16

Strong Emergence and Freedom: Comments on Stephan

Max Kistler

In his chapter 'An Emergentist's Perspective on the Problem of Free Will', Achim Stephan asks whether new light can be shed on the free will problem by considering it from the viewpoint of different conceptions of emergence. The idea is certainly promising: part of what makes free will puzzling is the difficulty of understanding the relations between processes of deliberation and decision taking place at the psychological level and the neural processes underlying them in the brain. As Stephan reminds us, traditional answers to the free will problem can be sorted into three categories: determinism (according to which our conviction of being free is illusionary), compatibilism (according to which we can be free although all events and processes in our body are determined by earlier events and processes), and libertarianism (according to which our decisions are not determined by any earlier states of ourselves, be they psychological or physical). The hope is that the conceptual difficulties that bedevil any or all of these approaches can be partly or wholly solved in the framework of one or other conception of how psychological properties and processes *emerge from* brain properties and processes. Stephan himself takes the result of his inquiry to be negative. However, I shall at the end suggest a way in which emergence may help us make sense of freedom in a compatibilist way. Let me first make some remarks on Stephan's theory of emergence. According to Stephan (this volume), there are two independent reasons for considering a property as strongly emergent, in other words, more than weakly emergent. They correspond to two fundamental types of strong emergence, synchronic and diachronic.

The first way in which a property P can be strongly emergent is by being weakly emergent and *synchronically irreducible*: the fact that object o is P *at time t* cannot be deduced from the properties the object's parts possess *at t* together with their mutual relations *at t*.

The second way is to be weakly emergent *and (diachronically) unpredictable*: the fact that o is P *at t* cannot be deduced from the micro- and macro-properties o and its parts possess *at some earlier time t**.

1. SYNCHRONIC EMERGENCE AND IRREDUCIBILITY

Let us first look at the concept of strong emergence in terms of synchronic irreducibility. The problem for this concept is to reconcile the irreducibility of emergent properties with the hypothesis that they are synchronically *determined* by the system's parts. Synchronic determination, in the sense of nomological dependence of a systemic property on the properties of the system's parts and their interactions, is part of the concept of weak emergence: 'A system's properties and dispositions to behave depend nomologically on its micro-structure, that is to say, on its parts' properties and their arrangement' (Stephan 1999: 50–1).

Stephan takes synchronic determination to be compatible with 'synchronic irreducibility'.[1] This is indeed part of the doctrine of classical British emergentism.[2] However, today many doubt that there are any absolutely irreducible properties.[3] This change in mind is in large part due to quantum mechanics' achievement of reductively explaining chemical properties, which had been taken by emergentists such as Broad to be paradigmatic cases of properties that are irreducible although synchronically determined. It now seems that, if one takes it for granted that a given macroproperty is objectively synchronically determined by underlying (physical) microproperties, then it is a mere question of time when that determination relation will be discovered by scientific means. When we do not know how to reduce a given systemic property, this is not due to any objective feature of that property but only to our present ignorance and the imperfection of today's theories. Given any systemic property, there seems to be no reason to deny the possibility, at least in principle, that science eventually discovers its synchronic determination relation. That discovery provides scientists with the means of producing a reductive explanation of that property, in terms of the properties of the system's parts and their interactions.

One of Stephan's most interesting contributions to the analysis of emergence is his distinction between two steps of synchronic determination. This distinction might help us find out whether there can, after all, be good reasons to expect there to be absolutely (and not only provisionally) irreducible though synchronically determined properties.

Figure 16.1 (Boogerd et al. 2005) shows two ways a property can be emergent, corresponding to two steps of synchronic determination. For each step of

[1] See Stephan 1997. [2] See, e.g., Broad 2000 [1925]. [3] See McLaughlin 1992.

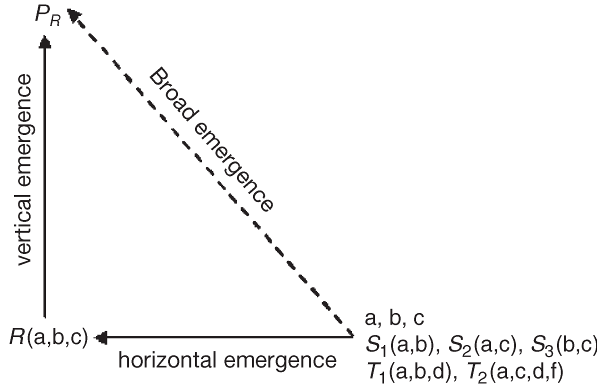

Figure 16.1. a, b, and c are the parts making up the system. S_1(a,b), S_2(a,c), and S_3(b,c) are simpler, binary, wholes including these parts. T_1(a,b,d) is a system with the same number of parts, and T_2(a,c,d,f) is a system with more parts than R(a,b,c). P_R is a systemic property. The diagonal arrow represents Broad's idea of emergence. The horizontal and vertical arrows capture the two conditions implicit in Broad.

determination it seems conceivable that it is impossible to reductively explain it. Thus it appears that there are two ways in which a systemic property P_R can be emergent, in the sense of being synchronically irreducible to properties and relations of systems' parts.

Systemic property P_R is synchronically irreducible either (1) because it is impossible to deduce the state R(a,b,c) of the interacting whole (where R(a,b,c) is taken to give rise to P_R) from the properties the parts *a*, *b*, *c* have when they are isolated, or from the properties of other systems (represented by S_1(a,b), S_2(a,b), . . .) which contain some of these parts. Boogerd et al. (2005) argue that this is the case for complex biochemical systems. In this case, P_R is a case of what they call 'horizontal emergence'. Or (2) P_R is synchronically irreducible because it is not 'behaviorally analyzable' (Stephan 1999: 52) in terms of R(a,b,c), which makes it a case of 'vertical emergence'.[4]

Let us look a little closer at the concepts of 'horizontal' and 'vertical' emergence. Take horizontal emergence first. Stephan takes the horizontal determination relation between the properties the parts have in isolation or in other circumstances (i.e. in systems S_1(a,b), S_2(a,b), . . .), and the state R(a, b, c) of the whole under consideration to be a case of 'synchronous emergence'. This raises the following difficulty. Following Humphreys (1997), we may call 'fusion' the process during which parts *a*, *b* and *c* come into interaction, and then eventually come to form a system. The problem is this: the horizontal relation cannot be synchronous because fusion takes time. Let me explain this in a little more detail.

[4] The diagonal line in Figure 16.1 represents Broad's concept of emergence, which conflates, according to Boogerd at al. (2005), horizontal and vertical emergence.

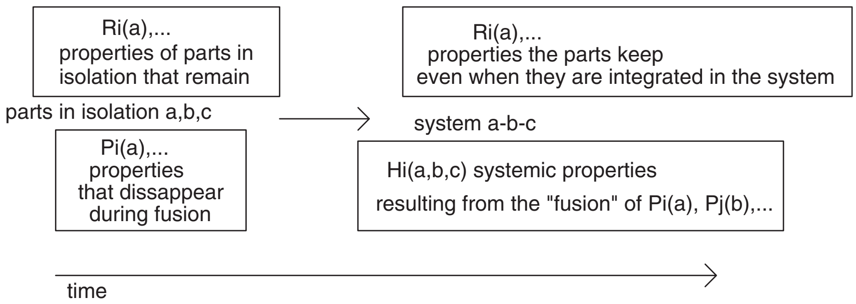

Figure 16.2. Horizontal emergence.

Under certain conditions, a stable global structure R(a,b,c) emerges ('horizontally') in such a way that, when it has emerged, *some* of the properties necessary for the interaction between the parts *a*, *b* and *c* are *lost*. As an example, take the formation of molecule electron orbits. They arise from the fusion of atomic electron orbits that *disappear* during the formation of the covalent chemical bond.

Figure 16.2 distinguishes properties P_i (a), . . . of parts that *disappear* in the process leading to an emergent property H_i (a,b,c) from properties R_i (a), . . . that remain. It is clear that those properties P_i that disappear before the emergent properties H_i comes into existence *do at no time coexist* with these emergent properties. Therefore, the relation between the properties P_i (a), . . . of the parts at the beginning of the fusion and the emergent property H_i (a,b,c) that exists at the end of the fusion process *is not synchronous*.

Let us now look at Stephan's second step of synchronic determination, and the 'vertical emergence' that arises from the impossibility to explain it reductively. He shows that there are really two ways for a systemic property P_R to be 'vertically emergent' (i.e. not to be 'behaviorally analyzable' in terms of the global state R(a,b,c)). They result from the impossibility to carry out one of the following two steps of reductive explanation.

In step (i), a systemic property is 'functionalized' in Levine's (1993) and Kim's (1998) sense, i.e. characterized by its functional role. In step (ii) it is then shown that 'the specified functional role [. . .] result[s] from the properties and behaviors of the system's parts and their mutual interactions' (Stephan, this volume: 233).

I suggest analysing step (ii) further in two substeps, so that there are more than two ways in which a property can be irreducible that make it 'vertically' emergent.[5] The first of these substeps consists in:

[5] I have argued elsewhere that neglecting this distinction creates problems for Chalmers and Jackson's (see Kistler 2005a) and Kim's (see Kistler 2005b) accounts of reductive explanation.

(iia) finding the property filling the role identified in step (i). This role-filling property is a systemic property of the interacting system, just as the role itself is played by the whole system. The second substep then consists in:

(iib) showing how properties of parts of the system and interactions between those parts bring the systemic role-filler property identified in step (iia) into being. One especially important way of doing this is by discovering a mechanism.[6]

Thus, there are really three, not just two, steps in vertical reduction.

(1) In a first step, a systemic property is functionalized by showing that the predicate expressing it does not directly denote a first-order property, but rather a role: it is equivalent to an existential quantification over some property or other that has certain causes and effects among system level properties.
(2) In a second step, the first-order property that fills the role specified in (1) is identified. This role-filler is a system level property, i.e. it belongs to the system as a whole.
(3) The role-filler property is analysed in terms of a mechanism.

Here is an example. 'Haemoglobin', though it appears to denote a substance (and 'being haemoglobin' a first-order property), really denotes the role F of being some substance or other transporting oxygen in mammal blood. What fills that role are those chemical properties M of different haemoglobin molecules that have (among others) the causal power of binding O_2 molecules. M's power of binding O_2 can then be reductively explained by interactions among the haemoglobin molecules' parts p_1, p_2, p_3. . . .[7] The mechanistic explanation of M's power of binding O_2 shows how the amino acids composing haemoglobin molecules interact ($R(p_1, p_2, \ldots)$) in such a way that the interaction gives rise to the conformation of the molecule which then explains M's binding power.

Terminology tends to obscure the difference between (1) the relation between role (F) and role-filler (M) and (2) the relation between a systemic property (M) and the elements p_1, p_2, $p_3 \ldots$ and organization $R(p_1, p_2, \ldots)$ of the mechanism: both relations are sometimes called 'realization', the discovery of both can be called 'reduction', and both are said to give rise to 'multiple realization/reduction'.

There appears to be no reason to expect any of these steps of determination to elude scientific discovery for principled reasons. There is no reason to expect any role F to be irreducible in principle, either because it is impossible in principle to find a role-filler property M or because it is impossible in principle to find a mechanistic explanation of M in terms of the system's parts and their interactions.

[6] See Machamer et al. 2000; Craver 2001; Bechtel 2006.
[7] See Rosenberg (1985: ch. 4).

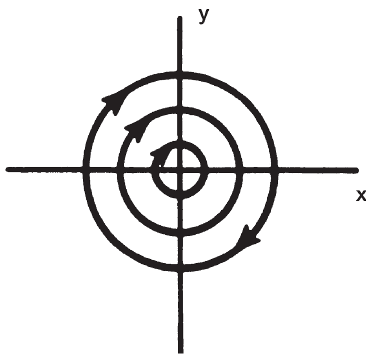

Figure 16.3. Phase space diagram of undamped harmonic oscillator (taken from Rueger 2000: 476).

It might be thought that this means that there are, after all, no strongly emergent properties. If strong emergence requires irreducibility and if there are no in principle irreducible properties, then there are no strongly emergent properties. However, this consequence is not inevitable: we can avoid it by construing strong emergence in a way compatible with reduction.

Let me mention one promising proposal of a criterion of emergence that does not require irreducibility. Rueger (2000) has suggested a *topological* criterion for diachronic emergence, which can also be used as a tool for constructing a concept of synchronic emergence compatible with reduction. A change between two dynamic states is *quantitative* if the corresponding trajectories are *topologically equivalent*. If the change is *qualitative* because the trajectories are *not* topologically equivalent, this may be taken as a ground for judging that this qualitative change is a case of emergence. Here is an example: Figure 16.3 can represent the trajectory, in phase space, of an undamped pendulum, i.e. a pendulum swinging without friction in a vacuum, if we take x to indicate angular deviation from the rest position and y to indicate angular speed. Figure 16.4 shows the trajectory of a damped oscillator. Introduction of damping causes a topological change in the form of the trajectory: it switches from circular to spiral. In terms of Rueger's criterion, this is a ground for taking the change to be qualitative. Such qualitative change can then be taken to be sufficient, together with weak emergence, for strong emergence.

2. DIACHRONIC STRUCTURE EMERGENCE

Let us now turn to the second of Stephan's concepts of strong emergence, diachronic structure emergence.[8] According to Stephan, a weakly emergent

[8] Stephan's concept of diachronic structure emergence is similar to Bedau's (1997) notion of 'weak emergence'.

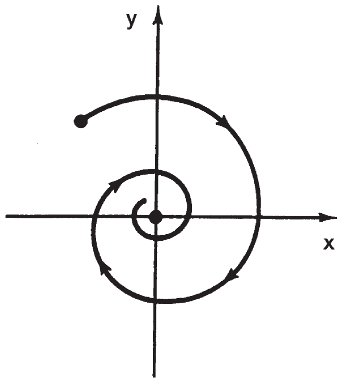

Figure 16.4. Phase space diagram of damped harmonic oscillator (taken from Rueger 2000: 475).

systemic state is diachronically structure emergent if and only if its formation obeys laws of deterministic chaos and is unpredictable unless by simulation. It is open to the charge that unpredictability unless by simulation seems to be neither necessary nor sufficient for emergence.

It is not necessary because the emergence of many properties is predictable without simulation: take a crystal that appears in the process of cooling a liquid. The coming into being of the crystal's observable macroproperties such as its colour, form, and hardness is a paradigmatic case of diachronic emergence. Although the movement of the molecules obeys deterministic chaos, the presence of attractors in such chaotic systems makes their evolution predictable.

Unpredictability is not sufficient either for emergence: if there is no point attractor, the evolution of a system may be unpredictable though nothing emerges. Take the system of air molecules in the atmosphere. The trajectories of the air molecules are not in the basin of any point attractor so that it is impossible to predict them in the long run. However, no qualitatively new property emerges from the evolution of this chaotic system.

3. FREEDOM AND CONSCIOUSNESS

Let me now turn to the main issue raised by Stephan's paper. All versions of the three replies to the free will problem Stephan mentions acknowledge the supervenience of mental processes on physical processes in the brain. Figure 16.5 sketches the compatibilist position advocated by Beckermann (2005). To take a free decision is a mental process, represented by a sequence of mental events m_0, $m_1, \ldots$, which supervenes on a parallel series of physical events $p_0, p_1, \ldots$.

Keil's (2007) 'libertarian' conception, sketched in Figure 16.6, does not differ from Beckermann's in respect of the relation between mental and underlying

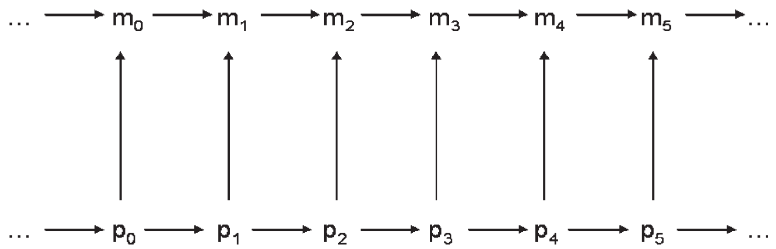

Figure 16.5. Schema of Beckermann's compatibilism (taken from Stephan, this volume).

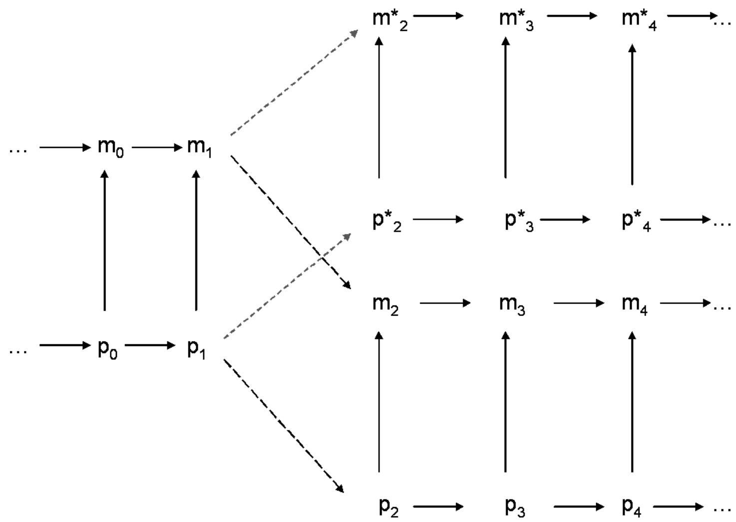

Figure 16.6. Schema of Keil's libertarianism (taken from Stephan, this volume).

physical properties: a series of mental events supervenes on a parallel series of physical events. The difference lies in Keil's thesis that both series contain bifurcation points. Free decisions are supposed to take place at these points.

Stephan's as well as Beckermann's and Keil's formulations leave open two interpretations of the relation between the physical events p_i and the mental events m_i in these schemas. According to the first, the mental events are distinct from the physical events though the former 'rest on' and supervene on the latter. The second interpretation has it that there is just one process that is both mental and physical, its constitutive events having both mental and physical properties. These interpretations correspond to two ways of conceiving of events: on Davidson's (1989 [1970]) conception, mental and physical events are token identical, in the sense that there is really just one event that can be described alternatively in mental or physical vocabulary. On Kim's conception, the relation between a mental event *m* and the physical event *p* it supervenes on is that

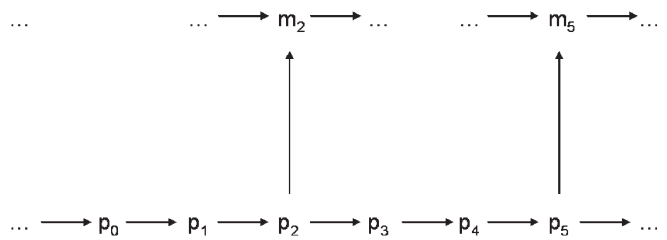

Figure 16.7. Stephan's (this volume) sketch of Singer's hard determinism.

between a role and what fills the role: *m* is the functional description of a mental role and *p* the physical role-filler. However, both interpretations have in common that there could be no causal interaction between the mental events m_i in the process leading to an action and the physical events p_i on which they supervene.

Let us now turn to Singer's position, according to which our belief in free will is an illusion. Singer holds that what distinguishes apparently free actions from other actions is that part of the process leading to the action is *conscious*. Stephan's interpretation, sketched in Figure 16.7, misrepresents Singer's position by assimilating it to Beckermann's and Keil's with respect to the relation between mental events and their underlying physical events.

It is incompatible with Singer's position to represent, as does Stephan in Figure 16.7, the relation between event p_2 (that is part of the chain causing some muscle movement constitutive of a given action) and event m_2, which is the event of the agent becoming conscious of p_2, by the same symbol (vertical arrow) as the relation between p_i and m_i in the schemas (Figures 16.5 and 16.6) representing Beckermann's and Keil's positions.

The incompatibility stems from the fact that the relation between a mental event m_i and the underlying physical event p_i is, as we have seen, not causal, whereas the process making a state conscious *is* causal.

According to the main psychological model for (access-)[9] consciousness, a mental state *m* is conscious if its content is *accessible* for information processing. In functional terms, *m* is conscious if and only if it is situated in a 'global work space'.[10] If a state is situated in the global work space, it has the capacity to interact with many functional subsystems of the mind/brain—both on the input (vision, language understanding) and output (motor system) side. This capacity rests on the configuration and strength of neural connections. Therefore, an event m_i of becoming (access-) conscious of an event p_i is not only mentally but also physically different from p_i, and the relation $p_i \rightarrow m_i$ is causal.

9 Block (1995) has introduced a famous distinction between access and phenomenal consciousness. Psychological research focuses on access consciousness.

10 See Baars 1997; Dehaene et al. 2003.

4. FREEDOM DESPITE DIACHRONIC AND SYNCHRONIC DETERMINATION

Let me end with sketching a way in which the concept of emergence might indeed be used in constructing a compatibilist solution to the free will problem. Emergence can help us understand how a complex system such as a human being could be free, although it is composed exclusively from physical parts subject to deterministic laws.

The first hypothesis I will use is that the human body, and in particular its brain, is a complex system that obeys the laws of deterministic chaos. In that case, for a given precision of description, it is impossible to predict, from a description of the conjunction of the states of the parts of the system (the neurons and synapses) at time t, a description of the conjunction of the states of the parts at $t + \Delta t$ if the time span Δt is long enough. This impossibility to predict is common for complex systems exhibiting deterministic chaos. The evolution of the conjunction of microstates is 'undetermined' in this sense. For any given set of global states of the body $\{S_i(t)\}$ that belong, at time t, to a given type T, these states have at much later times evolved into states $\{S_i\ (t+\Delta t)\}$ that do not any longer belong to any common type T*.

The second hypothesis is that mental states emerge from brain states, either by 'horizontal' emergence, i.e. through fusion, or by 'vertical' emergence, i.e. by the systematic interaction of the parts of the brain in a mechanism.

The third hypothesis is that these mental states obey to 'system laws',[11] in this case psychological laws. Those laws impose constraints on the evolution of the system and thus contribute to determining the evolution of (1) the system properties and (2) the state of the system's parts (neurons and synapses).

In this framework, the conviction that our actions are free, i.e. determined at a psychological level by our preferences and beliefs, can be reconciled with the conviction that all parts of our bodies and brains obey deterministic laws. The state of the body of a person at time t_3 is determined jointly by two constraints: (1) by physical laws in virtue of the physical properties of the parts of the system at t_2 (this is short-term diachronic determination because t_2 must immediately precede t_3), and (2) by the psychological laws applying to the person by virtue of systemic properties of the system at t_1 (preceding t_3 by a longer time span). A description of all parts of the person's body and their interactions does not suffice to predict a description of all parts at a much later time. In this sense, the state of our body does not on its own determine the state of our

[11] The term has been coined by Schurz (2002). Such laws are valid for specific types of system, such as the organisms of a certain biological species or ruby lasers. If the evolution of a system is regular enough that it obeys such a law, it is what Cartwright (1999) has called a 'nomological machine'.

body over long time spans. In particular, it does not on its own determine our actions. The determination of our actions is mediated by emergent psychological properties of our body and by psychological processes such as deliberation and decision.[12]

REFERENCES

Baars, B. 1997. *In the Theater of Consciousness*. New York: Oxford University Press.

Bechtel, William 2006. *Discovering Cell Mechanisms*, Cambridge: Cambridge University Press.

Beckermann, A. 2005. 'Free Will in a Natural Order of the World'. In C. Nimtz and A. Beckermann (eds), *Philosophy—Science—Scientific Philosophy. Main Lectures and Colloquia of GAP 5*. Paderborn: Mentis, 111–26.

Bedau, M. A. 1997. 'Weak Emergence'. *Philosophical Perspectives: Mind, Causation, and World* 11: 375–99.

Block, N. 1995. 'On a Confusion about a Function of Consciousness'. Repr. in N. Block, O. Flanagan, and G.Güzeldere (eds), *The Nature of Consciousness*. Cambridge (MA): MIT Press, 1997, ch. 20.

Boogerd, F.C., Bruggemann, F.J., Richardson, R.C., Stephan, A., and Westerhoff, H. 2005. 'Emergence and its Place in Nature'. *Synthese* 145: 131–64.

Broad, C. D. 2000 [1925]. *The Mind and its Place in Nature*. London: Harcourt, Brace and Co.; repr. London: Routledge.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Craver, C. F. 2001. 'Role Functions, Mechanisms and Hierarchy'. *Philosophy of Science* 68: 31–55.

Davidson, D. 1989 [1970]. 'Mental Events'. In idem, *Essays on Actions and Events*. Oxford: Oxford University Press, 207–27.

Dehaene, S., Sergent, C. and Changeux, J.-P. 2003. 'A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data during Conscious Perception'. *Proceedings of the National Academy of Science* (USA), 100.14: 8520–8525.

Humphreys, P. 1996. 'Aspects of Emergence'. *Philosophical Topics* 24: 53–70.

—— 1997. 'How Properties Emerge'. *Philosophy of Science* 64: 1–17.

Keil, G. 2007. 'Mythen über die libertarische Freiheitsauffassung'. In D. Ganten, V. Gerhardt, and J. Nida-Rümelin (eds), *Die Naturgeschichte der Freiheit*. Berlin and New York: de Gruyter.

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

Kistler, M. 2005a. 'Réduction "rôle-occupant", réduction "micro-macro", et explication réductrice a priori'. *Dialogue : Canadian Philosophical Review* 44: 225–48.

—— 2005b. 'Is Functional Reduction Logical Reduction?'. *Croatian Journal of Philosophy* 5: 219–34.

—— 2006. 'The Mental, the Macroscopic, and their Effects'. *Epistemologia* (Genoa, Italy), 29: 79–102.

[12] I have developed this idea in Kistler (2006) and (2007).

—— 2007. 'La réduction, l'émergence, l'unité de la science et les niveaux de réalité'. *Matière Première* (Paris, France) 2: 67–97.

Kronz, F. M. and Tiehen, J. T. 2002. 'Emergence and Quantum Mechanics'. *Philosophy of Science* 69: 324–47.

Levine, J. 1993. 'On Leaving out What it's Like'. In M. Davies and G. Humphreys (eds), *Consciousness: Psychological and Philosophical Essays*. London: Blackwell, 121–36.

Machamer, P., Darden, L., and Craver, C. F. 2000. 'Thinking about Mechanisms'. *Philosophy of Science* 67: 1–25.

McLaughlin, B. 1992. 'The Rise and Fall of British Emergentism'. In A. Beckermann, H. Flohr, and J. Kim (eds), *Emergence or Reduction?—Essays on the Prospects of Nonreductive Physicalism*. Berlin and New York: de Gruyter, 49–93.

Rosenberg, A. 1985. *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Rueger, A. 2000. 'Robust Supervenience and Emergence'. *Philosophy of Science* 67: 466–89.

Schurz, G. 2002. '*Ceteris paribus* Laws: Classification and Deconstruction'. *Erkenntnis* 57: 351–72.

Stephan, A. 1997. 'Armchair Arguments against Emergentism'. *Erkenntnis* 46: 305–14.

—— 1999. 'Varieties of Emergentism'. *Evolution and Cognition* 5: 49–59.