

DUALIST EMERGENTISM

Martine Nida-Rümelin
University of Fribourg (Switzerland)

When consciousness arises in the phylogenetic or in the ontogenetic evolution of a biological system, something fundamentally new comes into existence. Once it has arisen, consciousness causally influences the functioning of the biological system that gave rise to it. These are typical emergentist ideas about consciousness. There are dualist versions as well as materialist versions of emergentism. I will focus in this paper on a dualist version of emergentism.¹ I will describe its intuitive motivation and sketch some arguments in its favor.²

1. Conscious Individuals and Consciousness Properties

The term “conscious” is used in many different ways. In one of its senses we can use the term to mark the fundamental distinction between those concrete individual things that have experiences (that have ‘an inner life’, ‘a point of view’, that are such that it is something like to be them) and the rest of concrete things or matter.³ Thus, the notion of “consciousness”, or, to be more precise, the notion expressed by the adjective “conscious” as it will be used in this text is applicable to individual things only. It marks a distinction between e.g. humans, dolphins and many other animals on this and hopefully some other planets on the one side and tables, stones, mountains etc. on the other. The term will not be used here to mark a difference between states of individuals or between processes or events. Furthermore, in the sense at issue a dolphin is a conscious being even while in dreamless sleep.⁴ The capability of having experiences is necessary and sufficient for an individual to be a conscious being.

¹ David Braddon-Mitchell argues in this volume against any emergentist view that tries to avoid dualism and yet to incorporate its intuitive merits. The present version of emergentism is not among the views he attacks since it implicitly endorses a dualist ontology. Contrary to Braddon-Mitchell I claim that the present dualist proposal does deserve the label ‘emergentism’ for reasons that will I hope get apparent in my description of the view.

² I am grateful to Terence Horgan, Brian McLaughlin and Barry Loewer for very helpful comments on an earlier version of this paper that motivated a number of changes and saved me from several mistakes.

³ It is hard to explain this general notion of consciousness non-metaphorically. This may invite the conclusion that we need a clear definition or at least some explication before we may use the term in philosophical theory. Of course it is in order to ask for clarifications in some sense, e.g. to ask for an analysis of how the term is conceptually related to other notions and of how it is distinct from similar concepts. But we should not expect too much. When we attribute some specific experience to an individual we thereby already presuppose that ‘it is something like to be that individual’. (The locution has been introduced by Thomas Nagel in his famous paper Nagel 1974.) Arguably, the general notion of consciousness at issue here is *conceptually prior* to any specific notion of any kind of experience. If this is so, then it should not be expected that someone might ever come up with any illuminating, non-circular definition of what it is to be conscious. But this does not mean that the term is in any sense obscure. To the contrary, or so I claim, we do have – upon reflection – an intuitive notion of what it is for an individual to be a conscious being which is quite clear and easy to grasp. Any proposed definition would have to be tested on the basis of this pre-theoretic intuitive understanding.

⁴ This needs to be pointed out since “conscious” is sometimes used in the sense of “awake”.

Conscious beings have properties that no individual without consciousness could possibly have. I will call these properties *consciousness properties*. I cannot use the common term 'mental properties' instead since it is controversial whether all mental properties require consciousness. Having a propositional attitude (having beliefs, desires, intentions, and the like) is an example. If propositional attitudes can be defined in functional terms, and if any property that can thus be reduced to its causal role does not require consciousness for its instantiation, then having a propositional attitude is not a consciousness property.⁵ Neither can I use the term 'phenomenal properties' to replace the term 'consciousness properties'. As will be explained below (see section 5), the property of being active by doing something is a consciousness property in the sense just explained. Only conscious individuals can be active in the relevant sense. But it is quite clear that properties that consist in being active in a particular way (e.g. the property of running or the property of taking a decision) are not phenomenal properties.⁶

2. The evolution of consciousness

At some point in the evolution of life some specific pieces of matter got arranged in a way that led to the occurrence of consciousness. At some point in the development of individual humans and other conscious animals the same kind of change takes place. This radical change may be interpreted in two ways. According to the first interpretation, the change involves new individuals, conscious beings, coming into existence (this is the view I favour and the one the substance dualist accepts). According to the second interpretation, no new individuals come into being. Rather the organism at issue acquires qualitatively new properties, consciousness properties. The emergentist believes that this change occurs as a result of physical conditions satisfied by the biological system. A certain arrangement of matter leads with nomological necessity to the existence of conscious individuals with qualitatively new properties. The following two claims partially characterize a substance dualist version of emergentism:

Claim 1 (Emergence of new Individuals): There are specific physical conditions C such that the following holds: at any time t, if t is the time at which a particular material system M (e.g. a biological organism) first satisfies C, then with nomological necessity a subject of experience (a conscious being that belongs to an ontological category different from the one of material objects) comes into existence at t and starts at t to have M as its body.

Claim 2 (Emergence of Consciousness Properties): A subject cannot have consciousness properties unless the subject's body has corresponding physical properties.

⁵ In my opinion the first claim is wrong and the second true.

⁶ This is not to deny that having these properties is accompanied or even requires some specific phenomenology. - I use the term "running" as a description of an activity and not as the description of a kind of bodily movement. In this sense, no non-conscious robot can run, only subjects of experience can run or swim or do anything (see section 5 below).

No change in consciousness properties is nomologically possible without a simultaneous change in corresponding physical properties of the subject's body. No two nomologically possible individuals (whether in the same world or in different worlds) can differ in their consciousness properties without a difference in the physical properties of their respective bodies.

Both of these claims need some explanation and additional remarks:

- Nomologically possible worlds are worlds with the same laws of nature as the actual world. These laws include psychophysical laws that are – according to the emergentist view here presented – fundamental laws of nature.
- According to claim 1, subjects of experience are a product of nature. The existence of a subject without a body which satisfies the physical conditions C is nomologically excluded. Subjects of experience come into existence when their body satisfies certain physical conditions. This implies that the conscious being at issue has not existed before: reincarnation is nomologically excluded.⁷
- Claim 1 is a substance dualist claim: the conscious being that comes into existence at t₁ is not identical to the system that gives rise to its occurrence. According to this view the occurrence of consciousness is more than the instantiation of qualitatively new properties. The occurrence of consciousness requires the coming into being of individuals belonging to the special ontological category of experiencing subjects.
- To call the view 'substance dualist' is not meant to imply that there are two kinds of stuff involved (see section 3). It is however meant to imply that the subject is something over and above its body in a sense in which a statue is *not* something over and above the corresponding lump of clay.
- These new individuals have the system at issue as their body. What is it for a subject S to have the organism O as its body? This means, roughly, that (a) the consciousness properties of S causally depend in the right way on the physical properties of O (e.g. if O is damaged, S feels pain) and that (b) S does what it does with the organism O (e.g. O's hand goes up, if S raises its hand).⁸
- According to claim 1, the occurrence of a conscious being is nomologically necessitated by the conditions C. An emergentist might however consider the possibility that there is – within limits – a certain amount of real chance involved: it might not be nomologically determined at what point exactly the individual at issues comes into existence. The claim could be reformulated accordingly (but for simplicity I will not include this complication here).

⁷ A weaker claim that one might still call emergentist would be silent about whether subjects can change their body and about whether subjects can exist without a body. This weaker claim would only state that certain physical conditions C are nomologically sufficient for there being a subject that starts at t to have the system as its body.

⁸ It is tempting to say that S's activities cause certain changes in O. But this would not be quite right in many cases. My raising my hand does not cause my hand to go up; rather it is partially constituted by my hand's going up.

- Talk of 'physical' conditions should be understood in a broad sense. Biological, chemical as well as functional properties are included. Claim 1 is compatible with the idea that the occurrence of consciousness depends only on functional properties of the system. It is thus compatible with the claim that conscious individuals might occur on the basis of a non-biological system made up of some non-biological stuff.⁹
- Claim 2 states a close dependence between consciousness properties and physical properties. It is impossible e.g. to take a decision without a simultaneous change taking place in the body (presumably the brain). Mental events need a physical basis. However, claim 2 does not imply that consciousness is in any sense causally inert. Claim 2 is compatible with causal influence in both directions: physical changes cause changes in consciousness properties. Claim 2 allows for the possibility that the subject itself influences via simultaneous causality the processes in its brain by taking decisions, considering hypotheses, directing its attention and moving its body (see section 5 below).
- The ideas formulated in claim 2 can be captured in part by stating a thesis of strong nomological supervenience: there are no nomologically possible worlds w_1 and w_2 , and subjects s_1 and s_2 and times t_1 and t_2 such that there is no physical difference between s_1 at t_1 in w_1 and s_2 at t_2 in w_2 but yet there is a difference between the two subjects at the times in these two worlds with respect to their consciousness properties.¹⁰

3. Substance Dualism

With the emergence of consciousness new individuals of a special ontological category, conscious beings or subjects of experience (I will use these terms interchangeably) come into existence. A philosopher who accepts this claim (formulated above as claim 1) endorses some version of substance dualism. Substance dualism is often presented in a way that makes the view appear clearly unacceptable and quite ridiculous. It is therefore necessary to make a few remarks to avoid possible misunderstandings.

- Substance dualism – as I use the term – is characterized by the claim that the subject of experience (the thing that has consciousness properties, the thing a person refers to using the first person pronoun, the thing people refer to using a name of a person

⁹ Claim 1 thus allows for the possibility that future robots will give rise to the occurrence of consciousness. However, once the subject of experience at issue were to become active (see section 5 below) its body would cease to be a mechanically functioning system.

¹⁰ Emergentism is often characterized by the combination of a thesis of metaphysical supervenience with an anti-reductionist claim. The dualist emergentist, however, has no reason to endorse metaphysical supervenience. It is often assumed that dualism can be partially characterized by the denial of metaphysical supervenience. But while it is obvious that the denial of metaphysical supervenience implies dualism I doubt the reverse implication. The issue of metaphysical supervenience is therefore left open in the present characterization of emergentist dualism. For the role of supervenience in an explication of emergentism see Beckermann (1992), Kim (1999), Stephan (1997) and Stephan (2002).

or a name of an animal) is not composed of matter. The experiencing subject, according to this view, is not a body or a brain or a system composed of anything; nor is it an abstract entity. What a subject of experience is can best be positively characterized by saying that it is capable of having consciousness properties and by describing the special ontological status of its identity across time and of its identity across possible worlds.¹¹

- For contingent historical reasons substance dualism is often associated with the view that animals are mere automata (mechanically functioning bodies) while the human animal alone has a different ontological status. There is no systematic reason for a substance dualist to be tempted by this idea. We know that we are not alone in the animal kingdom in being conscious.¹² But it is the mere fact of the existence of consciousness in a particular given individual A that justifies a substance dualist view with regard to A. So a reasonable substance dualist will not restrict his or her claim to the human case.
- The present version of substance dualism does not imply that the human person is composed of a material and a non-material part, a body and a soul. According to the present version of substance dualism there is no need to talk of composition in this context. The person *is* the subject of experience and he or she *has* a body. A given agglomeration of molecules A is the body of the subject of experience S if and only if there is the right kind of relation between the experiences and the activities of S and physical changes and/or movements within/of A. (If for instance A is damaged, then S feels pain and if S is engaged in running then A moves in a 'running way').
- According to a traditional religious view a person *has* a body and *has* a soul. According to the present view a person has a body but no person *has* a soul. At best persons *are* souls. However it would be misleading to use the term "soul" in the description of the kind of emergentist dualism I have in mind. The soul is supposed to be able to exist without a body. The emergentist dualist does not endorse the claim that subjects can exist without having a body. The soul is supposed to be immortal. The emergentist dualist view does not include the claim that subjects of experience cannot cease to exist. The soul is often thought of as being composed of some thin non-material stuff. The emergentist substance dualist does not postulate the existence of thin immaterial stuff. The soul is often thought of as being able to be located in space (people think of the soul as something that can leave the body and

¹¹ The special ontological status of identity across time of conscious individuals is the topic of my book "Der Blick von Innen. Zur transtemporalen Identität bewusstsensfähiger Wesen", Suhrkamp 2006 (an English translation is in preparation and will be published under the title "The View from Inside".)

¹² I am leaving it open how far "down" in the animal kingdom there are conscious beings. - Of course there is an interesting philosophical problem about how the claim of the existence of consciousness in other individuals can be justified. But there is also an interesting philosophical problem about how our belief in the existence of the external world can be justified. It would be inappropriate and irrational if someone withheld opinion about the existence of the external world as long as no generally accepted justification has been explicitly developed. The same is true for the case at issue.

fly away). The emergentist dualist should resist these ideas with respect to subjects of experience. Subjects of experience are located only in the sense of having a body with a spatial location.

- The emergentist substance dualist claims that Peter, a subject of experience (a person) is not identical to his body. This claim, however, does not imply that Peter is in any sense 'hidden' in his body. The substance dualist can endorse the natural view that you see *me* when you look at my body. The substance dualist can justify this claim in the following way. When looking at my body (e.g. into my face) you can see non-inferentially that I have certain properties. This is why it is appropriate to say that you see me by looking at my body. You can see non-inferentially that I am laughing by looking into my face while I am laughing. When you look into my face while I am laughing you see *me* laughing.
- Another misleading term often used in this context is the term "the Self". Here again it is sometimes said that a person *has* a self. The emergentist substance dualist position I wish to defend rejects this idea. The referent of the term "the Self" is the referent of the first person pronoun used by some person. But the referent of the first person pronoun used by a person P is simply P. So 'the self' (in a given case) is simply the person (or the subject of experience).
- Talking of 'the self' in this context invites another idea that we should reject. Some philosophers argue that selves are somehow constituted by their capacity to refer to themselves in "I-thoughts".¹³ But subjects of experience may and do exist without having the capability of entertaining I-thoughts.

The idea of subject causation developed below does not make sense if the subject is numerically identity to its body or a part of the body. Therefore, every argument for subject causation is also an argument for substance dualism. The most powerful arguments for substance dualism are related to the philosophical problem of identity across time.¹⁴

4. Qualitatively new properties

According to the view defended in this paper new *individuals* come into existence when consciousness arises. But contemporary discussions about emergentism focus on a different idea of novelty. As emergentism is commonly understood the novelty brought about by the occurrence of consciousness consists in the instantiation of qualitatively new properties by individuals that already existed before. According to widespread opinion conscious individuals are biological organisms; consciousness properties thus are properties of a

¹³ A view of this kind may be found in Lynn Rudder Baker (2000) and in Jonathan Lowe (1996).

¹⁴ For reasons of space these arguments cannot be presented in the present paper. Some of these arguments are developed and discussed in detail in M. Nida-Rümelin (2006a).

complex material thing that is composed of smaller parts (e.g. of cells that stand in a great variety of causal interactions). With this in mind the question about the relation between consciousness and its physical basis becomes a question about the relation between properties of the whole organism and properties and relations instantiated by the smaller parts that make it up. The emergentist is portrayed as saying that consciousness properties are emergent in this sense: they are properties of a whole that cannot be reduced to the properties of its parts and the relations between them but they ‘emerge’ on the basis (because of) the properties of the parts and the relations obtaining between them. The intuitive idea is that the instantiation of emergent properties of a whole *consists in more than* in the instantiation of the property of being composed of parts that satisfy certain conditions. The shape of an object and its weight are examples of properties of a whole that are *not* emergent. The functional properties of a biological system provide further examples of non-emergent properties of a whole.

It is not a trivial task to account in a precise manner for the distinction between emergent properties of a whole and non-emergent properties of a whole (relative to certain micro-properties and micro-relations of micro-parts that make up the object).¹⁵ An emergentist who accepts that conscious individuals are biological organisms must however rely on some account of this distinction in order to give a precise meaning to his claim that consciousness properties are qualitatively new properties.

But this approach to the emergentist idea of novelty is erroneous according to the view defended in this paper. One reason is that – according to this view - subjects of experience are not composed. So the sense in which some of their properties are qualitatively new or emergent cannot be understood according to the model of emergent properties of composed wholes.¹⁶ Another and maybe more important reason is this: to explain the novelty intuition along these lines misconstrues the character of the change that takes place when consciousness arises. Let me try to explain.

Suppose consciousness arose for the first time on our planet in the moment in which a particular quite primitive organism somewhere in some ocean began to feel comfortable warmth when it moved by chance into warmer water. According to the emergentist an astonishing and radical change took place in this moment. But let us ask what exactly it is that makes the change a radical change and a change that deserves amazement. It is not the instantiation of the particular phenomenal property of feeling warmth. What makes the change amazing has nothing to do with this special phenomenal character. Rather, the

¹⁵ See e.g. Ansgar Beckermann (1992).

¹⁶ According to the view I have in mind here, subjects of experience are not composed of matter but also they are not temporally extended (they do not have temporal parts although they persists through time).

astonishing fact is this: since, as we assumed, a feeling of warmth has occurred, there is 'someone' who feels the warmth. The fact that 'someone' came into existence is the astonishing aspect of the change and the aspect that makes the change a radical change. Before the first occurrence of a faint feeling, no one was on our planet to experience the world. In that moment, a subject capable of experience came into existence.

According to this view any justification of the emergentist claim that a radical change has taken place when consciousness occurred must be based on the fact that the first instantiation of consciousness properties requires the coming into being of subjects of experience. It is this coming into being of conscious individuals that deserves astonishment. If this is correct then an appropriate formulation of the emergentist intuition of novelty does not require, contrary to what is commonly assumed, any general theory about reduction. We need not define what it would be for a property to be irreducible to other properties or relations in order to explicate what makes consciousness properties qualitatively new and radically different from physical properties. We can understand what makes these properties qualitatively new by taking into account the following two elements: (a) the instantiation of consciousness properties requires that the instantiating being is a subject of experience and (b) subjects of experience are radically different in kind from all other kinds of entities. The main task if we wish to get a clear understanding of the novelty of consciousness is then to get clear about the special ontological status of conscious individuals.¹⁷

Given what has been said in this section we can add the following claim to the characterization of dualist emergentism:

Claim 3 (Qualitatively New Properties)

Consciousness properties are qualitatively new properties. The instantiation of consciousness properties does not consist in the instantiation of physical properties of parts of the organism at issue and/or relations between them. The novelty of consciousness properties is due to the fact that they are instantiated by subjects of experience which are not identical with any physical thing.

¹⁷ At this point someone might reply in the following way: the difference is simply that the former but not the latter can instantiate consciousness properties. So the task is quite trivial. We can make a list of consciousness properties and say that a being is conscious if and only if it is capable of instantiating at least one of these properties. The list will be an open list and we might want to add something like 'or properties similar to those on the list'. There will be a certain amount of arbitrariness, so the reply might go on, in this procedure. There simply might not be any fact of the matter about whether a given property deserves to be added to the list. Accordingly in many cases there is no fact of the matter as to whether a given being is a conscious being. According to this view, to understand the difference between conscious individuals and other concrete individuals is to have an appropriate list of this kind in mind and to know that a being is conscious if and only if it is capable of instantiating some of these properties.

This proposal is fundamentally misguided according to the view I propose. For each property it is a substantial factual question whether it should be added to the list. The answer depends on whether the property requires a subject of experience for its instantiation. Furthermore, the list cannot be used to clarify what it is to be a subject of experience. If there are only consciousness properties on the list (as it should be) then to understand what having one of the properties on the list consists in already requires a grasp of what it is to be a subject of experience.

5. Subject causation

One important characteristic of conscious individuals is their capacity to have experiences like the feeling of warmth, the visual experience of an approaching object, the complex experience of listening to a piece of music and the like. To have a particular experience sometimes involves being active in a specific way. Listening to a piece of music attentively and with the intention to enjoy its particular musical qualities involves e.g. directing one's attention towards aspects of the piece. In this sense many experiences are not passive. The subject of experience is itself active in the experiencing. The same holds for thinking. To think about a philosophical puzzle involves actively considering different theoretical possibilities and actively directing one's attention upon a specific subpart of the problem. Subjects *do* something in their experiencing and in their thinking. Subjects of experience are even more obviously and more visibly active in their bodily doings. Only some of our doings are actions but all actions are doings. Doings can be 'mental' (e.g. the forming of an intention, the direction of attention in thought or perception) or they can involve bodily movements (e.g. turning one's head, smiling, walking). Not all our movements and bodily changes are doings (breathing can be automatic; digestion involves movements but is not a doing). There might not be a sharp line between doings and non-doings and there might not be a sharp line between actions and other doings. I will not presuppose any particular view about the difference between actions and mere doings. Although actions are (on our planet) probably restricted to the human case, doings are not. Many animals (maybe even all animals) are active in their doings too.¹⁸

With these preparations a further characteristic of conscious individuals can be described. Doing something requires a subject of experience who does the doing. Only conscious individuals can be active in the relevant way. To be active in the relevant way means - according to the view here proposed - that the subject is itself a causal origin of what happens. The subject is a causal origin of changes in the brain when it directs its attention to a particular aspect of a problem and it is a causal origin of changes in the brain that initiate and that uphold a movement when the subject does something involving a bodily movement. I will call this kind of causation *subject causation*.

¹⁸ Actions are normally (but maybe not always) done for some reason. Doings are often done without any reason. While sitting in a train a person may turn her head from time to time. These movements are doings (it is the person who does them, the movements are not in that sense automatic) but they can be done without any reason. In actions we are normally in some way aware of what we do. We need not be aware of our doings. While giving a talk a person might move her hands without being aware of the fact that she does. - I leave it open here whether the capability of having an experience can occur without the capability of being active and vice versa. I also leave it open how far down in the evolution animals are active in the sense at issue.

A similar idea is known under the heading 'agent causation'. Some incompatibilists with respect to determinism and human freedom defend the view that in acting the person is itself the cause of some event.¹⁹ According to their view, agent causation is not to be confused with event causation. The person is not an event, but the person is a (or the) causal origin of her action (or of some event preceding or accompanying her action). So the causal relation does not obtain in this case between two events. The view here proposed has some similarity to these theories. Like agent causation theorists, I claim a causal relation between subjects and events that are caused by the subject and I subscribe to the view that the causal relation at issue is different from event causation. However, there are also several important differences:

First, the idea of agent causality is normally assumed to be restricted to the human case and it is assumed to occur only in the context of human action. The view here proposed is in a sense more radical than this. It includes the claim that conscious individuals in general are active in all their doings. The claim is thus neither restricted to action nor to the human case.²⁰ According to the view I advocate the jump of a squirrel or the barking of a dog are examples where 'someone' is active in the sense of subject causation.

Second, agent causation theorists normally think of the person as intervening at a given isolated moment t . Up to that moment t several options are open (it is causally underdetermined which of them will be realized). The intervention of the person results in a realization of one of these options. After t things develop in the normal causally determined manner. According to this picture agent causation is a temporally quite local phenomenon. The agent intervenes at specific points in time while leaving the rest of the causal chains intact. This is not the picture I wish to propose. I think of subject causation as *continuously* and *simultaneously* influencing some of those physiological events in the brain that are the basis of mental doings (like considering a hypothesis or directing one's attention) or of bodily expressed doings (like jumping out of joy or playing a piece of Mozart on a piano). According to this view the events resulting from subject causation are not to be thought of like exceptional isolated miraculous 'little bangs'. Rather, subject causation is present virtually all the time while a conscious being is awake and it continuously influences in a complex way what happens in the brain and in the subject's body. When at some point in its development a brain brings it about that a subject of experience comes into existence then the brain itself thereby undergoes a fundamental change. It ceases to be a physically determined system

¹⁹ Agent causation theorists disagree about the kind of event that is directly caused by the agent. For a brief survey see section 2.4. in Timothy O'Connor (2002).

²⁰ It could be claimed that all conscious individuals are in fact or even necessarily active. This claim is plausible but I cannot see how the speculation could be justified in a convincing way.

and a great variety of processes underlying the many activities of the subject develop in a way in which they could not develop without the subject's causal influence.

Third, libertarians (who believe in human freedom and in the incompatibility of freedom with determinism) sometimes seem to think that in the case of free action a *preceding* mental event (e.g. a decision) causes a *later* physical event. Agent causation theorists (who are libertarians of a special sort) sometimes seem to endorse this view too (adding that the person herself causes the mental event which in turn causes the physical event). If this idea of a preceding mental event causing a later physical event is combined with the dualist claim that the mental event is non-physical then the following picture emerges: there are non-physical mental events that happen without any physical basis. The corresponding physical change happens only a bit later. The present view does not imply this problematic result. The result is avoided by the idea of simultaneous causation. In taking e.g. a decision the person simultaneously causes changes in her brain. In general, the person cannot cause anything without thereby *simultaneously* causing a change in her brain.²¹

Forth, if the problematic picture just sketched were correct, then consciousness properties would not supervene on the physical. There could be two individuals at some time *t* with the same physical properties that differ with respect to the decision they take at *t*. Some libertarians and some agent causation theorists therefore seem to be forced to deny the claim of nomological supervenience of consciousness properties on physical properties. The present view about subject causation does not exclude nomological supervenience. To the contrary, the overall view here proposed explicitly endorses nomological supervenience of consciousness properties on physical properties or more precisely on neurophysiological properties of the brain (compare claim 2 above). This claim of nomological supervenience is well-motivated within the present approach: (a) all differences with respect to phenomenal properties are brought about by differences in physiological properties of the brain (brain processes cause the instantiation of phenomenal properties) and (b) the subject's activities are always accompanied by corresponding physiological processes since the subject cannot cause anything without causally influencing processes in its own brain.²²

According to the emergentist view here proposed the subject can causally influence physical events happening in its own brain. These physical events would not occur if the subject was not active in the relevant way. It follows that these events are not causally determined by preceding physical events. There is no overdetermination involved here. Subject causation is

²¹ The term 'change' might invite misunderstandings. The person can cause that a certain state continues, she then causes (in a sense) that no change takes place. (Of course the upholding of a particular state involves neural activities that again involve a great variety of changes.)

²² The theoretical motivation for the acceptance of some supervenience claim is normally quite different. Philosophers hope to express some form of dependence of the mental upon the physical without thereby endorsing any causal relation between the mental and the physical. Contrary to this, the claim of nomological supervenience is combined in the present view with a form of interactionism: the subject itself causes physical changes and physical changes cause the instantiation of certain consciousness properties by the subject. Compare for the discussion of supervenience Kim (1993) and McLaughlin (1995).

incompatible with the claim that every brain event has sufficient physical causes and it is also incompatible with the claim that all brain events have only physical causes. Since brain events are physical events, the thesis of the existence of subject causation here proposed is incompatible with the principle of the causal closure of the physical.²³

6. Causal relevance of consciousness properties

The causal relevance of consciousness properties is an empirical question to be treated in psychology and in neurophysiology. The philosopher, however, can and should contribute by describing various ways in which consciousness properties could in principle be causally relevant to the behaviour of the subject and to the development of brain processes. A philosopher who accepts the version of dualist emergentism proposed in this paper has the additional task to explain how he or she can account for the obvious causal relevance of consciousness properties in a way that fits into his or her overall view and that in addition does not contradict and is in the best case already supported by available empirical data. In this section, I will only mention a few elements that would have to be developed in a more comprehensive presentation of dualist emergentism.²⁴

In some cases the causal relevance of given instantiations of consciousness properties is due to subject causation. In these cases the instantiations of consciousness properties owe their causal relevance to the fact that the subject is active in a particular way.

Example 1: I see an apple in front of me and I desire to eat it. I reach out for the apple.

In this case I act on the basis of a conscious perception with a particular content and on the basis of a conscious desire. My conscious perception and my desire do not directly cause my act. But they are both causally relevant: I would not have acted in the way that I acted if I did not have these consciousness properties. It is me who does the arm movement by causing specific changes in the brain. But I would not cause these changes if I did not have those consciousness properties.

To say that in this case the perception and the desire are both causally relevant might invite the idea that there are three partial causes: the subject (who causally influences certain processes by way of subject causation), the perception and the desire (which causally influence the result by way of event causation). If we think of partial causes as cooperating to

²³ The principle of the causal closure of the physical as I understand it here states the following: for every physical event E, if X is a cause of E, then X is physical too. Subject causation is, however, compatible with the following principle of causal closure: for every physical event E1, if the event E2 is a cause of E1, then E2 is physical too.

²⁴ I hope to develop some of this in my paper "Doings and Subject Causation", in preparation for a special volume of *Erkenntnis* edited by Michael Esfeld, Albert Newen and Vera Hofmann.

bring about a result like several fires may ‘cooperate’ to warm up a room, then this picture is quite clearly inappropriate. The perception and the desire do not do any causal work in addition to motivating me to act in a particular way.²⁵

But this way of describing what happens in example 1 may cause still another possible misunderstanding. If the desire and the perception do not have any causal impact by themselves but owe, as I said, their causal relevance entirely to subject causation, then one might be tempted to conclude that the present view implies the following claim: the subject brings it about all by itself that the processes in the brain responsible for the triggering of the movement occur and develop in the way they do. But this cannot be so. Learned motor programs realized in the brain are obviously necessary for me to be able to reach out for the apple. I do not cause the brain process that triggers the movement out of nothing. The present view does not imply the denial of the following obvious truth: complex physiological processes have to occur in the preparation of any bodily movements and these complex physiological processes are in great part predetermined by the ‘programs’ realized in the motor cortex. The present view only implies that subject causation is a necessary condition for the occurrence of these brain processes in a given case.

Another way in which consciousness properties can be causally relevant has to do with the programming of motor programs and other programs in the brain.

Example 2: While practicing, Anna, a pianist, carefully listens to the sounds she produces with her fingers. The way the piece sounds to Anna will influence in a complex manner the way Anna moves her fingers. After a few months Anna will be able to play the piece in a way that conforms to her musical judgement.

We sometimes say that a pianist plays a piece ‘automatically’ which is in a sense partially correct. It is impossible to have the movements of the fingers in all musically relevant aspects under conscious control while playing a complex piece rapidly. Often the pianist would not know how to play on without ‘leaving the fingers decide’ what to do next. (This is why in order to remember the movements you go back some steps in the piece and ‘let the fingers do’.) But still the movements are not automatical in this sense: the pianist does the playing. Anna is herself causally relevant for the physical events that bring about the movements. She causes the movements by way of continuous and simultaneous subject causation. But, of course, she has no conscious control over all the relevant details. She causes the movements of her fingers, but the way the fingers move depends in great part on a learned program. The learned program, and this is what the example is supposed to illustrate,

²⁵ Motivation cannot be explicated in terms of causation. That certain psychological preconditions motivate a person to act in a particular way does not mean that they cause the person to act in that way.

depends itself on consciousness properties. The learned program would be different if Anna had played the piece differently in the past. And she would have played the piece differently in the past if it had not sounded to her the way it did while she played. We may say that the way the piece sounded to her motivated Anna to move her fingers in a specific way. But the case is quite different from example 1. In example 1, the agent decides to reach out for the apple on the basis of what she sees and wants. There is no time for taking decisions about finger movements in the piano playing case. You just listen carefully, attend to the musical qualities you are interested in and try to make it sound a particular way. Often you do not know what exactly you do when it begins to sound all right.²⁶ But the fact that it begins to sound all right is causally relevant: you will try to do it again this way next time and if you succeed then a new detail of the program is beginning to be implemented in your brain.²⁷

We often say that we did something automatically when there was no time to think.

Example 3: John is lost in his daydreaming while driving a car. He almost overlooks a red light. He sees it just in time to jump on the break. There is no time for reflection. He jumps on the break without any thought intervening between the seeing and the jumping.

If John had not seen the red light he would not have jumped on the break. His seeing of the red light is causally relevant for his jumping. Still, even in this case, we should not say, or so I claim, that the seeing caused the jumping all by itself.²⁸ Even in a case of a rapid reaction it is still the person (or subject) who does the doing. Jumping on the break is a doing (even if one doubts that it is an action). If it is a doing, then the person causally brings it about that the body makes the movement. The fact, however, that John can react so quickly and without reflection in the right way is due to a program developed in a previous process of learning.²⁹

The three examples considered so far seem to suggest that the causal relevance of the instantiation of a consciousness property is either due to subject causation (in this case a consciousness property or event inherits its 'causal powers' from subject causation) or it is due to the physiological processes underlying the consciousness properties (in this case a consciousness property inherits its causal powers from the causal powers of the underlying physical process). This naturally raises the question of whether consciousness properties

²⁶ To say that you do not know exactly what you do is to say that you would not be able to give an independent description of the movements. Of course this is the case too in virtually all our daily actions. We have to do some reflection in order to describe the movements we execute when we pull water in a glass or open a window.

²⁷ Something like this happens in most cases of acquisition of motor skills. The joy of a child that learns to walk when it realizes "now it works!" is causally relevant for the learning just like the phenomenal character of the sounds is in Anna's case.

²⁸ One might think that the seeing is nonetheless a cause of the jumping since it causes a reaction in the subject. But I hesitate to agree. The subject's causal influence is not caused by any preceding event, not even 'in part'. The subject does the doing *on the basis* of the perception where – even in this case – the 'on the basis'-relation is not to be confused with being caused.

²⁹ I am not claiming that every such program that we 'rely on' in our doings is due to learning. There may be a great variety of innate motor programs that still require an active subject to be 'executed'.

can have causal powers of their own, causal powers that are not inherited (neither from subject causation nor from physical event causation). A potential example might be the case of an insight on the basis of thinking.

Example 4: Elisabeth has been thinking about a philosophical puzzle again and again for many weeks. One morning suddenly a simple solution pops up in her mind. Elisabeth carefully considers this way to solve the puzzle and finds it intuitively highly attractive. Elisabeth forms the belief that this is the correct solution.

The fact that the solution appears intuitively correct to Elisabeth is causally relevant for the formation of the belief. This claim should be true according to any acceptable theory about thinking and believing. If we exclude (as I think we should) that this is a case of over-determination then either the instantiation of the consciousness property (being an intuition with a specific content) has its own causal powers (over and above the causal powers of the instantiation of physical properties by the underlying brain processes) or its causal power is entirely due to subject causation (like in at least some of the other cases considered before). But the latter possibility is excluded: the formation of belief (maybe in contrast to acceptance) is not an action and it is not a doing. So the causal relevance of intuitions for the formation of belief cannot be due to subject causation. But intuition should not be epiphenomenal. I therefore tend to think that consciousness properties in some cases have causal powers of their own.

Let me summarize the theses developed in the last two sections:

Claim 4 (Subject Causation):

Whenever a conscious individual does something then it is itself a causal origin of the doing. This causation (subject causation) is not a case of event causation. Subject causation is continuous and simultaneous causation.³⁰ Subject causation is incompatible with the causal closure of the physical.

Claim 5 (Causal Powers of Consciousness Properties)

In many cases consciousness properties are causally relevant. They often (but not always) owe their causal powers to subject causation and/or to the underlying physiological processes.³¹

³⁰ By calling subject causation ‘continuous and simultaneous’ I mean to express the idea that subject causation is not to be thought of as an initial cause of a physical process but rather as an influence stretched out in time while the physical process is happening. Subject causation is ‘continuous’ in the sense that a whole physical process between t_1 and t_2 is brought about by a subject who is active between t_1 and t_2 and it is simultaneous in the sense that details about the physical process realized at t' (between t_1 and t_2) are caused by the subject at t' .

³¹ It should be added here that causal powers owed to the underlying processes are not genuinely causal powers of the consciousness properties at issue. I am convinced by the reasoning developed by Jaegwon Kim according to which consciousness properties would be epiphenomenal if all their causal powers were ‘inherited’ from the causal powers of the physical. A parallel reasoning however does not apply to causal powers that are due to subject causation.

7. Why believe in subject causation?

If you observe a squirrel jumping from one branch of a tree to another, then the squirrel does not look to you like a mechanism that jumps as the result of some inner 'mechanical' process. It looks to you as though the squirrel itself, the subject of experience, does the jumping. When you see a conscious individual that looks around, sits down, turns its head in the direction of a noise, then you do not see these events as the result of a mechanical process. We see the movements of biological organisms that we implicitly accept to be conscious as being done by the conscious individual itself. A related claim is true for the way we experience our own doings. We experience our doings as brought about by ourselves. To assume that some inner processes cause our doings is incompatible with the content of the phenomenology of our experience.³² If these experiences of ourselves when we are active and our perceptions of others as being active are not illusory, then conscious individuals *are* active in their doings. They are not, in that sense, biological 'automata'. We should not be ready to accept a philosophical theory that implies that our way to perceive the world (ourselves and other conscious beings) is fundamentally mistaken in a radical way. This is, in a nutshell, the most powerful argument, I claim, for the acceptance of subject causation.

There is no room here to defend the view in detail. But let me mention the elements that would have to be present in an elaborated version of the argument. (a) The content of the experiences at issue must be further analysed. It has to be shown that the experiences just mentioned really have the representational content that I just claimed they have: they represent the other subject (or ourselves) to us as being active in the sense of subject causation and thus in a way that is incompatible with the assumption of causal determination. If this is the correct analysis of the content of our experiences then it follows that our daily experiences cannot be veridical unless conscious individuals really are active in the sense of subject causation. (b) The experience of others as active and of ourselves as active is deeply entrenched in our world view and in our emotional and intellectual life. In a second step it has to be shown that this fact justifies us in taking the corresponding conviction (the conviction that subject causation is real) as epistemically central in this sense: we should not be ready to abandon the claim of subject causation unless we are forced to do so by extremely powerful counterevidence. (c) In a third step the argument has to show that there is no such extremely powerful counterevidence. In particular it has to be argued that, contrary to what many people assume, there is no powerful *empirical* evidence for the non-existence of subject causation.

8. The Adequacy of Amazement

³² A similar point is made by Terence Horgan (2006) and 2007).

For the emergentist consciousness is an astonishing phenomenon. There is a puzzle about how nature is capable of 'producing' this 'new' phenomenon on the basis of something quite different: the arrangement of molecules in a particular way and their causal interaction. Many philosophers accept that there is *prima facie* an explanatory gap. We do not seem to be able to *understand* why a certain complex arrangement of molecules leads to the occurrence of consciousness (*general explanatory gap thesis*) and why a certain complex arrangement of molecules leads to the occurrence of a particular kind of experience (*specific explanatory gap thesis*).³³ Most philosophers who accept these 'gaps' argue, however, that the puzzlement dissolves once we understand what it is about our cognitive make-up that makes it difficult or even impossible to understand why consciousness occurs (given a certain physical basis) and why specific conscious states (or events) are correlated with specific physical states (or events). A number of proposals have been made to explain the existence the so-called explanatory gap thereby providing an illusion theory: they explain why consciousness *appears* mysterious to us given our cognitive architecture although there really *is* no mystery about consciousness.³⁴ According to these philosophers, from an objective point of view, there is nothing to be puzzled about.

The emergentist rejects the idea that our natural puzzlement about consciousness is illusionary. It is an essential part of the emergentist position to insist on the *adequacy* of our amazement when we reflect about the phenomenon of consciousness. The emergentist understood in this way not only subscribes to the explanatory gap claims just mentioned. These claims merely *describe* our cognitive situation. The emergentist adds a normative claim: consciousness *deserves* astonishment. According to that view, consciousness is *objectively* an astonishing phenomenon and it is therefore a mistake to think that our puzzlement is the result of some kind of illusion. To the contrary, our amazement about the occurrence of consciousness is a symptom of our grasp of the phenomenon. A person who understands what consciousness consists in will see upon reflection how amazing it is that consciousness arises on the basis of some arrangement of matter. The emergentist so understood insists that it is perfectly *appropriate* to be puzzled about the occurrence of consciousness and that a person who does not see any puzzle here thereby shows a lack of understanding of what it is for an individual to be conscious.

References

³³ Levine argues forcefully for the explanatory gap thesis (Compare Levine 1993 and Levine 2001).

³⁴ As an example see Papineau (2002), chapter 5.

- Baker, Lynne Rudder (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press.
- Beckermann, Ansgar (1992), "Supervenience, Emergence, and Reduction", in A. Beckermann, H. Flohr & J. Kim, *Emergence or Reduction? - Essays on the Prospects of Nonreductive Physicalism*: 94-118.
- Chisholm, Roderick .M., (1976), *Person and Object*, Dordrecht.
- Horgan, Terence (2006), "Causal Compatibilism about Agentive Phenomenology", in Marcelo Sabates, David Sosa, Terence Horgan, *Supervenience an Mind. Festschrift for Jaegwin Kim*, MIT Press, in preparation.
- Horgan, Terence (2007), "Is Agentive Experience Veridical?", in preparation for a special issue of *Erkenntnis* collecting the papers presented at the conference "Mental Causation, Externalism, and Self-Knowledge", Tübingen (October 2005), organized by Michael Esfeld, Albert Newen and Vera Hofmann.
- Kim, Jaegwon (1993), *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge University Press.
- Kim, Jaegwon (1999), "Making Sense of Emergence", *Philosophical Studies* 95: 3-36.
- Levine, Joe Levine, J. (1993), "On Leaving Out What It's Like", in *Consciousness: Psychological and Philosophical Essays*, G. Humphreys and M. Davies, Hg., Oxford: Basil Blackwell, 1993, 121 - 136. Wiederabgedruckt in *The Nature of Consciousness: Philosophical Debates*, N.J. Block, O. Flanagan, and G. Güzeledere, eds. Cambridge, MA: MIT Press, 1997.
- Levine, Joe (2001), *Purple Haze: The Puzzle of Consciousness*, Oxford University Press.
- Lowe, Jonathan E. (1996), *Subjects of Experience*, Cambridge: Cambridge University Press.
- MacLaughlin, Brian (1995), "Varieties of Supervenience," in E. Savellos, and U. Yalcin, eds., 16-59.
- Nagel, Thomas (1974), "What Is It Like to Be a Bat?", *Philosophical Review* 83: 435-450.
- Nida-Rümelin, Martine (2006a), *Der Blick von Innen. Zur transtemporalen Identität bewusstseinsfähiger Wesen*. Suhrkamp Verlag (Frankfurt am Main). English translation in preparation (title: "The View from Inside. Transtemporal Identity of Conscious Individuals").
- Nida-Rümelin, Martine (2006b), "Doings and Subject Causation", contribution in preparation for Michael Esfeld, Albert Newen & Vera Hofmann (eds.), Special Issue of *Erkenntnis*
- O'Connor, Timothy (1994), "Emergent Properties", *American Philosophical Quarterly* 31: 91-104.
- O'Connor, Timothy (2000), *Persons and Causes: The Metaphysics of Free Will*. Oxford Univeristy Press.
- O'Connor, Timothy (2002), "Libertarian Views: Dualist and Agent-Causal Theories", chapter 15 of Robert Kane (ed.), *The Oxford Handbook of Free Will*, Oxford University Press 2002.

Papineau, David. (2002), *Thinking about Consciousness*, Oxford University Press.

Savellos, E. and Yalcin, U., eds., (1995), *Supervenience: New Essays*. Cambridge: Cambridge University Press.

Stephan, Achim (1997), "Armchair Arguments Against Emergentism", *Erkenntnis* 46: 305-314.

Stephan, Achim (2002), "Emergentism, Irreducibility, and Downward Causation", *Grazer Philosophische Studien* 65: 55-93.