

## The Explicable Emergence of the Mind

**Elizabeth Schier (lizschier@gmail.com)**

Macquarie Centre for Cognitive Science, Macquarie University  
NSW 2109, Australia

### Abstract

The goal of the symposium 'Integrating Perspectives on the Relation between Mind and Brain' was to get people with different views and from different disciplines to open up a dialogue by focusing on answering a set of questions. In this paper I present a view of the relation between the mind and the brain that is informed by recent work in the philosophy of science. The basic idea is that the mind is more than the brain because mental states are identical to the activity of groups of organized neurons. Unlike the standard non-reductive materialism irreducibility is not seen as related to multiple realizability. The upshot is that we can bring the relation between the mind and the brain in line with other clear cases of ontological emergence, we can see how psychology can be an independent science, and yet how important explanatory connections can be made between psychology and neuroscience.

**Keywords:** Reduction; emergence; levels of composition; levels of explanation; multiple realizability.

### Introduction

Normally debates about the reducibility or not of the mind focus on the relation between different sciences. So Fodor (1974) argues that the laws of psychology is independent of the laws of physics and more recently Coltheart (2006; This Issue) has asked whether cognitive neuroscience has informed cognitive psychology. In this paper I want to present an alternative approach to these questions. In particular, the idea is to focus not on the relation between sciences but on the entities and properties in the world itself (for more on these different ways of understanding levels see Opie, This Issue). The utility of this approach is first of all that there is a clear story to tell about simple examples and secondly that it is obviously different from an approach that focuses on the relation between disciplines. I will begin by considering the lessons learned from the simple example of the relation between water and H<sub>2</sub>O. Then I will consider how the account that is developed out of this example can be applied to the mind. Finally I will show how the account developed leads to a different set of answers to the questions posed in the symposium.

### Explicable Emergence

It has become standard in philosophy to base the irreducibility of mental states in their multiple realizability (MR). The basic idea is that what makes something a mental state if its causal (functional) role in the system. So a belief or desire can be realized in many different ways; in neurons, or on a silicon chip. Similarly it seems that an and-gate can

be constructed out of an electronic circuit or cats and mice (Block 1995). The upshot is that there are many different ways of building a mind; with neurons, with silicon chips and even Swiss Cheese (Putnam 1973).

The debate about whether the mental states are functional (Block 1980; Searle 1981; O'Brien and Opie 1999), whether MR actually occurs (Shapiro 2005) and whether MR is sufficient for the mind to be more than the brain (Kim 1993) rages. However I want to suggest that we can side step this debate. When we look at a simple example it seems that MR can be beside the point for emergence. Of course it is far beyond the scope of this small paper to spell out in details the differences between the two approaches. Rather, the goal is to demonstrate the promise of this apparently different, more mechanistic account of emergence.

### The Lesson from Water

Consider for example the relation between water and H<sub>2</sub>O. Although it is undeniably correct to say that water is composed of H<sub>2</sub>O molecules, it is an oversimplification to say that water is simply H<sub>2</sub>O.<sup>1</sup> For starters a single molecule of H<sub>2</sub>O is not a liquid. Secondly a group of H<sub>2</sub>O molecules need not form any macro-substance, say because they are distributed evenly throughout the universe. Even if we stipulate that the atoms are in "close quarters" such that they form a macro substance we are not guaranteed that this substance will be water. This is because ice and steam are also composed of H<sub>2</sub>O molecules in close quarters. Obviously we do not want to identify water and ice and steam. So we need to stipulate not only that the substance is made of H<sub>2</sub>O molecules but also that they are bonded in a particular way (see below). The point is that behind even simple identities we see that there are actually examples of an organized whole having properties that its parts and its parts in a different organization lack. I want to suggest that in these simple cases we see examples of emergence.

But why say this is a case of emergence? First of all it is clear that the organized whole and its parts have different properties. One test that seems to provide at least a sufficient condition for such novelty is Alexander's dictum (Kim 1992). This principle states that to be real is to have causal powers, so for a whole to be different it needs to have different causal powers to its parts.<sup>2</sup> It should be clear that

<sup>1</sup> Note I am glossing over the fact that "pure" water will also contain D<sub>2</sub>O (i.e. hydrogen atoms that have not only a proton but also a neutron in the nucleus).

<sup>2</sup> Note that I am talking of causal powers here for the sake of convenience and this should not be taken to indicate a deep

water has casual powers that are distinct from those of its parts in a different organization. For example, it is crucially important for the person about to dive off the 10m platform that the H<sub>2</sub>O in the pool is water and not ice or steam.

So we can see that organized wholes, such as water, have properties, such as liquidity, that the same parts in a different organization lack. But to say that a property is emergent is also to say that it is at a higher ontological level. Why think these new properties of organized wholes are at a different level to their parts? One common characterization of levels is in terms of parts and wholes; the parts are at one level and the whole is at another level.<sup>3</sup> So for example protons, neutrons and electrons are at one level and atoms are at a higher-level. Given the prima-facie plausibility of this notion of levels, and the fact that organized wholes, such as water or ice have properties that their parts lack, it seems plausible to put these new properties of the wholes at a higher level to the parts. As will become apparent below it is important to note that although the whole is at a higher-level to the parts, the whole is also identical to the *organized* parts. So water is just a group of H<sub>2</sub>O molecules arranged in one way, but the same group of H<sub>2</sub>O molecules could be ice if arranged in a different way.

Finally, I want to suggest that this is a genuine example of emergence, despite the fact that the emergence of the higher-level properties is explicable. ‘Emergence’ was historically used by the British Emergentists to indicate a relation between wholes and parts where the whole not only has novel properties to the parts but also is inexplicable in terms of its parts (McLaughlin 1992; Schier 2007). For example they accepted that consciousness was a property of the brain, but held that it was impossible to explain how appropriately organized groups of neurons could be conscious. It should be clear that the notion of emergence at play here relinquishes any claim to the inexplicability of emergent properties in terms of their parts. For example it is possible to explain the different densities of ice and (cold) water in terms of the nature of the hydrogen bonds between the molecules. In ice the molecules are (on average) moving relatively slowly, so the bonds between the molecules are straight and strong, resulting in ice having an open lattice-like structure. In contrast in water, the molecules are moving faster such that some of the hydrogen bonds are broken. The result is that some molecules can move into the cavities of the lattice-structure in ice such that there are more molecules per unit volume. We can see that a more modern notion of emergence retains the notion that wholes have properties that their parts lack while rejecting the claim that the emergence of such properties is inexplicable (for more details see Schier 2007).

---

commitment to understanding causation in terms of causal powers rather than say, causal relevance. It should be clear that the organization of the parts is a causally relevant factor. For example the organization of a group of H<sub>2</sub>O molecules is relevant to accounting for why they caused a burn rather than soothing it.

<sup>3</sup> See Craver (2007) for why this notion of levels is distinct from the formal mereological notion of levels.

It seems that the simple example of water demonstrates that there are emergent properties. First of all organized wholes can do things that their parts can't and so are ontologically distinct from their parts. Secondly, given that one common conception of levels is in part whole terms, it seems plausible to put these new properties of wholes on a higher-level to the parts. Finally, although such emergent properties are explicable in terms of their parts and the relations between their parts, we need to reject the British Emergentist notion of emergence which held not only that wholes had novel properties but also that the emergence of such properties was inexplicable. Now that we have a general notion of what is involved in such “explicable” emergence I want to turn to consider how the mind might be explicably emergent with respect to the brain.

### **How the Mind Explicably Emerges**

We have seen that the property of being liquid water is explicably emergent because the organization of the constituent H<sub>2</sub>O molecules is a determinant of the properties of the whole. I now want to consider how such an account could be extended to the mind. One feature of minds is that they are able to represent the world, that is, mental states are about the world. So my current visual experience is about the text on the screen in front of me and my current auditory experience is about the Gomez album that I am listening to. Importantly it seems that representation provides a good example of how the mind could explicably emerge from the brain.

A point of almost universal consensus amongst psychologists and cognitive neuroscientists is that the brain uses distributed or population coding. This is in contrast to local coding, where the activity of an individual neuron codes for the presence or absence of a particular stimulus, say your grandmother. However it now seems clear that no such “grandmother neurons” exist, and that neurons will change their firing rate to a range of stimuli. For example a single neuron in the Fusiform Face Area (FFA) will respond not only to the presence of your grandmother's face but also to the presence of your mother's or aunt's face. Your grandmother's face is represented by the collective activity of a group of neurons in the FFA and your mother's face is represented by a different overall type of activity of the same set of neurons in the FFA. Looking at the activity of an individual neuron is like looking at an individual pixel in the text of this page, it won't tell you what the letter or word is. This is because an individual pixel is not a representing vehicle; rather it is part of a representing vehicle. Similarly it seems that the brain's representing vehicles are not the activity of individual neurons but rather the activity of a population of neurons.

If the brain uses distributed representations, then representational content explicably emerges from individual neurons. Just as an individual H<sub>2</sub>O molecule is not a liquid, so too does the activity of an individual neuron not represent anything. Similarly, just as a group of H<sub>2</sub>O molecules can be a liquid, solid or gas, so too can a group of firing neurons

represent your grandmother, mother or aunt. It is only when a group of neurons are organized in the appropriate way that they represent your grandmother. The same neurons in a different organization can represent nothing (say because they are all firing at resting rate) or can represent something different (because they are firing in a different pattern). As with water, the arrangement of the constituents determines the properties of the whole and so, as with water, it seems that at least representational content is an explicably emergent property of the mind.

Now that we understand how the mental properties could be explicably emergent, I now want to turn to consider how the questions posed in the symposium would be answered on this account.

## The Questions

In this section I will focus on the questions that explicable emergence gives a clear answer too. In particular I will focus on the first four questions.

### Question 1

*Are the explanations provided by cognitive psychology and cognitive science indispensable? Or are they replaceable in principle by explanations in neuroscience? If not, why not?*

To claim that the mind is explicably emergent is to claim that new mental properties, such as the property of having representational content, only occur when firing neurons are arranged in the right way. Importantly these new mental properties will require explanations that are as distinct from neural explanations as the mental properties themselves are.

However, the fact that the bearer of the new mental properties is identical to a complex neurological state complicates the picture. As we saw above, the thing that has representational content just is a complex brain state, that is, a set of neurons firing in the right sort of way. So although there are new mental properties, it is far from clear that these must be referred to in purely cognitive terms. Note I am not suggesting that it will be possible to remove all talk of the property of having representational content. Rather the suggestion is that the explanation of that property may be neuroscientific.

Further exploring the case of the property of being liquid water may help. This is a new emergent property that is not had by individual H<sub>2</sub>O molecules and so requires a new explanation over and above any explanation of the nature of H<sub>2</sub>O molecules. But because the property of being liquid water is just a property of a group of H<sub>2</sub>O molecules arranged in the right way it is possible to explain the nature of water in terms of the nature of the H<sub>2</sub>O molecules. Moreover it seems possible, at least in principle, to remove all talk of the property of being liquid water and replace it with talk of the property of being a collection of H<sub>2</sub>O molecules arranged in such-and-such a way.

What we can see is that explicable emergence leads to a very different picture of the relations between explanations at different levels than is provided by the standard Fodorian picture. According to Fodor (1974; 1997) higher-level

properties are “radically” multiply realizable in that the realization base of such properties has nothing in common from the point of view of the lower-level science. So consider for example the property of being a form of monetary exchange. A clay tablet, a copper coin, a piece of plastic, a cow or a pattern of activity across silicon chips can all be forms of monetary exchange. It is quite clear that the various realizations of such a property have nothing in common from the point of view of physics. So there is no way to use the vocabulary of physics to refer to such properties. In contrast it seems that it is possible, at least in principle, to use the vocabulary of physics to refer to the complex arrangement of H<sub>2</sub>O molecules that realizes the property of being liquid water.

What we can see is that explicable emergence provides a very different view of the relationship between ontological and explanatory levels. On such an account it seems quite plausible to say that one explanation is at a higher-level to another because the former explanation refers to the whole that is at a higher-level to the (unorganized) parts. But because higher-level wholes can be explained in terms of their parts, a higher-level explanation need not be in a radically new vocabulary to a lower-level explanation. The upshot is that the properties referred to in cognitive psychology will not be eliminable, but it might be possible to refer to and explain such properties in neuroscientific, or at least cognitive neuroscientific terms.

The point is that because an account in terms of explicable emergence begins with a focus on the entities themselves, there is not a simple relation between levels defined in terms of those entities and their parts and the vocabularies and domains of existing sciences. For example cognitive neuroscience can explain mental representations because they are identical to patterns of activity across organized layers of neurons. But because representations are real and distinct from the firing of individual neurons, the phenomena of cognitive science are not eliminated in favor of a non-cognitive neural explanation. So the current position is in between a Fodorian inspired MR account where explanatory connections between levels are impossible and a Bickle-style radical eliminative account where we can eliminate any reference to cognitive phenomena altogether. Importantly, the notion that the relations between sciences are either ones of indispensability or replacement seems mistaken. Instead it seems that we may end up with a range of interconnected yet independent sciences such as that envisaged by Darden and Maull (2000).

### Questions 2 and 3

*Are the phenomena referred to in explanations in cognitive science realizable or implementable in many different ways at the neurophysiological level? If so, is this why these explanations are irreplaceable?*

*How do mental or psychological states relate to neural states? For example, are they identical to neural states?*

For explicably emergent properties the answers to these two questions are quite intertwined. This is because there seem to be cases where the realiser of explicably emergent properties that are identical to organized lower-level parts but there are also examples of explicably emergent properties that are multiply realizable. The unanswered research question is what category the mind may fit into.

Consider identity first. Although it is not true that the property of being liquid water is identical to the property of being an H<sub>2</sub>O molecule, it is the case that the property of being liquid water is identical to the property of being H<sub>2</sub>O molecules arranged in a particular way. Similarly, although the property of having representational content is not identical to the property of being a firing neuron, it is identical to the property of being a group of firing neurons arranged in a particular way. So there is a sense in which it seems plausible to say that psychological states are identical to neural states. However there is a problem. Liquidity may be a better analogy for psychological properties than water.

Explicable emergence puts the focus on parts and their organization and it seems that there are cases where different parts can have the same organization. Consider for example liquidity. Many different types of particles can be liquids. So the property of being a liquid is multiply realisable. Similarly, it seems hasty to rule out the possibility that mental states can be realized in, say artificial neural networks. If this is the case then it would be as inappropriate to identify mental properties with neural properties as it would to identify the property of being a liquid with the property of being a group of H<sub>2</sub>O molecules arranged in the right way.

But it seems that because there is a structure in common, and because the structures which parts can form is going to depend on the nature of the parts themselves, this sort of multiple realizability will be no bar to an explanation of the properties of the whole in terms of the parts. So although both water and milk are liquids, we can explain why water boils at a much higher temperature than milk because of the relative strength of the inter-molecular hydrogen bonds. Moreover, although different types of parts can have the same property, this does not mean that the difference in the nature of the parts is irrelevant. For example although both water and petrol are liquids, this does not mean that your car will work if you fill the petrol tank with water. So it may be the case that for some explanatory purposes, the precise nature of the multiply realizable property is irrelevant. But for other explanatory purposes the difference between the different types of realisers may be important.

What we can see is that regardless of how the data turns out, it is clear that if explicably emergent phenomena are multiply realizable, their multiple realizability plays a very distinct role from that imagined by non-reductive materialists, such as Fodor, who base irreducibility in multiple realizability. Explicably emergent properties are irreducible not because they are multiply realizable but because the organization of the parts is a determinant of their nature.

#### **Question 4**

*Is it possible or likely that cognitive psychology will reduce to neuroscience? If so, what is the relevant sense of reduction? If so, will a successful reduction mean that phenomena (the entities/properties/processes) at the cognitive level do not exist?*

It seems that for explicably emergent properties it is possible to get a 'reduction without leveling' (Bunge 1977). That is, as argued above, we can explain the nature of the whole in terms of the parts, and yet the whole has new properties that the parts lack. So obtaining an explanatory connection between levels does not lead to the elimination ('leveling') of the higher-level property. Yet, unlike in a Fodorian account, it is possible to get some minimal form of reduction between levels in the sense that it is possible to form explanatory connections between levels. So cognitive psychology will reduce to neuroscience to the extent that it is possible to explain cognitive phenomena in neural terms.<sup>4</sup> But establishing such an explanatory connection between levels will not lead to the elimination of higher-level properties or of explanations that refer to them. This is simply because properties at a higher-level are real and distinct from properties at the lower-level.

#### **Conclusion**

We can see that an account which takes the mind to explicably emerge from the brain provides a form of non-reductive materialism that is distinct from the standard Fodorian account. First of all irreducibility is not based in the multiple realizability of higher-level properties, rather it arises from the fact that organized wholes have properties that the parts in different organizations lack. So the whole is more than the parts because the whole is the parts in a particular organization. This means that although the mind may turn out to be multiply realizable, this is not only irrelevant to its emergence, it may also be possible that the mind is not multiply realizable but rather identical to the brain. Secondly levels are defined in part-whole terms rather than in terms of properties and their realisers (for more on these different notions of levels see Kim 1993). So there is no clear mapping between sciences as they exist now and ontological levels. For example a mental state such as representing the screen in front of me could be described either in psychological or neuroscientific terms. Finally, although explicably emergent properties are reducible to the extent that they can be explained in terms of the nature of their constituents this in no way means that such properties or explanations that refer to them can be eliminated. If the

---

<sup>4</sup> Note that a reduction to neural states looks more plausible for representation than for computation. This is because computation is the manipulation of representations. This adds an extra level of organization, namely the temporal organization of particular representations. Accounting for this extra level of organization in purely neural terms may prove even harder than accounting for the intermediate representational level in neural terms. (Thanks to Glenn Carruthers for pointing this out)

mind is explicably emergent then the psychological and neuroscientific levels neither float free nor collapse. Rather they are connected but distinct.

### Acknowledgments

I want to thank the other participants in the symposium as well as members of the levels group at Macquarie University, particularly Glenn Carruthers, Wayne Christensen, Max Coltheart, Peter Menzies and John Sutton. Finally I want to thank Kellie Williams for bringing all of these issues into clear focus.

### References

- Block, N. (1980). Troubles with Functionalism. *Readings in Philosophy of Psychology*. (N. Block. 1.) Cambridge, MA, Harvard University Press
- Block, N. (1995). The Mind as the Software of the Brain. *An Invitation to Cognitive Science, Volume 3*. (D. Osherson, L. Gleitman, S. Kosslyn, S. E. and S. Sternberg.) Cambridge, MIT Press
- Bunge, M. (1977). 'Levels and Reduction'. *American Journal of Physiology*, 233(3), 75-82.
- Coltheart, M. (2006). What has Functional Neuroimaging Told us About the Mind (so far)? *Cortex*, 42, 323-31.
- Coltheart, M. (This Issue). Levels of Explanation in Cognitive Science. *ASCS 2009 Proceedings*.
- Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford, Clarendon Press.
- Darden, L. and Maull, N. L. (2000). Interfield Theories. *Readings in the Philosophy of Science*. (T. Schick.) London, Mayfield Publishing Company
- Fodor, J. (1974). Special sciences, or the Disunity of Science as a Working Hypothesis. *Synthese*, 28, 97-115.
- Fodor, J. (1997). Special Sciences: Still Autonomous After all These Years. *Philosophical Perspectives*, 11, 149-63.
- Kim, J. (1992). "Downward Causation" in Emergentism and Nonreductive Physicalism. *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. (A. Beckermann, H. Flohr and J. Kim.) New York, Walter de Gruyter
- Kim, J. (1993). *Supervenience and Mind*. Cambridge, Cambridge University Press.
- McLaughlin, B. P. (1992). The Rise and Fall of British Emergentism. *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. (A. Beckermann, H. Flohr and J. Kim.) Berlin, Walter de Gruyter
- O'Brien, G. and Opie, J. (1999). A Defence of Cartesian Materialism. *Philosophy and Phenomenological Research*, 59, 939-963.
- Putnam, H. (1973). Reductionism and the Nature of Psychology. *Cognition*, 2(1), 131-46.
- Schier, E. (2007). *Emergence Colour and the Knowledge Argument: So what if Mary didn't know*. Doctoral Dissertation, The University of Adelaide.

- Searle, J. R. (1981). *Minds, Brains and Programs*. *Mind Design*. (J. Haugeland.) Cambridge, MA, MIT Press
- Shapiro, L. A. (2005). *The Mind Incarnate*. Cambridge, MIT Press.

### Citation details for this article:

Schier, E. (2010). The Explicable Emergence of the Mind. In W. Christensen, E. Schier, and J. Sutton (Eds.), *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science* (pp. 306-310). Sydney: Macquarie Centre for Cognitive Science.

DOI: 10.5096/ASCS200946

URL:

<http://www.maccs.mq.edu.au/news/conferences/2009/ASCS2009/html/schier.html>