture of the relation between mental and physical properties. In the case envisaged, the physical causes of bodily movements will have to ensure that the right action is performed, 'right' here meaning an action that intelligibly flows from that intention. The puzzle is this: how do the different levels 'march in step', so that upper-level causes mirror lower-level causes in bringing about effects that supervene in the right way? One way of ensuring harmony between levels is to rely on natural design (Papineau, this volume, Chapter 12). If the levels are arranged so that they 'fit', the right causal profile of supervening and subvening properties is assured. The question then becomes one of ascertaining how such a fit is obtained. Darwinian natural selection is one answer; it operates on the effects of certain causes, such causes being grouped together, and so co-typed, *just because* they produce the desired (fitness-enhancing) effect. Such biofunctional properties can be multiply realized, the realizing properties having one thing in common: they cause the 'right' effect.

It is an open question as to how much of the causal harmony between the putatively non-reducible levels is explicable by the mechanism of a selection process. Hendry (this volume, Chapter 14) argues for strong non-reducibility for chemistry without relying on any selection process; Pettit (this volume, Chapter 17) outlines conditions under which group-rationality may be said to emerge, where a group may be said to exhibit rational agency, also apparently without the operation of selection, though he does include a 'disciplining' condition which could have the same effect as natural selection. This general topic of the compatibility of differing causal processes is relevant to issues as diverse as whether the kinds delivered by the taxonomies of the special sciences are natural kinds (Papineau), or whether there can be room for freedom given the causal closure of the physical (Stephan, this volume, Chapter 15).

4. CONTRIBUTIONS

We conclude with a brief summary of the lead contributions to this volume.

Tim Crane defends the claim that any genuinely physicalist position must distinguish itself from (what traditionally has been known as) emergentism; it cannot afford to postulate inexplicable or 'brute' correlations or identities. As a result, he argues, physicalism is necessarily reductive in character—it must either give a reductive account of apparently non-physical entities, or a reductive explanation of why there are non-physical entities. Crane claims that many recent 'non-reductive' physicalists do not do this, and because of this they cannot adequately distinguish their view from emergentism. This, he argues, is the real challenge posed by Joseph Levine's 'explanatory gap' argument: if physicalists cannot close the explanatory gap in Levine's preferred way, they must find some other way to do it. The price of failing to close the explanatory gap is to give up on non-reductive physicalism, since the resulting position will be indistinguishable from emergentism. Emergentists can embrace the generality of physics, with emergent properties being the supervenient properties of a thing not identical to any properties of its parts and supervenience being inexplicable in physical terms.

The attempt to reconcile non-reductive physicalism with a causal powers metaphysics is the target of the chapter by **Timothy O'Connor** and **John Ross Churchill**. The authors first outline Kim's attack on the non-reductionist's thesis, noting its dependence on a 'causal exclusion' principle. They argue that this assumes what they call a 'causal-powers metaphysic', a metaphysics requiring the exercise of ontologically primitive causal powers or capacities of particulars, these powers providing the 'oomph' in causation. Kim's argument is then reformulated to make this dependence explicit (hence the 'power-exclusion argument'). O'Connor and Churchill claim that if (as they think) this causal-power metaphysics is correct, then the non-reductivist position is incoherent, short of accepting systematic overdetermination of the mentally caused effects. Their argument is strengthened by examining a notable attempt to defend non-reductionism assuming the causal-power metaphysic, that of Shoemaker. They find this defence unpersuasive, concluding that non-reductive physicalism and the causal-power metaphysic are incompatible. The non-reductionist is forced, as a result, to give up on non-reduction and opt for either reductionism or eliminativism. Their preferred alternative is to sacrifice some of the premises leading to the unwelcome conclusion, specifically premises asserting the realization of the mental by the physical and the causal completeness of the physical. This inclines them towards the acceptance of an ontological variety of emergence, one respecting 'the distinctive character and efficacy of the mental'.

The focus of **Paul Noordhof**'s chapter is on the conditions that need to be satisfied if we are justifiably to say that we have emergent property causation. A necessary condition for property causation is that an instance of the property cited is a cause. Property causation also involves a certain kind of generality, say a causal law linking cause and effect. Noordhof identifies a set of narrowly physical property causes, a subclass of the class of properties identified by current physics or a future physics sufficiently resembling our own, containing just those properties which are property causes in this way. These are not exhaustive of the physical properties: there are broadly physical properties that supervene on the narrow physical properties. There is an important distinction to be made between the broadly physical properties and emergent properties. Both supervene on the narrow physical properties, but the former do so with metaphysical necessity, the latter with nomological necessity. Noordhof uses this distinction to motivate a difference between non-reductive physicalism and emergent dualism: both suppose that the instantiation of narrowly physical properties determines the instantiation of the other target properties, the non-reductive or emergent ones. They differ over whether the instantiation of these other properties involves something genuinely new. Non-reductive physicalists deny this, whereas emergent dualists assert it. The problem is to make sense of when there is something new introduced.

Building on work done elsewhere, Noordhof develops a counterfactual theory of property causation, focusing on the possibility of emergent causation, and in particular on the question of whether all emergent causation involves emergent property causation. His negative conclusion leads him to identify a second kind of emergence: emergent non-reductive physicalism. He goes on to apply the conclusions of this discussion to two candidates for emergence: phenomenal consciousness and free will.

Causation is clearly a central concern for non-reductionists of all persuasions. **Peter Menzies** and **Christian List** contend that recent interventionist accounts of causation, and in particular that developed by James Woodward, help to shed light on the debates (introduced by Kim) surrounding causal closure of the physical. Menzies and List use Woodward's interventionist account to identify necessary and sufficient conditions for the causal autonomy of a higher-level property and to show that these conditions are satisfied when causal claims about higher-level properties have a special feature, that of realization-insensitivity. This feature consists in the fact that relevant causal claims are true regardless of the way the higher-level properties they describe are physically realized. If higher-level properties are realization-insensitive, then when such a property, say a mental property Ma causes an action Ba, the realizing physical property does *not* cause Ba. This 'Downwards Exclusion Result' ensures the causal autonomy of the realization-insensitive properties. Menzies and List go on to show that these findings are consistent with those of other philosophers (e.g., Alan Garfinkel), who have noted the realization-insensitivity of higher-level

causal relations as a distinctive feature of the special sciences, and who have suggested that this feature ensures their independence from lower-level causal relations.

In their contribution, **Cynthia Macdonald** and **Graham Macdonald** support a form of strong emergence, one where emergent properties are not just complex properties derivable from the properties of more simple parts and their relations. The reason for the non-derivability of such emergent properties may differ from case to case, so there may be a different explanation of the non-derivability of biological properties than there is for the non-derivability of mental ones. The Macdonalds proceed to defend their version of strong emergence against critics, such as Kim, who think that it must lead to downward causation, which is claimed to be incompatible with the causal closure of the physical. The Macdonalds provide a metaphysics of events to show why this claim is false, arguing that this metaphysics allows two properties, say a mental and a physical property, to be co-instanced in a single event, and so allows the cause-event to be a single exemplifying of both a mental and physical property. This ensures the causal efficacy of the mental property *instance* while opening up space for distinctive causal (and explanatory) work to be done by the mental *property*. The chapter concludes by arguing against recent objections to this approach voiced by Philip Pettit (1996) and the alternative position developed, in different ways, by him and by Carl Gillett (2006a, 2006b).

Much of the discussion surrounding non-reductive physicalism has focused on its ability to allow supervenient causation. As **David Papineau** notes, there has been far less discussion of whether non-reductive physicalism can accommodate non-physical *laws*; Fodor assumed there were laws, even though these may be 'loose', or *ceteris paribus*, ones. The problem facing the Fodorian that Papineau points to concerns the compatibility of there being both supervening and subvening laws with multiple realizability: the *different* realizations of a supervening state, S_1, must all result in some physical state that determines the *same* effect, say S_2. The subvening heterogeneity, required for non-reduction, sits uncomfortably with the supervening homogeneity. The choice is either to give up on there being special science (supervening) laws, or to accept that the supposed underlying variability is an illusion, and that type-type reduction is available. Papineau queries whether this reduction seriously threatens the explanatory autonomy of the special sciences, given the often insuperable practical difficulties of providing reductions. The only way to avoid reducibility is to appeal to selection processes, processes in which the different physical states are 'designed' to lead to the same effect.

Although such 'selection-based patterns' may allow projectible correlations, Papineau worries that they result in a taxonomy of natural kinds that are too 'thin' to figure in a substantial science; for paradigm natural kinds, a multiplicity of their properties will figure in laws, whereas with selection-based kinds, only the property yielding the relevant effect is projectible. Pain, for example, leads

to damage-avoidance, and so the concept *pain* will figure in a relevant law, but insofar as pain is variably realized, no further laws will be forthcoming; to the extent that it is not multiply realized (say, in humans), there will be further laws to be discovered. The special sciences may consist in some 'thick' natural kinds (where there is no multiple realization) and 'thin' selection-based kinds. As far as causation is concerned, Papineau argues (with Menzies and List, and Hendry, this volume) against the assumption that the subvenient, realizing, fact always usurps the causal power of the more general supervening fact. But even granting this, there is still the problem that with variably realized special facts there will be no uniform physical law linking cause to effect. And in this case selection-based laws don't help, since these are based on pre-existing causal powers and do not add to them. Papineau concludes that non-reduced special kinds are not causes. They range over cases with quite different causal structures.

Most of the argumentation concerning the causal autonomy of the properties of higher-level sciences has focused on either psychology or sociology. **Robin Hendry** turns to chemistry, arguing against what he sees as the orthodox view amongst physicalists, who hold (i) that successful quantum-mechanical explanations of chemical bonding render unlikely the existence of downward causation from the chemical to the physical, and (ii) that the very idea of downward causation is murky. In his chapter he argues against both claims: against (i) he argues that it does not withstand investigation of either the early history or the mathematical structure of quantum chemistry, and against (ii) he proposes a counternomic criterion for downward causation, one which attempts to capture the emergentist views of C. D. Broad. The emergentism Hendry argues for permits, he claims, the supervenience of chemical properties on physical properties, since the supervenience relation is consistent with downward causation. With regard to the specific case of chemistry Hendry makes a distinction between resultant and configurational Hamiltonians (which describe the evolution of the quantum-mechanical complex system). The reductionist claims that the quantum-mechanical explanations of chemical structure and bonding will involve only resultant Hamiltonians; the emergentist expects that they will involve configurational Hamiltonians. Only evidence supplied by science can decide the issue, and Hendry argues that the evidence supports the emergentist. He concludes by replacing the physicalist's 'completeness of physics' with the 'ubiquity of physics': physical principles constrain the motions of particular systems without necessarily fully determining them.

Within the debates about mental causation and emergence there has been little discussion of free will, and it is this gap that **Achim Stephan**'s chapter aims to fill. Stephan situates his discussion in the context of free-will debates between two German philosophers and a neuroscientist, these usefully lining up as a libertarian, a compatibilist, and a hard determinist. The central question for each is how to envisage the relation between 'person-level' psychological properties and processes, and the underlying subpersonal neurophysiological

processes. Stephan discerns a surprising commonality amongst the disputants: all accept a synchronic dependency claim, that there can be no psychological change without a neurophysiological change. All are also committed to a diachronic 'principle of alternative possibilities' in characterizing what free will requires: the agent must have been able to have done otherwise. As Stephan sees it, the hard determinist must see the personal/subpersonal relation as one of reducibility: the person-level properties are explained (away) by the neurophysiological properties. The libertarian sees this relation as a case of emergent properties, where a property is said to be emergent if it is had by a system but not by that system's parts. Here the synchronic dependency claim is protected by postulating that non-deterministic psychological processes must rest on non-deterministic neurophysiological processes. Stephan argues that the compatibilist must embrace reductionism and rely on a notion of mental causation, where the subsequent actions and neurophysiological processes are caused by the agent's psychological properties, for example beliefs and desires. The nub of the dispute, as he sees it, relies on the plausibility of psychological reduction; given the imperfect state of knowledge concerning brain processes and their relation to psychological processes, no convincing conclusion is presently available.

The free-will advocate connects our supposed freedom with the exercise of our psychological, specifically rational, capacities. For individuals, **Philip Pettit** argues, a rational configuration of propositional attitudes is maintained by our ability to reason. The question he poses is whether there can be *group agents* and, if so, whether the rationality of the analogue of propositional attitudes can be maintained without reasoning. The possibility of group agents, and group rationality, also raises the question about the relation between the group and its members: must a group judgement supervene on the judgements of its members?

Reasoning, for Pettit, requires being able to monitor one's own propositional attitudes and actions, adjusting these where rationality so requires in response to evidence, to other attitudes, and/or to proposed actions. Pettit argues that under a range of plausible conditions, groups lacking feedback on their own decisions will not be able to reason, and so will not function satisfactorily as agents. This is supported by results on the aggregation of judgements that show, under certain conditions, the impossibility of ensuring that group judgements over connected issues will be complete and consistent.

Revising the way that group judgements are arrived at in any way that does not include systemic feedback will also fall short of ensuring group rationality, suggesting that any 'no-feedback constitution', one aiming to have rational but unreasoning group agents, is likely to fail. With feedback there is the possibility for group rationality to emerge: the group may be able to exercise the sort of control over its processes of judgement formation analogous to the personal control over judgements characteristic of individual reasoning. This control makes the group agent responsible for decisions reached, and elevates the group above that of a mere self-organizing collective to that of a self-governing collective.

REFERENCES

Aizawa, K. 2009. 'Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale'. *Synthese* 167: 493–510.

——and Gillett, C. 2009. 'The (Multiple) Realization of Psychological and other Properties in the Sciences'. *Mind & Language* 24: 181–208.

Allen, G. 1975. *Life Science in the Twentieth Century*. Cambridge, Cambridge University Press.

Anderson, P. W. 1972. 'More is Different'. *Science* 177, no. 4047: 393–6.

Bechtel, W. and Mundale, J. 1999. 'Multiple Realizability Revisited: Linking Cognitive and Neural States'. *Philosophy of Science* 66: 175–207.

—— and Richardson, R. 1998. 'Vitalism'. In E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge and Kegan Paul: 9.

Bedau, M. 1997. 'Weak Emergence'. *Philosophical Perspectives* 11: 375–99.

Chalmers, D. 2006. 'Strong and Weak Emergence'. In P. Clayton and P. Davies (eds.), *The Re-Emergence of Emergence*. Oxford: Oxford University Press, 244–54.

Davidson, D. 1970. 'Mental Events'. In L. Foster and J. Swanson (eds.) *Experience and Theory*. London: Duckworth, 79–101.

Endicott, R. 2005. 'Multiple Realizability'. In D. Borchert (ed.), *The Encyclopedia of Philosophy*, Supplement, 2nd edn. New York: Macmillan, 427–32.

Fodor, J. 1974. 'Special Sciences: or the Disunity of Science as a Working Hypothesis'. *Synthese* 28: 77–115.

Gillett, C. 2002. 'The Varieties of Emergence'. *Grazer Philosopische Studien* 65: 95–121.

——2006a. 'Samuel Alexander's Emergentism: or, Higher Causation for Physicalists'. *Synthese* 153: 261–96.

——2006b. 'The Hidden Battles Over Emergence'. In P. Clayton and Z. Simpson (eds.), *Oxford Handbook of Religion and Science*. Oxford: Oxford University Press, 261–96.

Heil, J. 2003. *From an Ontological Point of View*. Oxford: Clarendon Press.

Hitchcock, C. 2001. 'A Tale of Two Effects'. *Philosophical Review* 110: 361–96.

Horgan, T. 1993. 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World'. *Mind* 102: 555–86.

Humphreys, P. 2008. 'Computational and Conceptual Emergence'. *Philosophy of Science* 75: 584–94.

Johnson, S. 2001. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Scribner.

Keeley, B. 2000. 'Shocking Lessons from Electric Fish: the Theory and Practice of Multiple Realization'. *Philosophy of Science* 67: 444–65.

Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

——1998. *Mind in a Physical World*. Cambridge, MA: MIT Press.

——2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

Macdonald, C. and Macdonald, G. 2006. 'The Metaphysics of Mental Causation'. *Journal of Philosophy* 103: 539–76.

McLaughlin, B.P. 1992. 'The Rise and Fall of British Emergentism'. In A. Beckermann, H. Flohr, and J. Kim (eds.), *Emergence or Reduction?*, Berlin: de Gruyter, 1992, 49–93.

——1995. 'Varieties of Supervenience'. In E. Savellos and U. Yalcin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 16–59.
——1997a. 'Supervenience, Vagueness, and Determination'. *Philosophical Perspectives* 11: 209–30.
——1997b. 'Emergence and Supervenience' *Intellectica* 25: 25–43.
O'Connor, T. 1994. 'Emergent Properties'. *American Philosophical Quarterly* 31: 91–104.
Pettit, P. 1996. *The Common Mind*. Oxford: Oxford University Press.
Polger, T. 2007. 'Realization and the Metaphysics of Mind'. *Australasian Journal of Philosophy* 85: 233–59.
——2009. 'Multiple Realization and Evidence'. *Synthese* 167: 457–72.
Putnam, H. 1967. 'Psychological Predicates'. In W. H. Capitan and D. D. Merrill (eds.), *Art, Mind and Religion*, 37–48. Reprinted. as 'The Nature of Mental States', in H. Putnam, *Mind, Language and Reality* (1975), 429–40.
Shapiro, L. 2000. 'Multiple Realizations'. *Journal of Philosophy* 97: 635–54.
——2008. 'How to Test for Multiple Realization'. *Philosophy of Science* 75: 514–25.
Stoljar, D. 2008. 'Distinctions in Distinction'. In J. Kallestrup and J. Hohwy (eds.), *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*. Oxford: Oxford University Press, 263–79.
Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.