

Searle on Emergence

Vladimír Havlík

The Academy of Sciences of the Czech Republic, Prague

Abstract: Searle's conception of ontological emergence is a basis for his explanation of mind and consciousness in the physical world. In this article, I try to show that a closer examination uncovers some possible ambiguities in Searle's conception of emergence. First, I try to show that Searle's distinction between emergent₁ and emergent₂ leads to a distinction between a strong and a weak interpretation of a causal consequence of interactions among constitutive entities and that from this point of view the existence of emergent₂ is improbable only in the strong sense. Second, I attempt to clarify Searle's distinction between *explanation* and *deduction* of consciousness in his claim for the non-deducibility of consciousness. At the end I try to show in what sense is Searle's concept of emergence loaded with a form of mechanicism, one which is being abandoned in more recent ontological conceptions.

Keywords: emergence, mind, consciousness, emergent property, system property.

In recent discussions of emergence, Searle's concept of emergent properties is classified as a mainstream theory of ontological emergence. Searle uses ontological emergence as the basis for his concept of mind and consciousness, which are, for him, entirely biological properties, fundamentally dependent on their biological bearer – the brain. Mind and consciousness belong only among the properties of the whole complex system, not of the constituents of that system. In order to explain his notion of the specificity of mind, Searle often uses an analogy with physical properties (Searle 1984; 1992; 1997; 2004). He refers to physical properties such as liquidity, solidity or transparency, which can be attributed to matter only at a certain macro-level of reality while for the micro-level constituents it is impossible to embody them. The molecule of water is neither “wet” nor “liquid” and it also

lacks many other physical characteristic of water understood as a macroscopic liquid, e.g. capillarity. Analogically, consciousness is a causal emergent property of the complex brain system at the macro-level and there is no such property at the micro-level of separate neurons. The role of physical analogies is important for our understanding of emergent properties as a common thing in nature. It is usual that complex systems embody novel properties or behaviors which their constituents do not embody at elementary levels. From this point of view, the existence of mind and consciousness is nothing special because there are causal powers which lead to these emergent properties. Their distinctiveness lies in the fact that they are pure biological properties and cannot be created out of their biological bearer.

Searle's physical analogy with liquidity has often been criticized. Liquidity cannot be in the same relation to water as consciousness is to brain, because liquidity can be deduced from the properties of elementary particles whereas consciousness cannot be deduced from the properties of neurons.¹ This objection seems to me somewhat questionable. It is true that physicists could explain why and under what exact conditions there occurs the phase transition of numerous molecules of water to liquidity whereas the exact conditions for the transition of a system of neurons to the state consciousness are not known. Yet I do not think that this could be the principal point of Searle's analogy. His argument lies in emphasizing the interconnection between micro-level and macro-level, such as the emergence of liquidity of water, and this is a common physical phenomenon, manifested at many different levels of reality.

Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense of "higher-level" or "emergent" in which solidity is a higher-level emergent property of H₂O molecules when they are in a lattice structure (ice), and liquidity is similarly a higher-level emergent property of H₂O molecules when they are, roughly speaking, rolling around on each other (water). (Searle [1992] 2002, 14).

¹ E.g. "What Searle fails to see is that liquidity can be predicted from the properties of elementary particles, whereas consciousness cannot be predicted from the properties of neurons." (<http://www.scaruffi.com/mind/searle.html>)

Drawing on this assumption, it is not surprising that a similar “mechanism” is present in the case of neurons and consciousness. Searle is well aware that we don’t know the exact conditions of emergence of consciousness; he presents the mechanism of emergence as an attempt to bridge the gap in our knowledge. He expects that a similar mechanism, common as it is in many other cases, will prove mind and consciousness to be biologically and physically natural things and the gap in our knowledge will be filled in the future by neuroscience and natural sciences in general.

These physical analogies have yet another important aspect. An explanation of all the macro-level properties of water as a liquid is in most cases possible only *ex post facto*, i.e. once we know them to be manifest at the macro-level, we can investigate under what conditions they occur. However, to ante-define or predict these system properties, i.e. to deduce them exclusively from knowledge concerning properties of micro-structure (e.g. from knowledge concerning the properties or features of the water molecule), is a completely different task. Moreover we now have what I take to be compelling evidence that it is impossible. From this point of view, the emergence of liquidity of water as well as the emergence of consciousness in the neuronal structure of the brain is similarly emergent.

Searle investigates the problem of emergence in more detail in *The Rediscovery of the Mind* (see Searle [1992] 2002, chap. 5). He considers emergent properties as a type of system properties (features) that are not, or not necessarily, properties or features of elements creating the system. E.g., the shape and the weight of a stone are properties which the molecules creating the stone do not have. In fact, Searle distinguishes two types of system properties – *deducible* and *non-deducible*² – in a manner very similar to J. S. Mill.³ Some system properties can be deduced, figured out or calculated from their composition and ordering, and sometimes from their relation to the environment. However, there are system properties which cannot be deduced or calculated from lower-level orderings of elements and environmental relations. This type of properties then must be explained in terms of causal interactions between micro-level elements. Searle calls them “causally emergent

² Searle does not use these terms to label this distinction.

³ For Mill’s distinction between homopathics and heteropathics laws cf. Mill ([1843] 2005, chap. VI, 242).

system features". To this type of properties belong the already mentioned macroscopic physical properties, such as solidity, liquidity and transparency (Searle [1992] 2002, 111).

Although Searle's main task is not to analyze emergence by itself but rather to use it only as a means for an explanation of the existence of consciousness in the physical world (which is an overall acceptable procedure), a closer examination uncovers some ambiguities. In the following, I will try to demonstrate them and clarify some of them, while leaving some other open. At the end an attempt will be made to show in what sense is Searle's concept of emergence loaded with a form of mechanicism.

I will start with Searle's famous claim about consciousness, what I call the *non-deducibility of consciousness*. He says:

The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro-level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them. (Searle [1992] 2002, 112)

While emphasizing the importance of the additional causal relations among entities constituting the system, he simultaneously presupposes that these additional causal relations must be explained by causal interactions at the micro-level. The requirement of explicability of these causal interactions at the micro-level leads Searle to a distinction between two types of causal emergence, which he calls emergent1 and emergent2. Searle defines emergent2 in the following way: "A feature *F* is emergent2 iff *F* is emergent1 and *F* has causal powers which cannot be explained by causal interactions of *a, b, c, ...*" (i.e. by causal interactions of system elements – V. H.; cf. Searle [1992] 2002, 112). By this definition, the distinction between the two types of emergence is dependent on the *ability to explain* a system property by causal micro-level interactions. However, the emphasis on *the explanatory capacity* could be misleading. Searle evidently does not want the distinction between emergent1 and emergent2 to be based in the epistemological availability of an explanation. Rather, the possibility or impossibility of an explanation ought to derive from the objective state of things, and in this particular case from causal relations. Thus it is necessary to understand the distinction between emergent1 and emergent2 in the following sense: "Feature *F* is emergent2 iff *F* has causal powers *which are*

not causal consequences of causal interactions (occurring between system elements) *a, b, c, ...*"

Although Searle introduces this distinction himself, he considers most emergent phenomena – including consciousness – to be of the emergent1 type. As for emergent2, he even voices some doubts whether it could exist, as he thinks that the existence of such properties “would seem to violate even the weakest principle of transitivity of causation” (Searle [1992] 2002, 112). In other words, whereas simple transitive sequence is valid for causal relations (“if event *c* is the cause of *d* and *d* is the cause of *e*, then *c* is the cause of *e*”) such a principle would be invalid in the case of emergent2. In the case of emergent1, the principle of transitivity obtains, which in Searle’s opinion implies that consciousness (as an emergent1 feature of the brain) is a causal consequence of neuronal interactions. Also, it follows from the definition for properties of the emergent2 type that some system properties (causal powers) are not causal consequences of interactions among constitutive elements, ergo could not be explained by these causal interactions. Unfortunately, this claim could give rise to the following misunderstanding.

First, if we adopt Searle’s distinction between emergent1 and emergent2 in the ontological sense, then the explanation of system properties becomes dependent on objective causal relations. However, what is now the meaning of the claim that *a system property is not the causal consequence of interactions among constitutive entities*? What is now meant by the claim that *the system has a property which could possess causal power while this property is not a causal consequence of the microstructure*? In my opinion it is possible to distinguish between a strong and a weak interpretation of this causal consequence. In the strong version, this claim means that the examined property is not dependent on the state of its microstructure, and given that the system is in some state *S*, it *may have* but *need not have* this property. The presence of a system property with regard to the system’s microstructure is arbitrary. In the weak sense, the claim means that the property is dependent on microstructure (i.e. given that the system is in some state *S*, it *either does have* or *does not have* the property) yet it is not deducible or calculable from the microstructure. In this case, the presence of the system property is not arbitrary and, in a sense, it is necessary. Thus the system property *P* is not a direct causal consequence of the only system’s microstructure, but it is causally dependent on it.

Unfortunately, Searle neglects to take account of the possibility that there is a strong and a weak interpretation of causal consequence. He tends to think that any property which is not the causal consequence of microstructure is always emergent² and its existence is improbable. However, once we admit the distinction between the strong and weak interpretations of causal consequence, it appears that the existence of emergent² is improbable only in the strong sense, while in the weak sense it is not only highly probable but in many cases even emergently evident.

However, Searle uses a similar distinction between *explanation* and *deduction* when he claims (in the passage quoted above): "The existence of consciousness *can be explained* by the causal interactions between elements of the brain at the micro-level, but consciousness *cannot itself be deduced or calculated ...*" (Searle [1992] 2002, 112, my emphasis). While the distinction between *explanation* and *deduction* may seem somewhat bizarre due to the Hempel's symmetry between *explanation* and *prediction*, the fact is that with emergent entities, there is, in practice, a temporal asymmetry between the two terms. We can *explain* something once it occurs (*ex post*), whereas we must be able *deduce* something even before it occurs (*ex ante*). What Searle wants to stress is the impossibility to know in advance which mutual causal relation between entities occurs and which properties (features) will emerge. However, emergent properties (features) which already occur and can be investigated, could be explained by the interaction of entities at the micro-level.

Second, the distinction between emergent¹ and emergent² depends on whether given property is a *causal consequence of interaction among elements or not*. In his claim for the non-deducibility of consciousness, Searle says that consciousness is not deducible "*from the sheer physical structure of the neurons without some additional account of the causal relations between them.*" He thinks that the dynamics of mutual causal relations of the constituting system is crucial for the existence of system emergent properties (in the sense of emergent¹). Given the fact that consciousness, as a case of emergent¹, is the causal consequence of neurons, is there any possibility to deduce it from their sheer physical structure and an *additional account of the causal relations between them*? It remains an open question whether Searle would accept that consciousness could be deduced from mutual causal interactions among neurons in the brain. While he claims that this is not possible without these causal relations, he also never says if it is possible at all, and in fact

he knows that it is not. We cannot deduce or calculate consciousness either from the sheer structure of neurons or from their mutual causal relations. I think that in speaking of an additional account of causal relations, Searle means not only mutual interactions between individual neurons but also those causal relations which occur only at the macro-level and need not even be recognizable at the micro-level. Although Searle never says this explicitly, I think he would never admit that the non-deducibility of consciousness impacts attempts to deduce consciousness from sheer structure only, and not from causal interaction as well. If this conclusion is correct, then the term “additional causal interactions” is a sort of mysterious incantation whose content is only to be speculated about.

However, there is a possibility that someone wants to connect non-deducibility of consciousness with an argument for irreducibility of consciousness (the fundamental part of Searle’s conception). He can then claim: “If there is something irreducible to its microstructure, then it cannot be deducible from it too.” It does seem to be a very seductive possibility. Yet unfortunately, it is, once again, not evident or convincing in the case of emergent phenomena. There are features that can have causal consequences at macro-level, could be calculable by simulation but at the same time cannot be fully reducible to its microstructure.⁴ While I do not have sufficient space here for a detailed argument, I do think that irreducibility cannot support non-deducibility.

I will now try to show that Searle’s conception of emergence includes a form of mechanicism, one which is being abandoned in the more recent ontological conceptions. Searle presupposes that entities of microstructure, such as molecules or neurons, are identical both as isolated entities and as entities involved in the establishing of the system as a whole. All that can change is the mutual causal relation, while entities by themselves stay rigid. Causal interactions between entities are the only cause and the only source of system properties or system features. Although Searle never analyzed these relations in detail, it is plausible that he would accept the following schema, founded on the transitivity of causality: 1) entities never lose their identity (such as shape, structure, features) and they manifest it even when involved in the establishing of a system; 2) entities and their properties or features are the cause of the causal relations which occur among them in the system; 3) thus,

⁴ E.g., the so-called weak emergence (see Bedau 1997).

these causal relations are the cause of the emergence of system properties. System properties are in principle explicable, yet at the same time system features and the system structures are not (directly) deducible or calculable from a mere knowledge of the constitutive entities. Based on the principle of transitivity of causal relations, we must assume that the sources of consciousness are causal features at the micro-level (i.e. electrochemical states of neurons in the brain). Emergence is a common mechanism in our world, and emerging novel entities such as substances, properties, relations and behaviors are thus situated at a higher system level from the original level of its constituents. This universal scheme of emergence is realized at various hierarchical levels of reality and it is responsible for the emergence of complex systems, from chemical compounds via life all the way to consciousness. Although Searle would probably agree with this scheme, he assumed that his concept of emergent1 is sufficient for an explanation of all emergent phenomena including consciousness. However, it is possible to demonstrate that there exist emergent physical phenomena for whose explanation the emergent1 concept is insufficient due its form of mechanicism. Many phenomena in physics demonstrate that entities are a different thing when they stand alone and when they take part in the creation of a system. These evidences are frequently brought up not only by physicists working on solid state physics and condensed matter physics (e.g. Anderson 1972, Leggett 1987, Laughlin 1998, Healey 2010). Recently, research in quantum theory has also provided similar evidence (e.g. Healey 1991, Silberstein - McGeever 1999, Kronz - Tiehen 2002, Hüttemann 2004).

We could maintain consistency with Searle's assumption regarding the identical mechanism of emergence of novel entities in nature, including consciousness, even in the light of the new concepts of emergence, such as the concept of "fusion" which claims that constitutional entities change radically when taking part in the creation of a system. What leads to the emergence of novel system features or properties is a fusion of elements together with their mutual causal interconnections and their interactions with the environment. We could claim that the same mechanism is responsible for, or at least present at, the emergence of consciousness, and that Searle's vision of emergence as a general mechanism is thereby not weakened but rather enhanced.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
havlik@flu.cas.cz

References

- ANDERSON, P. W. (1972): More is Different. *Science* 177, 393-396.
- BEDAU, M. A. (1997): Weak Emergence. *Philosophical Perspectives* 11, 375-399.
- HEALEY, R. A. (1991): Holism and Nonseparability. *The Journal of Philosophy* 88, No. 8, 393-421.
- HEALEY, R. A. (2010): Reduction and Emergence in Bose-Einstein Condensates. *Found. Phys.*, DOI: 10.1007/s10701-010-9481-8.
- HÜTTEMANN, A. (2004): *What's Wrong with Microphysicalism*. London: Routledge.
- KRONZ, M. F. – TIEHEN, T. J. (2002): Emergence and Quantum Mechanics. *Philosophy of Science* 69, 324-347.
- LAUGHLIN, R. B. (1998): Fractional Quantization. *Nobel Lecture*.
- LEGGETT, A. J. (1987): *The Problems of Physics*. New York: Oxford.
- MILL, J. S. ([1843] 2005): *System of Logic, Rationative and Inductive. Being a Connected View of The Principles of Evidence and the Methods of Scientific Investigation*. New York: Harper and Brothers, Publishers.
- SCARUFFI, P. (2006): *The Nature of Consciousness. The Structure of Life and the Meaning of Matter*. Omniware.
- SEARLE, J. R. (1984): *Minds, Brains and Science*. Harvard University Press.
- SEARLE, J. R. ([1992] 2002): *The Rediscovery of the Mind*. Cambridge: the MIT Press.
- SEARLE, J. R. (1997): *The Mystery of Consciousness*. New York: The New York Review of Books.
- SEARLE, J. R. ([2004] 2007): *Freedom and Neurobiology. Reflections on Free Will, Language, and Political Power*. New York: Columbia University Press.
- SILBERSTEIN, M. – MCGEEVER, J. (1999): The Search for Ontological Emergence. *The Philosophical Quarterly* 49, No. 195, 182-200.