CONSCIOUSNESS AS AN EMERGENT PROPERTY:
Why Consciousness Is Not A Property of the Brain, but that of a Human Being.

Joseph Wang

UDK 17.024.1

[Special editor's remark: due to technical necessities, we were unable to use the standard superscript marks for note. Since this text contains many numeric marks, we resolved to use the following format for note: (nNumber). So, for example, the first note will be marked (n1).]

0. *Introduction*

One of the possible solutions to the mind–brain–problem in the philosophy is the so–called *emergentism*. It is characterized as a middle way between two extreme positions: "substance dualism" and "eliminative materialism". (n1) There is a general agreement among all emergentists, that there are entities which are solely *based* on lower physical materials, but they have properties which go beyond that of their physical parts. This is the position I want to defend here. But since there are many possible metaphysical theories which fit to this first characterization, an overview over different theses in emergentisms should be presented at first [1]. They should provide a better understanding of 'emergentism' which I want to defend [2]. I will then try to discuss the notion of 'consciousness' as a property and the neuro–anatomical basis for it. Here I shall provide some basic biological facts about human organism and introduce the distinction between 'body schema' and 'body image' [3]. On the account of the information provided in [2] and [3] I conclude that 'consciousness' is an emergent property of an emergent substance, the human being, but not that of a human brain [4].

1. *What is Emergentism?*

Confusions arise when one reads about emergence theories. Many philosophers call their position 'emergentism' while their positions even contradict with each other. In my judgment not every possible position within the sub-

stance dualism and the eliminative materialism should be called 'emergentism'. To bring in an order into this clutter we must first meet some distinctions.

According to A. Stephan the emergence theory was first developed by Samuel Alexander, Conwy Lloyd Morgan and Roy Wood Sellars, the so–called 'early twentieth–century British emergentists'(n2). It is brought about to answer two major questions in the philosophy. The first one is of cosmological importance: How is 'change' in the world possible? Is there a pattern to which the change occurs? How does the world 'evolve'? As far as an emergence theory is supposed to answer this question, I will call it 'evolutionary emergentism'; it is mainly a theory about how novelty comes into the world. I will not address the evolutionary emergentisms in this paper.

The other main question emergentists try to explain is that about metaphysics. If an emergence theory deals with this question, I will call it 'metaphysical emergentism'. This kind of emergentisms tells us about the nature of things in the world. Specifically it is supposed to answer to these questions: "Are there different kind between material beings?" "Can some of the properties instantiated in the world be reduced to other properties or not?"(n3) P. Clayton distinguishes five different meanings of 'emergence' according to the different coverage where this word is used. (n4) Clayton characterizes the "metaphysical emergence" in this way:

Emergence in this sense is a metaphysical theory, in the sense that physicalism and dualism are also metaphysical theories. It claims that the nature of the world is such that it produces, and perhaps must produce, continually more complex realities in a process of ongoing creativity, and it is a thesis about the nature of what is produced. [...] Metaphysical theories are not limited inferences from the available evidence; they are hypotheses about the nature of reality as a whole. (n5)

After we have sorted out the metaphysical emergentisms from the evolutionary ones we can meet another distinction made by M. Silberstein and J. McGeever. We can separate the 'ontological emergence' from the 'epistemological emergence'. (n6) While the epistemological emergence deals with our epistemic performances, like the relation between different nature laws or the problem with the predictability of a system, the ontological emergence deals with the relation between a systemic whole and its parts.

It can be objected here that like the British emergentists most contemporary metaphysical emergentists (e. g. Clayton (2004)) address both ontological and epistemological aspects of a metaphysical theory. Therefore the distinction between ontological and epistemological emergence theories are merely academical. I think though even in cases where an emergentist dedicates herself to both aspects of metaphysical emergentism it is still possible to analyze the ontological part of her emergence theory separately.

After narrowing down the possible emergentisms discussed here to the ontological emergentisms there are still many different emergence theories we need to discriminate. I think the best way to do this is to follow Stephan and analyze the main theses maintained by the emergentists. (n7) Though most emergentists share the main theses they have individual understandings about them; and sometimes the different meanings even contradict with each other. Before I come to each thesis I want to clarify my usage of the notion 'strong' and 'weak'. Since nearly every author I read uses the term 'weak emergentism' to refer to a theory that is closer to the reductivism, I will call an emergence theory 'weak', if it has more similarity to the reductive physicalism. On the contrary the 'strong emergence' refers to a position closer to the property dualism. (n8) Since Stephan does not distinguish between the ontological from the epistemological emergentisms I will sort those main theses out concerning the epistemological issues. (n9)

1.1 Main Theses of Emergence Theories

There are six main theses I want to discuss here: naturalism, systemic properties, hierarchical existence, (synchronic) determinateness, irreducibility and the so–called downward–causation.

Naturalism

Stephan claims that naturalism is presupposed by emergentisms. The problem is what is *meant* by 'naturalism'? W. Löffler distinguishes between three different kinds of 'naturalisms': the methodological, the semantic and the ontological naturalism. (n10) Though most naturalistic philosophers are committed to all three kinds of naturalism, for our purpose the ontological naturalism is the most important one. According to Löffler, the ontological naturalism is a kind of monism: Everything which exists has the property A; and so there is not a thing with the property B. (n11) The ontological naturalists deny the existence of such thing as 'mental entities' or 'Cartesian souls'.

What is the link between naturalism and emergentism? While epistemological emergentists can commit to all three kinds of naturalism, an ontological emergentist must specify how to understand the ontological naturalism. In my opinion all emergentists share this idea:

(1) Every entity which exists is based on the basic particles.

The sentence (1) is the monism–claim, as I want to call it. With (1) the demand for monism should be satisfied. Since (1) is in no way special, as reductive physicalisms, supervenience theories, some kinds of property–dualisms, and some kinds of neutral monism are also committed to it, the only

possible metaphysical position which can be sorted out by (1) is the substance dualism. (n12)

Systemic Property

Emergentists support the thesis that when the basic particles can form a systemic whole, there are properties attributed to the system, but not to the parts of it. This thesis can be summed up as:

> *(2) There are systems with systemic properties: These properties are had by the systems, but not by their (proper) parts.*

The sentence (2) can be accepted within many different philosophical schools. We can find systemic properties easily: The weight of a tree e. g. is not had by any of its parts; though water (H_2O) is liquid, its parts (two H–Atoms and O–Atom) are not; a living cat lives, but none of her parts is actually alive. (n13) This can be called the system–claim. In my opinion the system–claim is also unproblematic, since even reductive physicalists can agree with it.

Hierarchical Existence

The emergentists are committed to the thesis of hierarchical existence, i. e. the claim that there are entities of different "levels". Basic particles, for example, belong to the most basal level, while systems formed by the basic particles belong to an upper level. A complex system belongs to an upper level, while its parts belong to lower levels. This can be summarized by (3):

> *(3) Each natural thing belongs to a certain level of existence. When regarding a system s, consisted of different parts, we can say that, when the part of s with the highest level belongs to the level n, then s belongs to the level n+1.*

There is a different assertion to this issue: According to Stephan it is not the part, but the property a thing has, which determines the existence level of the thing. (n14)

> *(3a) Each natural thing belongs to a certain level of existence according to the properties it has.*

In my opinion while (3) is more ontological guided, (3a) takes the account of epistemology more seriously. But more importantly (3) and (3a) do not contradict with each other. For the purpose followed in this paper the differences between (3) and (3a) can be neglected.

One can object here, that both (3) and (3a) cannot provide an exhaustive list of levels of existence, and therefore we can not identify the things according to their level of existence. I must confess that though some (n15) have

tried to enlist every level they can find, none of them has succeeded. But perhaps this list is not a thing emergentists need to provide actually. For the thesis of the hierarchical existence to be sound we do not need an exhaustive list of a level, but a robust notion of 'level of existence' which can cope with different kinds of systems and properties.

Synchronic Determinateness

We have to distinguish 'determinism' from 'determinateness'. 'Determinism', as a metaphysical theory, is the philosophical claim, that all events happening in a later time are predetermined by the events of a *prior time*. So determinism is a theory of how things evolve *in time*. 'Determinateness', on the other hand, claims something else. In context of ontological emergentism only the synchronic determinateness is of importance: (n16)

(4) *The properties and the behavior of an upper level entity (system) depend on the properties and the behavior of its parts and on the relation between them.*

Clearly, for (4) to be sound the notion of 'dependence' must be made clear here. There are differences in the interpretation of the determinateness–claim. Those philosophers supporting a weaker claim of (4) content that the determination is *nomological*, i. e. the notion of 'dependence' is bound to *nature laws*:

(4a) *The systemic properties and the systemic behavior are fully determined by the properties and the behavior of its parts, according to the nature laws.*

From (4a) we can follow the epistemological thesis that if we know all properties had by the parts and the natural laws, we can predict the properties and the behavior of the system. Clearly, if the 'nature laws' in question are the laws in the science of physics (n17), a philosopher arguing for (4a) would not be an emergentist, but her thesis would be a reductive theory. For ontological emergentisms (4) should be read in another way. A stronger claim of (4) that finds its supporters among the moderate emergentists is (4b):

(4b) *If a system s consisting of parts* $c_1, c_2 \ldots c_n$, *while* $c_1, c_2 \ldots c_n$ *are ordered in the way o, has the (systemic) property* p, *then every system* s_i *with* $< c_1, c_2 \ldots c_n \mid o >$ *has the property* p.

(4b) abandons the nomological claim; there is no nature law which determines the behavior and the properties of an emergent system. But (4b) insists on the thesis that the parts of a system fully determine the behavior of a system. It must be noted that though (4b) is compatible with the diachronic de-

terminism, it does not *entail* the thesis of diachronic determinism. Because though the systemic properties are fully synchronically *determined* by the parts of a system, it does not follow from this that the determinism is true. As to my knowledge most emergentists agree with (4b). There is an even stronger claim of ontological emergence, though:

> *(4c) There is no regularity between the properties of a system and those of its parts. Even two systems with the same parts and the same organization of the parts may have different systemic properties.*

I think (4c) is the most 'dualistic' interpretation an emergentist can have without abandoning (1). Indeed (4c) is compatible with both property— and substance–dualism. But not every emergentist agrees with (4). In their 'Dynamical Emergence Model' T. O'Connor and H. Wong (n18) suggest that the relation between a system *s* and its parts must be a *causal* and *diachronic* one:

> *(4d) The parts* ($c_1, c_2 \ldots c_n$) *ordered in a way o at a time t0 cause an emergent system s at a later time* t_1 *to exist. Therefore s is in a sense determined by its parts.*

Whether (4d) is compatible with (4a), (4b) and (4c) depends on the interpretation of the notion 'causal relation'. Since O'Connor and Wong themselves also reject the causal closure of the physical world, their emergentism is rather strong. But with (4d) alone a weak emergentism is also thinkable. (n19)

Irreducibility

This is probably the most important thesis within the emergence theory. Interpreted ontologically the irreducibility axiom separates theories of supervenience from those of emergence. (n20) Generally the ontological irreducibility can be stated like this:

> *(5) The systemic property* p *of a system s, consisting of* $< c_1, c_2 \ldots c_n \mid o >$, *is irreducible and therefore emergent, if s's having* p *cannot be deduced from the properties and the behavior had by* $c_1, c_2 \ldots c_n$.

Before I can introduce the different readings of (5), we must first ask what the notion of 'irreducibility' means. According to Stephan only (nature) laws can be deduced (and therefore reduced), but not properties. (n21) Thus, "*s*'s having *p*" must be thought as a (nature) law: If *p* is truly an emergent property, then we cannot deduce that *s* ($= < c_1, c_2 \ldots c_n \mid o >$) has *p* from the properties and behavior of $c_1, c_2 \ldots c_n$. (n22)

We can ask *why* (5) is true. By answering this question we can distinguish different kinds of emergentisms. The weakest reading is the solely epistemological interpretation.

(5a) "The system s has an emergent property p*" cannot be deduced from the known laws of nature which apply to* $c_1, c_2 \ldots c_n$.

If (5a) is true, then it will be possible that eventually, by getting more knowledge about the nature, we will finally be able to deduce the emergent property from the property of the parts. In this sense 'emergence' is a term relative to the knowledge we have. But an ontological emergentism, which is the main concern in this paper, requires a stronger reading of (5) than (5a). An ontologically more robust reading of (5) is (5b):

(5b) We cannot deduce the emergent systemic property from the properties of the parts, because the nature law applies to s cannot be deduced from the nature laws applied to $c_1, c_2 \ldots c_n$, *when* $c_1, c_2 \ldots c_n$ *are jointed in the o way.*

Though (5b) makes a stronger claim, the irreducibility is still an epistemological one, since (5b) still argues with the notion of 'nature laws'. (n23) In my opinion, what is actually needed in an ontological emergence theory is an *ontological* irreducibility, as M. Silberstein and J. McGeever(n24) claimed. The ontological irreducibility could be stated like (5c):

(5c) The emergent systemic property p of a system s, consisting of $< c_1, c_2 \ldots c_n \mid o >$, *is genuine irreducible and therefore emergent, because s is a complete new entity with new properties. The system s follows new nature laws and can even override the properties had by the parts of s.*

It is important that (5c) makes an *ontological* claim, while (5b) does not do the same. It seems clear that some systemic properties are reducible to the properties of the parts. The weight of a book, for example, is nothing but the *sum* of the weight of its proper parts, *viz.* the pages and other smaller parts. So, how much a book weighs is reducible to the weights of the pages (and other parts) of the book. This kind of systemic property will not be emergent in the sense of (5c). Whether there are some properties in the world which satisfy (5c) or not, might be an empirical question. Silberstein and McGeever regard some phenomenon in the Quantum Mechanics, *viz.* the EPR–Paradox, as an example for a genuine emergent property. (n25)

Downward Causation

Let us turn to the last and probably most irritating thesis emergentism holds: the downward causation. Let us first formulate a thesis what 'downward causation' means. According to Stephan, the thesis of 'downward causation' is this(n26):

(6) An emergent system is able for 'downward causation', if it can determine the behavior of its parts in a way, which cannot be reduced to behavior of a system with less complexity.

This 'feature' of an emergence theory has been questioned ever since it has been formulated by the British emergentists. I will discuss arguments in favor and against downward causation in the next chapter. What I want to do here is to discuss the different readings of (6). Again there will be weaker and stronger interpretations. First the weakest interpretation of (6), the epistemological reading, could be like this:

(6a) The notion of 'downward causation' is relative to our knowledge. Though a system seems to cause its part to behave in a unexplainable way, this behavior can be explained in a later time.

Certainly, (6a) will be too weak for an ontological emergentism. But (6a) is motivated by the advances of the science. While in earlier times chemistry is thought as a special science(n27) about atoms, molecules and their special behaviors, today most chemists admits that the fundamental chemical laws can be deduced from the quantum mechanics. (n28) But (6a) is not the only reading of (6), we can make a stronger claim about 'downward causation':

(6b) The behavior of the parts of an emergent system s depends not only the nature laws applying to the parts, but also on emergent properties of s. The emergent system can therefore be said to be able to cause downward, because s's emergent properties help to determine the behavior of its parts.

(6b) claims an ontological downward causation, but does not deny the causal role of its parts. There is still a stronger interpretation of (6) available:

(6c) As $c_1, c_2 \ldots c_n$ *join to the emergent system s, nature laws prior applicable for the parts lose their power for the parts, and s and* $c_1, c_2 \ldots c_n$ *now comply to a new, irreducible law. The system s is therefore able for the downward causation, because* $c_1, c_2 \ldots c_n$ *behave now in a complete different way.*

This would be the strongest ontological claim for emergentism(n29). Since the downward causation is a special issue, I want to discuss it more thoroughly.

1.2 The Downward Causation

There are many philosophical arguments against the existence of an event *e* which involves a 'downward causation'. I want to present two major arguments here:

1. The Pepper–Kim–Dilemma

I dub the notion of 'Pepper–Kim–Dilemma' from Stephan(n30). This dilemma is concerned about the notion of 'downward causation' and the plausibility of ontological emergentism. In short, two theses (a) and (b) cannot be stated consistently.

(a) The causal closure of the physical world.

(b) The ontological reading of the downward causation: (6b) or (6c).

We have to add two other premises to the list, if we want to show the inconsistency, which according to Kim emergentists are committed to.

(c) Alexander's Dictum: A property is only real, if it has causal powers. (n31)

(d) There are no overdeterminating causes in the world.

What is the dilemma after all? An emergentist must on the account of the existence of downward causation, either admit that her emergentism is a kind of epiphenomenalism, or she must deny the causal closure of the physical world. This can be shown in this way:

According to (c), if an emergent property p of an emergent system s is real it *must* have emergent causal powers. But if (a) is true, then p will be an over-determining cause for the behavior of the parts of s. This would contradict with (d). If the behavior of the parts of s is fully determined by the causal closure of the world, we do not need another causal power coming from s. So, p doesn't have any causal power, and therefore p is epiphenomenal, i. e. not real. S. Pepper follows from this that emergentism cannot be true, or (b) must be rejected. And if (b) is rejected, what is left of emergentism would just be another version of supervenience.

This epiphenomenalism can only be evaded if (d) or (a) can be rejected. Since (d) seems to be true, an emergentist will have to reject the causal closure of the world. This is, in short, the Pepper–Kim–Dilemma.

2. *Downward Causation is Unintelligible*

The most prominent modern adversary of ontological downward causation is probably J. Kim. He has presented another argument against the downward causation, and by doing so he rejects the idea of emergence. (n32)

For the sake of the argument, let us assume that there are really events including downward causation. This can mean either as an "ontological synchronic downward causation", or as the "ontological diachronic downward causation". In the synchronic reading the system s is said to have causal power

to determine the behavior of its parts at the same time, as s is determined (constituted) by its parts. This reading contradicts with the following thesis:

(e) The causal relation is asymmetric.

If a causes b at a time t, then by no means will b cause a at t. Either the parts (c_1, c_2... c_n) are caused by s, or s is determined by its parts, but not both. The synchronic downward causation is therefore unintelligible.

What about the second thesis: "ontological diachronic downward causation". In this reading an emergent system s will determine the behavior of its parts at another time. If c_1, c_2... c_n somehow cause the emergent system s to exist at a time t_0, then at some other time t_1 after $t0$ s causes c_1, c_2... c_n to behave differently than they would do due to their nature. In this reading the concordance to (e) is preserved.

According to Kim, there are still difficulties for the diachronic downward causation. Here the most severe one should be presented: The serious problem for the diachronic downward causation is the causal overdetermination. According to Kim, if ontological diachronic downward causation is taken for granted, we can simplify the situation of the downward causation to this: At time t_0 the base P determines(n33) an emergent system M; at a later time t_1 the base P^* determines M^*. Now, if s has any causal power, M should cause P^* or M^*. But M^* is already determined by P^*, so we do not need the causal power of M for M^*. But if P^* is caused by M, what should we do with the causal power of P? Kim concludes:

This appears to make the emergent property M otiose and dispensable as a cause of P^*; it seems that we can explain the occurrence of P^* simply in terms of P, without invoking M at all. If M is to be retained as a cause of P^*, or of M^*, a positive argument has to be provided, and we have yet to see one. (n34)

Defending Downward Causation

The arguments of Kim seem to be sound. Since I want to defend the ontological downward causation, I want to provide an argument for it. In my opinion the best way to do this is to reject the thesis (a), the causal closure of physical the world. (n35) Here I want to present two approaches explaining the plausibility of 'downward causation'. The first one comes from O'Connor and Wong (n36), they suggest a dynamical model view. The other one is an analysis of the notion 'causal power' embedded in the Aristotelian hylemorphism (n37).

As stated before in (4d), O'Connor and Wong regard the relation between the emergent system (E) and its parts (P) as a diachronic causal one. In their opinion E is not supervenient from, but caused to exist by an earlier P. O'Connor and Wong show us their idea by using the following notation.

$P@_n$: the overall summation of the base of E at time n.

P^*: the complex physical configuration, which is able to persist through time and cause E to exist.

P_n: the remaining physical aspect of E ($P@_n$ — P^*) at the time n.

E_n: An emergent state at the time n.

$\Rightarrow$: (Minimally sufficient) causation.

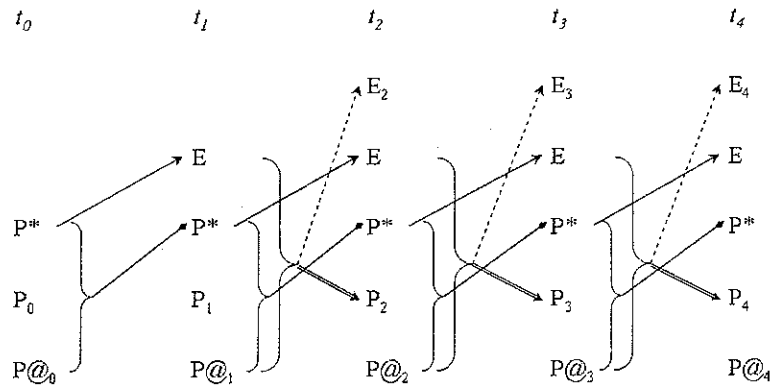
According to O'Connor/Wong the following causations should be regarded as premises for emergent downward causation:

$$P^* \text{ at } t_0 \Rightarrow E \text{ at } t_1.$$

$$P^* \mid P_0 \mid P@_0 \text{ at } t_0 \Rightarrow P^* \mid P_1 \text{ at } t_1.$$

Generalizing the idea in a time indexed diagram, we can show that the downward causation can be thought like this:

Figure 1



Dynamical evolution of system *S* over time

Upward causation of baseline emergent state E

Upward causation of super-emergent states E_n

Maintenance causation of emergence-sustaining configuration P*

Wide horizontal causation (including downward causation)

The dynamical model view of the emergence is able to avoid the difficulties mentioned above. Since O'Connor and Wong reject the causal closure of the world they won't have to worry about the overdeterminating cause. So their emergentism is neither epiphenomenalism, nor causal overdeterminating. Since this view is diachronic, the causation in this model is asymmetric, and therefore logical consistent. This is not the only solution to Kim's arguments.

2. *A Theory of Ontological Emergentism*

Now I come to the ontological emergentism I want to defend. (n39) The core idea is to regard the theory of emergence as one of the way an Aristotelian substance can be realized. So I want to present a certain kind of ontological(n40) emergentism, which does not contradict, but is even compatible with the Aristotelian hylemorphism.

According to the Aristotelian Metaphysics and its reception by the analytic non–physicalistic ontologists, J. Hoffman and G. S. Rosenkrantz (n41), only entities which are 'substances' can be said to be existent. Now, what is an Aristotelian substance in an emergentist's view?

2.1 Hylemorphism and the Doctrine of Potentiality and Actuality

Let us first recall what hylemorphism is. According to Aristotle (n42) hylemorphism is an ontological theory on how to interpret the world properly. Every substance existing in the world is determined by two concepts: The *hyle*, or *materia*, is the "stuff" of which a substance is made of; and the *morphe*, or *forma*, is the way the substance is "composed". We also have to regard another metaphysical theory, the doctrine of potentiality and actuality, developed by Aristotle. This thesis is a metaphysical framework within which we will be able to interpret 'changes'. According to Aristotle we can distinguish between the 'substantial changes', where a substance begins (or ceases) to exist, and the 'accidental changes', where a substance gains (or looses) an accidental property.

Because the topic of this paper is the human consciousness, let us review what Aristotle's hylemorphism has to say about the human kind of substance. In Aristotle's "On the Soul" (n43) and in the interpretation of Thomas Aquinas(n44) the human soul is the *forma substantialis individualis* of a human substance. What a human substance is, depends not only on the *materia* it is consisted of, but also on the human *forma* it has, and this is the (Aristotelian) soul.

It is important to catch the intuition that the *forma* of a substance determines which property a substance can gain or loose. E. g. while a parrot can produce sounds similar to the human language an orangutan cannot produce these sounds due to the anatomy of its pharynx. In terms of Aristotelian metaphysics it can be contented that the form of the substance orangutan does not contain the *potentiality* of speaking human tones, while the form of the substance parrot does.

2.2 Causal Power and Substance

Why bother with the Aristotelian notion of 'substance', 'potentiality' and 'actuality'? I think the ontological analysis provided by Aristotle can make the notion of 'causation' more robust. Let us analyze what is meant with the notion of causal power when we have adapted the Aristotelian view. First, causal power cannot be an Aristotelian substance, since causal powers cannot exist by themselves. (n45) Therefore they must be some kind of (accidental) quality. (n46) If this is accepted two more questions arise. First, what is the *bearer* of causal power? And second, how can we interpret the property of causal power properly?

According to the Aristotelian tradition, certainly substances can be seen as 'having causal power'. (n47) If properties are to be allowed for being able to cause 'causation' must be some kind of a second order property, which I will reject here. If this analysis is correct there is no need for 'downward causation' for the *emergent properties*. What we actually need for a strong ontological emergentism is *emergent substance*.

And this is precisely why I reject (c). Since according to my opinion no property has causal power, (c) should be rejected. The Alexander's dictum can be 'transformed' to the substance ontology, though.

(c') A substance is only real, if it has causal powers.

But what does it mean for a substance to have causal powers? If a substance *x* (e. g. a cat) has a certain causal power (e. g. causing the death of a mouse by catching it) then we can analyze the situation as follows:

(cp) A substance x has the power to cause y, if x causes y when the circumstances c's are fitting.

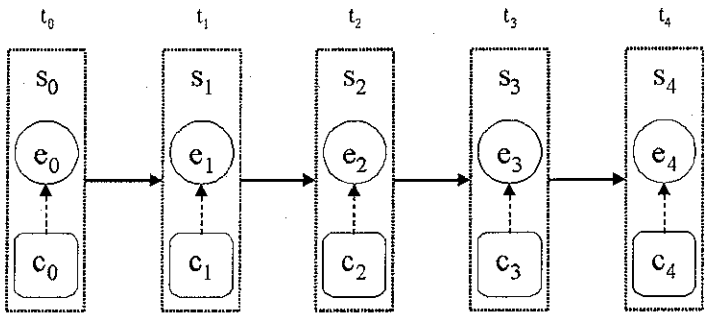
What (cp) contents, is not a definition of 'causal power', but rather the explication of 'causal power' as a *disposition*. If this can be accepted we will have another strategy to propagate the downward causation.

Let us review the thesis of 'downward causation' (6). What I want to defend here is the thesis (6c). When the parts *c1, c2... cn* join with each other and form a substance *s*, they behave in another way when they are not embedded in the system. Two questions arise with (6c): *How* can an emergent substance determine the behavior of its parts, and *why* does (6c) not violate with the causal overdetermination clause (d).

As much for the first question, I think the dynamical model view of emergence provided by O'Connor and Wong is convincing, though I should deny that the relation between the emergent system *s* and its parts is causal. In my view the base parts synchronic determinate the emergent system *s* at the time t_0; and the emergent system *s* as a *substance*, i. e. the whole system, causes

the emergent system s^* at a later time t_1 to gain or loose some properties. The figure 2 tries to make the relations between parts and emergent properties explicit:

Figure 2



s_n: Emergent substance s at time n.
e_n: Emergent property e at time n.
c_n: Emergent base c (={c_1, c_2 ... c_n}) at time n.

——→ Diachronic causation
- - - - - - → Synchronic determination

In comparison with Fig. 1, Fig. 2 seems much easier. As for the 'maintenance', i. e. the diachronic identity, of *s*, I shall follow E. J. Lowe and conclude that the diachronic identity of a three–dimensional substance is basic and not analyzable. (n48) If this kind of causation is sound we can conclude that parts belonging to an emergent system do not have any (diachronic) causal power at all, what they possess is the ability for synchronic determination. So the second question is in the Aristotelian view but a spurious question: Since the synchronic determinateness is not a causal relation, there can be no causal overdetermination for the emergent property *e*. And since the only (sufficient) cause for the existence of the substance *s* at t_1 is s_0, we do not have overdeterminating causes, neither for c_1 which is caused by s_0, nor for e_1 which is though determined by c_1, but caused by s_0.

2.3 Emergent Substances and Emergent Properties

If my analysis about 'causation' provided above can be accepted, I want to provide other theses concerning the irreducibility of emergent substances. Let us first try to define the notion of 'emergent substance': (n49)

(T1) Given a system s, consisted of material parts c_1, c_2... c_n *ordered in the way* o, *if s has at least one emergent property* p *which is irreducible in*

the sense of (5c) and s is not a (proper) part of an emergent substance, then s is itself an emergent substance.

What (T1) suggests is that if a system *s* is an emergent substance, it must have at least one emergent property *p*, and that the property *p* is emergent iff *p* is irreducible in the sense of (5c). With (T1) we can exclude systems which are only pseudo–substances. It seems clear, that not every system has irreducible properties, while 'irreducible' should be read in the sense provided by (5c). Let us regard the following examples:

A soccer–player *f* is consisted of, say, a head, a torso, and four extremities. But while *f* is indeed alive, none of her parts is really alive. Let us assume that 'being alive' is an irreducible property, so *f* is a system with irreducible properties and therefore a candidate for an emergent substance.

The soccer team *t*, on the contrary, is consisted of many parts ($f_1, f_2 ... f_{11}$). But let us assume for the sake of the argument that there isn't any property *t* possesses which cannot be reduced to the properties of $f_1, f_2 ... f_{11}$. So *t* is not a candidate for an emergent substance, but its parts, $f_1, f_2 ... f_{11}$ are.

What about the leg (*l*) of *f*? Let us assume that *l* consists of organic cells ($o_1, o_2 ... o_n$) and that *l* does have an irreducible property(n50). Because *l* is embedded in a system *f* which itself has irreducible properties we cannot regard *l* as an emergent substance, at least not as long as *l* is a part of *f*. So after this analysis we can surely regard *f* as an emergent substance, while neither *l*, nor *t* is actually an emergent substance.

Let us review what an emergent property is:

(T2) A systemic property p *is an emergent property iff* p *is irreducible in the sense of (5c).*

Because forming a system is somehow arbitrary(n51), (T2) does allow (proper) parts of a substance to have emergent properties. This should not pose any difficulties, since (T2) fits well with (3), the thesis of hierarchical existence. When defending (3), (T2) makes sure that in every level of existence there can be (at least in principle) a substance in that level of complexity. Though there might not actually be any emergent substance in the level of, say, the complexity of molecules, (T2) will not abandon the possibility.

2.4 Some consequences of the notion of 'emergent substances'

If this kind of ontological emergentism can be followed some questions and problems will arise. Some objection and counter–arguments should be presented here.

Is this kind of emergentism 'monistic'?

In order to understand this objection we have to review the thesis (1). As stated before (1) means that every existent being is based on the same particles. A philosopher committing to the Aristotelian hylemorphism contents that a substance is based on both *materia* and *forma*. This seems to contradict with (1), at least, if we read (1) as the following sentence:

(1a) Every existent being is based only on the same particles.

If (1a) is accepted then there will be no place for the *forma*. If a philosopher insists on the existence of the *forma*, she will be no monist, at least not in the sense of (1a).

This objection does sound. But perhaps we can moderate the ontological commitment of the Aristotelian emergentism. The *forma* does not really exist, just like the *materia* does not exist. At least they can not exist in the same sense in which a *substance* exists. There is a difference between the *forma* and the *materia* of a substance: their divisibility. When an emergent substance *s* lost a part *c*, (n52) *c* becomes itself a substance as long as *c* does fulfill (T1). So the basic particles which are not bound in an emergent substance can be regarded as a substance themselves. (n53) This cannot be said of the *forma*(n54). I think what ever the *forma* is, is best explained as the sum of the *essential* properties of a substance. Though these properties are essential to the emergent substance and therefore necessary conditions for the existence of the substance, they are not necessary conditions for the existence of the parts. Therefore, we do not need to proclaim the existence of the *forma*. We can still be weak ontological naturalist without giving up the existence of substances.

Where does the forma come from?

Another objection to this kind of emergentism aims at the existence of the *forma*. The ontological emergentism cannot make clear *how* the *forma* can really form the *materia*, if it does not exist. To their understanding the *forma* of an emergent substance must be somehow *preexistent* to the substance, so it can really help to *actualize* the substance.

I think that actually this question is aimed by the original British Emergentists. They try to answer the question *how* the evolution can bring out new entities. As far as the emergentism is concerned in this paper, the evolutionary emergentism cannot be dealt properly here. I hope it is sufficient here to answer this question with the evolutionary theory and those mechanisms (e. g. the mutation and the selection) mentioned there. So my answer to this objection will be that the evolution (or more proper: the mechanism in the evolution) is the source of the *formae*.

An emergent system is ontologically vague.

Similar to the argument of the "Ship of Theseus"(n55) someone might suggest the following argument: If the ontological emergence proposed here is true, we will have difficulties with the identity of an emergent substance. It seems clear that (at least) some emergent substances constantly gain and loose some parts. Suppose that an emergent substance s, consisting at t_0 of c_1, c_2, and c_3, looses c_2 and gains c_4 at t_1. If it is possible for an emergent system to be constituted by different parts, how can we identify both proper parts of s and the system itself? (n56)

This objection is serious. It seems that an ontological emergentist cannot cope with this objection without adaptation of the ontological emergence theory. We need either to pair the ontological emergentism with a strong evolutionary emergentism, by doing so the ontological and evolutionary emergentism can guarantee the ontological independence of the *forma*. Or we must adopt a dualistic conception of substances which abandons (1). When the monism claim can be dropped, a dualistic conception of substances which gives both *materia* and *forma* ontological independence can be suggested; and the problem of the vagueness of a substance will be evaded.

I am not sure if these are really the only options available. I would rather try to turn the tables. Indeed, we do not know for sure whether a specific part does belong to s or not, but this is just another emergent property of s which cannot be reduced to the properties of c_1, c_2... c_n. The identity criteria for an emergent system of a higher level are not the same to the identity criteria for its parts which are of a lower level of existence. So I would answer to the objection so: It is perfectly fine that we cannot deduce the identity criteria of a higher level system to the identity criteria of its lower level parts, and this is but another hint for the probability of the ontological emergentism. (n57)

There is no evidence for this.

One problem concerning ontological emergentism is that there is no evidence for this theory. Due to the advance of the nature sciences former candidates for emergent properties can now be reduced to the properties of some basic particles. One good candidate for the emergence was the molecules; they have properties which seemed to be irreducible from the properties of the atoms. But nowadays, as the science advances, we know that (at least some) basic properties of molecules are reducible to the properties of the atoms; and (at least some of the) properties of the atoms are reducible to the properties of elementary particles. It seems plausible that in future every systemic property that seems to be emergent can be reduced to the basic law of the physics.

One way to encounter this objection is to question evidences for alternatives of emergentism. Is there any evidence for, say, substance dualism or reductive physicalism? I will not discuss this any further.

Another way for ontological emergentism to survive this objection is to deny the following theses.

(5d) If the systemic property a of a system s, such that $s = < c_1, c_2 ... c_n \mid o >$ *can be explained with the properties of* $c_1, c_2 ... c_n$*, then a is ontologically reduced to the properties of* $c_1, c_2 ... c_n$.

But this will not take the emergentists very far, since by denying (5d) there can be no empirical objections to the ontological emergentism, and so a discussion about this subject will seem to be futile.

Yet the best way to counter this objection will be to provide a candidate for emergent property. According to most emergentists, the (human) consciousness would be a very promising candidate for an emergent property. If this kind of ontological emergentism is plausible, the consciousness will be a good candidate for a systemic property which cannot be reduced to the properties of the parts of a system. (n58) If it can be accepted that there is at least one property which is genuine emergent then we will have an evidence for the ontological emergence.

3. *Is 'Consciousness' an Emergent Property?*

As stated in [0], the goal of this paper is to defend the thesis that 'consciousness' is an emergent property. (n59) In order to make this thesis plausible, I want to proceed in four steps. First, I shall give an account on the notion of 'consciousness' in [3.1]. In this section I want to explicate what is meant by the notion of 'consciousness'. In [3.2] I provide some basic information about neuro–anatomy, as the nervous system is considered as the basis of the 'consciousness'. The so–called "second–order neural pattern"(n60) and the distinction between 'body schema' and 'body image' will play a central role in this section. In [3.3] I want to provide indications for the thesis that the property 'consciousness' depends on complex nervous systems. And in the last section [3.4] I want to provide an argument for the thesis that 'consciousness' is not reducible (and therefore irreducible) to the nervous system.

3.1 What Is 'Consciousness'?

I haven't said anything about 'consciousness' so far. Though I have presented what is meant by "emergentism" I want to provide at least some basic understandings of "consciousness", now, in order to defend the thesis that human consciousness is an emergent property. Because there are many aspects of 'consciousness' which should be captured in a general theory, and because many theories fail in capturing them all, most explications of this term are

slippery. In order not to get into this trouble, I think a rudimentary understanding of "consciousness" will suffice for the purpose of this paper. While there are many elaborated theories about (human) consciousness I only intend a hint here on what the consciousness is.

I cannot provide a full scope analysis of the notion 'consciousness'. But some basic aspects, which are–at least to my knowledge–widely accepted, will suffice here for our purposes. A first preliminary is that the consciousness in discussion here is the normal human consciousness. (n61) There are two aspects of 'consciousness' I want to discuss here; from these two I want to develop a heuristic definition of 'consciousness' that could help us in the search for the emergent property. The first one comes from Thomas Nagel. In his famous article Nagel explicates what is meant by 'consciousness':

But fundamentally an organism has conscious mental states if and only if there is something that it is to be that organism–something it is like *for* the organism. (n62)

I would like to call this thesis as the thesis of phenomenal consciousness(n63). The second aspect is the thesis that if anything is conscious, it must be conscious of 'something'. To dub Meixner's words:

Consciousness is a kind of a bipolar medium, of which the one pole (the *subject*) is referred to the other pole (the *object in a general sense*) by the relation of *"being–conscious–of"* [...]. (n64)

This should be called the "Bretano–Husserl–Thesis"(n65). If we combine these two aspects we get a heuristic definition of 'consciousness' (C). I regard the following thesis as granted:

> *(C) A conscious being x is conscious about y, if x is aware of y and x is aware of the awareness of x. (n66)*

Perhaps the best way to understand this explication of 'consciousness' is to contrast the conscious animals to the unconscious things. We normally do not assume that an ordinary chair is conscious, but human beings (and probably even bats) like ourselves are conscious. What a chair lack is both the self–awareness and the capacity to refer itself to other entities.

3.2 Neuronal Self–Representation

There is a wide agreement that the Central Neural System (CNS) plays a center role in the mind–brain–debate. I cannot go into details here, but I will try to give a good account of what is taught in the neuro–anatomy lessons at a medical school. Basically we can separate two different parts of the neuro–anatomy: First, there is a Peripheral Neural System (PNS) which consists of nerves go all over the body. The PNS is responsible for transferring information from the perceptive structures (e. g. the retina in the eye, the receptors in the skin

for pain, the neuromuscular spindle etc.) to the CNS, and from the CNS to executive structures (e. g. the muscles, the adenoids etc.). The other part of the neural system is the CNS. CNS consists of the spinal cord, the brainstem, the cerebellum and the cerebrum. (n67) There are overwhelming evidences that the nervous system consists of neurons which are interconnected with another in specific patterns. While this neural pattern as a whole seems to be fixed in the PNS, the neural pattern of CNS seems to change constantly.

Neurons

As the medical science advances, neurons can now be isolated and studied. What scientists have found out is that neurons relay electro–biochemical signals according to the all–or–none law. Signals are *either* blocked or *relayed*, but for neurons there are no 'weaker' or 'stronger' signals.

Another interesting thing about neurons is that they seem to be able to 'represent' something. We have now detailed knowledge about what happens between the 'activating' of the sensory cells, the signal transmission through the periphery neurons and the signal relaying up to the thalamus. When a sensory cell, say a rod cell in the eye, catches light, it 'sends' the biochemical signal to a periphery neuron in the optical nerve. In an 'appreciation' of values from different rod cells the neuron in the optical nerve terminates or relays the signal to thalamus. The person will notice that there is a light spot, if and only if the neuron in the optical nerve has transmitted the signal. When we stimulate the neuron in an unnatural way (e. g. per electric impulses), the person will be aware of a false light spot.

Representational Schema and Second–Order Neural Pattern

It is assumed that these signals from PNS will be transmitted via 'relay stations' (e. g. the thalamus and other diencephalon areas) to the cortex. In a simplified way we can content that there are neurons (or neuron groups), which are stimulated accordingly to specific neurons in the PNS. E. g. the neurons in the occipital lobe are activated, when neurons in the optical nerve are activated. (n68)

Now is the right time to introduce the distinction between proprioception and exteroception. Both terms refer to human senses: While the proprioception senses the own body (e. g. the muscle tonicity), the exteroception catches the outer world (e. g. the optical sense). It is possible to map the corresponding areas of the brain to the proprioception. By doing so we can get the so–called sensory body schema(n69). There are other body schemata as well, e. g. the motor body schema, by mapping the motor functions, and the functional brain map, by applying the major functions (e. g. the language centers or the visual centers) to the corresponding areas.

We can call the sensory body schema as the primary representation of one's body. According to Damasio, there must be a so–called "second–order neural pattern" (SONP), which represents the body schema(n70). What is SONP anyway? This is the explication Damasio gives:

The main characteristics of the second–order structures whose interaction generates the second–order map are as follows: A second–order structure must (1) be able to receive signals via axon pathways signals from sites involved in representing the proto–self *and* from sites that can potentially represent an object; (2) be able to generate a neural pattern that "describes," in a temporally ordered manner, the events occurring in the first–order maps; (3) be able, directly or indirectly, to introduce the image resulting from the neural pattern in the overall flow of images we call thought; and (4) be able, directly or indirectly, to signal back to the structures processing the object so that the object image can be enhanced. (n71)

It is important to note that SONP is neither a special brain region, nor is there only *one* layer of SONP. There are probably several different structures of SONP, and there are also SONP which represents other SONP–structures, i. e. SONP itself can be iterated.

Body Image

Beside the body schema there are also other concepts of 'self–representation'. One of them is called "body image"(n72). Stamenov explains: "Unlike body schemas, the body image is envisaged as *available for conscious experience* and as *possessing an integrated, unified multimodal character* due to the simultaneous representation of visual, tactile and motor information of corporeal origin"(n73). The crucial point of the difference between 'body schema' and 'body image' is their availability for consciousness. While 'body schema' *never* becomes conscious, the body image can. The body image is the "picture of one's self in the mind".

When people are asked to remember their last travel to the sea, some of them will have the image in the mind, how the sun shined and how cold the water was. They remember their experience from a first person point of view. But some of them will have a third person point of view about themselves. They will have the image in which they can "see" themselves on the beach, mostly from the behind. How can this be possible? These people surely have never *seen* themselves on the beach, aside perhaps on some photographs.

It is assumed that the body image is a self–representation out of the body schema. How this is done exactly is still under debate. Stamenov himself suggests that the so–called "mirror–neurons"(n74) might get involved with the creation of body image. (n75) In my opinion, which neuronal structure is ever

responsible for bringing out the body image, in any case it must be some kind of SONP.

Why should we distinguish between 'body schema' and 'body image'? As to the psychiatry this distinction can be the foundation of a model for the mental illness 'schizophrenia'. As A. Mishara points out, when the body schema and the body image of the patient have got separated, schizophrenia might be the consequence. (n76)

3.3 Consciousness Depends on the Neural Network

Many medical phenomena hint at the thesis that (human) 'consciousness' somehow depends on a functional neuronal network, which is the human brain. Studies from brain traumata, patho–neurological cases, and other intriguing reports about consciousness and CNS are well–known and do not need to be rehearsed here. Patients showing the phenomenon of blindsight(n77), suffering from chronic amnesia(n78), or being in anesthesia give us clues that human consciousness is linked with the bio–physiological functioning of the brain.

Here I want to present another approach to the dependence of the consciousness on the basic components. Let us review the interconnection between 'body image', 'body schema', and 'consciousness'. As stated in [3.1], something is conscious if and only if it is aware of something and it is aware of their awareness. As far as the *self–awareness* is concerned, the neuronal self–representation fits perfectly here. With the theory of body schema, we have a theory about how the self–representation of an organism with a CNS can be done. There are conclusive evidences that the body schema is linked with the self–representation. When physicians perform the epidural anesthesia on a patient, they cut the connection between the lower spinal cord and the upper parts of the spinal cord temporarily. With the epidural anesthesia the patient does not *feel* anything lower than the segment of the spinal cord which is blocked. She will not be conscious about her feet, for example.

With 'SONP' and 'body image' we have another theory which can explain *how* conscious self–representation could be possible. As suggested in [3.2] SONP could be the neuronal structure for the body image. Since SONP itself is still a theory, we do not have many evidences about the relation between SONP and self–representation. But perhaps we can contrast SONP to simple neuronal network. By doing so we might get a hint that with SONP the probability of 'self–consciousness' increases. Let us take Kinorhyncha(n79) as an example. The anatomy of this small worm is relatively simple and therefore very suitable for studying. Though a Kinorhyncha has got ganglions, and therefore it is possible that Kinorhyncha does have a kind of body schema, there

are not enough nerve cells to build up SONP. According to our theory, Kinorhyncha cannot have 'self–consciousness', though it will be able to be aware of its body.

I think most people do accept the thesis that a Kinorhyncha does not have (higher leveled) self–consciousness. If this is truly the case, SONP (or at least some structures similar to SONP) is required for the being conscious.

If these hints are conclusive, then the thesis that the systemic property 'being conscious' is *based* on basic neural parts. The first step to prove that consciousness is an emergent property is made.

3.4 Consciousness is Irreducible

The second step of the argument is more important. I want to show that the systemic property 'consciousness' can neither be reduced to the properties of neurons, nor can we deduce 'consciousness' out of the relation between neurons or neuron–groups. Though there are other arguments for the irreducibility of the mind(n80) an approach with the self–representation should be presented here.

Let us first regard the property "representing something", the first part of our explication of 'consciousness'. I think there is little doubt about our own consciousness, and that (at least) human mind is able to represent something. Prima facie no neuron, or any other material entities of a lower level of existence(n81) can actually be able to represent, i. e. there is nothing in a lower level of existence which *stands for* a specific entity. But the representation could perhaps be realized in the structural configuration of the neurons.

As stated in [3.2] the neurons in the occipital lobe get stimulated when the optical nerves are activated. Since there is a constant chain of neurons from the retina to the occipital lobe, we can content that the representation of visual sensation is done by the structural configuration of a whole neuron–group: The stimulation of the neurons in the occipital lobe *is* the representation of the stimulation of the neurons in the optical nerves. If this explanation is correct, then it seems that we *can* actually reduce the ability of *representation*. (n82) The structural configuration of neurons in PNS and subsequently the structural configuration in CNS are capable for representation and transportation of information. Though the ability of *representation* might not be fully reduced, this explanation is 'good enough'. Let us take the reduction of representation for granted at the moment.

In the next step, let us try to reduce the self–representation in a similar manner. Since the neuronal representation of the outer world is done by the neural configuration, it can be assumed that the self–representation is realized in neural configurations, too. As the concept of 'body schema' suggests, the neural self–representation can be grounded in the proprioception and the neu-

ral configuration of the neurons concerning the proprioception. We can call this the reduction of the *primitive*(n83) self–representation. With the reduction of the representation and the reduction of the primitive self–representation the second part of (C), the Brentano–Husserl–Thesis, could be seen as fulfilled. But what about the first part of (C), the phenomenal consciousness?

It would seem that the *complex* self–representation could be reduced to SONP, but this task is much harder to be completed. The problem here is the difficulty to decide, *what* is represented by the stimulation of a neuron in SONP?

Suggest that there are three neurons N_1, N_2, and N_a. Suppose that N_1 represents a proprioceptive sense, e. g. contraction of a muscle spindle, suppose that N_2 'encodes' another proprioceptive sense, e. g. pain in the muscle, and let us suppose that N_a–as a neuron in SONP–can be stimulated by both N_1 and N_2. What does the activation of N_a represent? The contraction of a muscle spindle? Or the pain in the muscle? Or "contraction or pain of the muscle"? In either case it is hard to imagine that from the structural configuration of neurons a 'body image' could come into existence.

The situation gets more complicated when we regard the distinction between one's self and the 'outer world'. Suggest N_3 represents the pain in the leg, N_4 an optical sensation, and N_b, as part of SONP, gets its inputs from N_3 and N_4. What does N_b represent when it get activated? With this neuron being stimulated we cannot even tell what N_b is suppose to do.

The perhaps most intriguing situation arises, when we take the circular stimulation into account. There are neuronal configuration which can re–stimulate themselves. According to the theory of synchronic oscillations, which is proposed as an answer to the "binding problem", there are neurons ordered as a circle. (n84) This neuronal circle is 'responsible' for the 'binding' of our sensations. Suppose three neurons Nc, Nd, and Ne form such a neuronal circle. Just like Na and Nb, whose existence are well documented, we cannot give an answer what Nc, Nd, and Ne are representing.

It might be objected here, that it is well possible that neurons in SONP do *not* represent anything, but they merely take neuronal inputs and put them together. If SONP is not *for* representing anything, then it does not make sense to ask what they are representing. If this is truly the case, then 'body images' will *not* be thinkable. The argument for the impossibility of 'body image' out of anything other than SONP goes like this: (1) The 'body image' is a conscious *representation* of one's self, and there is a phenomenon in the world including 'body image'; (2) the 'body image' is strongly linked with the body schema (n85), but there could also be a mismatch between the two kinds of body–representation, e. g. in case of schizophrenia(n86); (3) there is something we might call 'second–order neural pattern', and the existence of SONP is a necessary condition for a conscious self–representation(n87). If all three premises

are accepted it seems to be plausible to suggest that 'having a body image' implies 'having SONP representing one's self'. At least some part of SONP must represent one's self, otherwise there could not be a body image.

If this argument is sound, than it would suggest that 'having a body image' cannot be reduced to the structure of SONP. And we can follow that the property of a human being, 'consciousness', can neither be reduced to the properties of her basic components (e. g. neurons), nor can it be reduced to the structural configuration.

4. *Conclusion*

By now, I have provided an overview over ontological emergentisms (in [1]), and presented a draft of 'substance emergentism' (in [2]). In my opinion, an Aristotelian substance is a system with (at least) one irreducible systemic property. The empirical data from neurobiology, the two different kinds of self–representation, and an analysis of the notion of 'consciousness' (in [3]) suggest that 'consciousness' is an irreducible systemic property. If this can be accepted, the next question we have to answer is the question of the system with such an irreducible property.

Prima facie two systems come into questions: the human being and the human brain. There are arguments for the thesis that human beings, but not the human brains, should be regarded as substance. (n88) Here I want to present a corollary from other arguments. If the argument presented in [3] is correct, than there are two crucial distinctions: the one between 'self' and 'outer world', and the one between 'body image' and 'body schema'. Consider the brain in a vat, like the famous thought experiment by H. Putnam(n89): The brain cannot differentiate correctly between 'self' and 'outer world'. For it, every inputs must come from the out side. There can be no distinction between 'body schema' and 'body image', neither. So the argument presented in [3.4] cannot be applied to a brain correctly. Normally, the human being can distinguish between her self and the outer world. She has also two kinds of body–representation: body schemas and body images. So the human being is more suitable to be said to have the emergent property 'consciousness'.

Notes:

n1) Cf. O'Connor/Wong (2005), Heard (2006).
n2) Cf. O'Connor/Wong (2005).
n3) Cf. Stephan (1999), 6.
n4) Clayton (2004), 40–42. Clayton analyzes only the academic use of 'emergence'. In the ordinary language we can also find the word 'emerge', e. g. in the following sentence "a submarine emerges from the surface of the ocean". We should keep in mind that though the emergence theories developed by the early twentieth–century British emergentists give answers to both metaphysical and evolutionary questions, these questions are not interde-

pendent to each other by default. Some modern emergentists (e. g. O'Connor/Wong (2005), Silberstein/McGeever (1999)) develop only metaphysical emergence theories.

n5) Clayton (2004), 42. Clayton himself does not distinguish between the diachronic emergence in the evolution of the life-forms and the synchronic emergence, e. g. the emergence of the mental from the physical.

n6) This distinction is also made by D. Heard (Heard (2006)), Van Gulick (Van Gulick (2001)) and other philosophers.

n7) Cf. Stephan (1999), Chap. 3, 14–65. There is another way to bring in an overview. Van Gulick (Van Gulick (2001)) does it by separating theories of emergent properties from theories of emergent causal powers.

n8) O'Connor and Wong use the term 'left hand side' to refer to dualism and the 'right hand side' to refer to reductive physicalism (O'Connor/Wong (2005)).

n9) Specifically I will not deal with the "unpredictability" and the "irreducibility of the emergent laws".

n10) Cf. Löffler (1999), 32–35.

n11) Cf. Löffler (1999), 34, translation by jw.

n12) According to Stephan the naturlism-claim is composed of three different claims: (i) There are no other causal powers than the natural ones within the evolution; (ii) every thing in the universe is composed by the natural, i. e. the material, parts; and (iii) there is no extra parts in the living or mental systems (Stephan (1999), 15). Since (iii) follows directly from (ii) and (i) will be dealt with the thesis of 'downward causation', I only presented (ii) here.

n13) Not every emergentist uses the term 'systemic property' in their theory. According to Lowe (Lowe (1993), 634) and Hasker (Hasker (1999), 172ff) John Searle himself uses the notion of 'emergent1 property' to refer to 'weak systemic property'. Because I do not distinguish between weak and strong systemic properties, (2) can be accepted widely.

n14) Cf. Stephan (1999), 23.

n15) E. g. Clayton (in Clayton (2004), Chap. 5) makes some speculations about 'Emergence and Transcendence'; W. Kofler enlisted some 'evolutionary levels' (cf. Kofler (1993)) which is important for public health systems.

n16) According to Stephan the diachronic determinateness ("diachrone Determiniertheit") is a thesis about how the world is going to evolve. Determinism can be seen as a strong claim of diachronic determinateness. By contrast the synchronic determinateness ("synchrone Determiniertheit") is the thesis that the upper level entity is determined by the lower level entities. Cf. Stephan (1999), 26–31.

n17) In my judgment, by default the notion of 'nature laws' contains only the laws of 'physics'. It does not presuppose that the nature laws must be known.

n18) See: O'Connor / Wong (2005). E. J. Lowe also suggests a similar way to understand the relation between s and its parts; cf. Lowe (1993).

n19) In Kim's view (Kim (1999)) the synchronic determinateness is not a causal relation, because the notion of 'synchronic determinateness' is similar to the concept of 'supervenience', and 'supervenience' should not be seen as causal. O'Connor and Wong (O'Connor / Wong (2005)) explicitly deny this. In their view the parts of a system do cause the system to exist. If the relation between the parts and the system is a causal one, then we can affirm both the causal closure of the world and the ontological emergentism.

n20) The irreducibility thesis is often interpreted epistemologically only: What makes an emergent property irreducible is not that the emergent property is new, but the fact that we know too little about the parts. In this interpretation the irreducibility is mere a contingent fact; when we get enough information about the parts in the future, we will be able to analyze the emergent property to the properties of the parts.

n21) Stephan (1999), 35. He quotes the works of Popper and Nagel there.

n22) Some might object here that (5) already contradicts with (4a) and (4b). If we already know that (4b) is the case, then we can postulate that a part of s, say c1, has the following property: When c1 is joined with c2, c3... cn in the o way, then s (=, c2, c3... cn d o) will have the property p. With this kind of 'ad hoc properties' (cf. Beckermann (1992), 104) and with (4b) accepted we can reduce every emergent property to the property of the part. It is

clear that the emergentism can only make sense if ad hoc properties can be sorted out. I find it difficult to enunciate the intuition of 'irreducibility' properly, perhaps the formulation of Beckermann is useful here, cf. Beckermann (1992), especially 114f..

n23) Whether (5b) is a strong emergentism or not, depends on the ontological status of nature laws in question. If nature laws are supposed to be nothing else but ideas of human minds, then (5c) does not contradict, but presupposes (5b). If we assume that nature laws are universal entities, then (5b) is weaker then (5c), since (5b) cannot exclude that emergent properties can reduced to the submergent properties.

n24) Silberstein / McGeever (1999).

n25) Silberstein / McGeever (1999).

n26) Cf. Stephan (1999), 64. Again, I will leave out the passages concerning the evolutionary emergentism.

n27) The notion of 'special science' comes from the British emergentists. According to Broad (and also to Clayton (cf. Clayton (2004), 40ff)) we should have different sciences for different levels of emergence. Cf. Beckermann (1992).

n28) According to B. McLaughlin emergence theories have lost their plausibility when quantum mechanics has been 'discovered', cg. McLaughlin (1992), 54–55.

n29) One defender of this kind of 'downward causation' is W. Hasker, cf. Hasker (1999), 171–203, particularly 184–185.

n30) Stephan (1999), 197–218. 'Pepper' refers to Stephen Pepper who presented the epiphenomenal argument against the emergence theory in 1926; and 'Kim' refers to Jaegwon Kim. While for Pepper the emergentism is an epiphenomenalism, for Kim 'emergence' is only acceptable as a genuine ontological position, if emergentism denies the epiphenomenalism.

n31) Here I cite the version of Alexander's Dictum by Stephan, cf. Stephan (1999), 212. A more general formulation of the Alexander's Dictum which I do not reject contents that "all existents possess causal power" (O'Connor/Wong (2005), 663). Another wide understanding of Alexander's Dictum can be found in Clarke (1999), 296; it states: "To be real is to have causal powers."

n32) Kim (1999).

n33) Kim is perfect aware that we should distinguish between 'determination' and 'causation (cf. Kim (1999), 32 and Footnote 36). If this distinction is not met, the causal power of M will be useless. If this distinction is met, Kim asks for evidences.

n34) Kim (1999), 32.

n35) In addition I also want to reject the Alexander's dictum (c). This will be explained in [2].

n36) O'Connor / Wong (2005).

n37) I will deal with this in [2].

n38) Taken from O'Connor / Wong (2005), 666.

n39) J. Haldane has presented this approach in Haldane (1998). Similar ideas can be found in Brüntrup (1999).

n40) If emergentism is only understood as an epistemological theory there will be no direct contradiction between hylemorphism and emergentism. But, at least in my opinion, some ontological emergentism is incompatible with the Aristotelian hylemorphism.

n41) See Hoffman / Rosenkrantz (1997), Chap. 5, 150–187.

n42) Cf.: Aristotle: Metaphysics, Book VII–IX (1028a–1052a).

n43) Cf. Aristotle: De anima, Book 2, Chap. 1, 412b15.

n44) Cf. e. g. Thomas Aquinas: Summa contra gentiles, lib. 2, cap. 57. Though both Thomas and Aristotle have touched the subject of the independence of the soul, i. e. how a soul can survive even when the human being has died, I will not go into this. What I want to present here is an emergentism which is able to view consciousness as an emergent property, but not a speculation about the after-life.

n45) According to Hoffman and Rosenkrantz (Hoffman / Rosenkrantz (1997), chapter 2, 43–72, and 40f.) the independence of existence is one of the necessary criteria for being a substance.

n46) Since some substances can loose some of their causal powers without ceasing to exist 'having the causal power to do x' is perfectly accidental. But if Alexander's dictum is right then loosing all causal powers will be equivalent to non–existence, for this reason 'having (at least) one causal power' is an essential quality.
n47) The counterpart of the interpretation of 'substance causation' is the so–called 'event–causation'. In this paper I will defend the 'substance causation', but leave out the debate between the two concepts.
n48) Lowe (1998), 121–125.
n49) I must admit that there is the probability that there are no emergent substances at all. What matters in this part of the paper is that if there are emergent substances they must have emergent properties.
n50) I cannot think of any property which could be irreducible, but let us–for the sake of the argument–assume that this is truly so.
n51) Just like classes which can be formed arbitrarily one can form a system out of any parts. While the system 'liver' probably does have a 'principle of composition' this is not required. One can form e. g. the system 'el' out of one left eye and one left leg. The system 'el' will still have some systemic property, but it is not probable that 'el' will have any emergent properties which would qualify 'el' as a substance.
n52) In this paper there is sadly no place for a detailed discussion of the problem of gaining and loosing parts.
n53) Hoffman and Rosenkrantz also admit this, cf. Hoffman/Rosenkrantz (1997), 150–154.
n54) Aristotle has dealt with the possibility of an "independent soul" (cf. Aristotle: De anima, Book 3, Chap. 4, (429a); his argument is not quite clear, though.
n55) See e. g. Hoffman / Rosenkrantz (1997), 163–165.
n56) A traditional Aristotelian philosopher uses the forma to guarantee the diachronic identity of s, but this cannot be done within the ontological emergentism.
n57) For those who are not convinced by this argument: There are other answers to this, see below in [3].
n58) What I have said about the ontological emergentism can be applied to all emergent substances. So there might be other emergent properties beside the human consciousness.
n59) I will not defend other possible emergent properties (e. g. 'to be alive', cf. Stephan (1998), 108–113) here. If the ontological emergentism is true then it is well thinkable that there will be emergent properties other then the human consciousness.
n60) Cf. Damasio (1999), 177–182.
n61) In order to defend the thesis that human consciousness is an emergent property of the human being, it seems that the best way to do it is not to discuss the topic about the existence of animal consciousness, but to show that consciousness is a property which must be attributed to the whole organism. And for that I do not need to discuss pathological cases of consciousness, but the rather higher–ordered 'self–consciousness'.
n62) Nagel (1974).
n63) This notions is taken from N. Block (cf. Block (1995)). G. Brüntrup also speaks of two aspects of 'consciousness' we must keep in mind while speaking about 'consciousness', cf. Brüntrup (1999), 108ff.
n64) Meixner (2003), 327, translated by jw., italics in original.
n65) Cf. Meixner (2003), 327.
n66) I must apologize for the disadvantageous use of the notion 'awareness'. This notion is sometimes used to make the notion of 'phenomenal consciousness' explicit, and to contrast phenomenal consciousness against the access consciousness, cf. Block (1995).
Note that (C) is not a definition, since I do not use the equivalence ("if and only if"), but a simple subjunction ('if') in (C). There might be entities that fulfill both the 'phenomenal consciousness' and the 'Brentano–Husserl–Thesis', but do not have consciousness.
n67) There is a third part of the neural system, called "vegetative neural system" (VNS) which is sited in the stomach and the intestines. VNS is responsible for the bowel movement.

n68) A second way to recognize the corresponding areas is to study brain lesions, and this is the usual way to study functions of the brain. When a patient suffers from brain lesion, physicians document both the lesion and the neurological deficits the patient has. So a body schema can be made upon medical evidences.
n69) According to Stamenov (cf. Stamenov (2005)) W. Penfield and T. Rasmussen have published the famous sensory and motor homunculi in their work "The Cerebral Cortex of Man: A clinical study of localization of function", New York: MacMillan, 1950. A picture of these homunculi can be found in the Wikipedia, internet–resource: http: //en. wikipedia. org/wiki/Image: Sensory_and_motor_homunc uli. jpg.
n70) Cf. Damasio (1999), 177f. Note that Damasio himself speaks of 'proto–self', which is not the same as the body schema. To Damasio, there can be SONP for any neural structures, not only for the proprioception.
n71) Damasio (1999), 177.
n72) Cf. Stamenov (2005).
n73) Stamenov (2005), 28.
n74) Mirror–neurons are neurons that are activated when either one move herself or sees another one moves. E. g. when a certain mirror–neuron is stimulated, say, when the person waves with her hand, this neuron also gets activated, when she merely sees another person waving with the hand.
n75) Cf. Stamenov (2005).
n76) Cf. Mishara (2005).
n77) Cf. Block (1995).
n78) E. g. the case of David, cf. Damasio (1999), 43–47.
n79) 'Kinorhyncha' is a kind of very small worms (about 0, 2 to 1 mm long). For detailed descriptions of the anatomy of Kinorhyncha, cf. Müller / Schmidt–Rhaesa (2003), Nebelsick (1993).
n80) In my opinion the irreducibility of qualia, the so–called 'hard problem' (cf. Chalmers (1996), xiii), is one of the most discussed argument. Another type of argument stresses the irreducibility of the subjectivity to objectivity, cf. Feinberg (2001).
n81) Cf. thesis (3) in [1.1]. These entities are e. g. quanta, atoms, molecules, proteins, and cell organelles.
n82) Some might object here that the property, or (more precisely) the 'act of representing an object', cannot be reduced in this way. In fact it cannot be reduced at all, but has to be regarded as 'simple' or 'basic'. I cannot deal with this argument here properly, but in my opinion this argument presupposes other notion of 'irreducibility' than I do by stating the thesis (5c).
n83) The primitive self–representation is also 'possible' in lower life forms like Kinorhyncha. But the complex self–representation, as opposed to the primitive self–representation, involves the development of a 'body image'. It is widely accepted that lower life forms like Kinorhyncha do not have a complex self–representation.
n84) Cf. Crick / Koch (1990).
n85) Cf. Stamenov (2005).
n86) Cf. Mishara (2005).
n87) Cf. Damasio (1999), 176–182.
n88) E. g.: According to Hoffman and Rosenkrantz an organic substance must be an independent, living being. A human brain, if taken alone, is though the center of coordination of an organism, but cannot live without other parts of the organism. So the human being is a substance, while the brain alone is not. Cf.: Hoffman / Rosenkrantz (1997), Chap. 4 (91–149), and 177–179.
n89) Cf. Putnam (2000).

Literature:

Beckermann, Ansgar (1992): Supervenience, Emergence, and Reduction. In: Beckermann / Flohr / Kim (1992), 94–118.

Beckermann, Ansgar / Flohr, Hans / Kim, Jaegwon (Eds.) (1992): Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism. Berlin, New York: Walter de Gruyter.

Block, Ned (1995): On a Confusion about a Function of Consciousness, in: The Behavioral and Brain Sciences, vol. 18(2), 227–287.

Brüntrup, Godehard (1999): Das Geheimnis des Erlebens — Grundzüge und Grenzlinien einer fundamentalen Theorie des Bewußtseins. In: Quitterer / Runggaldier (1999), 104–129.

Clarke, Randolph (1999): Nonreductive Physicalism and the Causal Powers of the Mental. In: Erkenntnis, Vol. 51, 295–322.

Chalmers, David J. (1996): The Conscious Mind. In Search of a Fundamental Theory. New York, Oxford: Oxford Univ. Press.

Clayton, Philip (2004): Mind and Emergence. From Quantum to Consciousness. Oxford, New York: Oxford University Press.

Crick, Francis / Koch, Christof (1990): Towards a neurobiological theory of consciousness. In: Seminars in the Neurosciences, vol. 2, 263–275.

Damasio, Antonio R. (1999): The Feeling of What Happens. Body and Emotions in the Making of Consciousness. New York, London: Harcourt Brace & Company.

De Preester, Helena / Knockaert, Veroniek (Eds.) (2005): Body Image and Body Schema. Interdisciplinary perspectives on the body. Amsterdam, Philadelphia: John Benhamins Publishing Company.

Feinberg, Todd E. (2001): Why the Mind is Not a Radically Emergent Feature of the Brain. In: Journal of Consciousness Studies, vol. 8 (9–10), 123–145.

Haldane, John (1998): A return to form in the philosophy of mind. In: Ratio (new series), vol. XI, 253–277.

Hasker, William (1999): The Emergent Self. Ithaca, London: Cornell Univ. Press.

Heard, Daniel (2006): A New Problem for Ontological Emergence. In: The Philosophical Quarterly, Vol. 56, 55–62.

Hoffman, Joshua / Rosenkrantz, Gary S. (1997): *Substances, Its Nature and Existence*, London: Routledge.

Kim, Jaegwon (1999): Making Sense of Emergence. In: Philosophical Studies, Vol. 95, 3–36.

Kofler, Walter W. (1993): SUNDS und die Kybernetische Evolutionäre Medizin. In: Wallner, Fritz G. / Schimmer, Josef / Costazza Markus (Eds.): Grenzziehungen zum Konstruktiven Realismus, Beiträge zweier Kongresse. Vienna: WUV–Univ. –Verlag, (Cognitive Science 4), 200–227.

Löffler, Winfried (1999): Naturalisierungsprogramme und ihre methodologischen Grenzen. In: Quitterer / Runggaldier (1999), 30–76.

Lowe, Ernest Jonathan (1993): The Causal Autonomy of the Mind. In: Mind, vol. 102, 629–644.

Lowe, Ernest Jonathan (1998): The Possibility of Metaphysics. Substance, Identity, and Time. Oxford, New York: Clarendon Press.

McLaughlin, Brian (1992): The Rise and Fall of British Emergentism. In: Beckermann / Flohr / Kim (1992), 49–93.

Meixner, Uwe (2003): Die Aktualität Husserls für die moderne Philosophie des Geistes. In: Meixner, Uwe / Newen, Albert (Eds.): Seele, Denken, Bewusstsein. Zur Geschichte der Philosophie des Geistes. Berlin, New York: de Gruyter.

Mishara, Aaron L. (2005): Body self and its narrative representation in schizophrenia. Does the body schema concept help establish a core deficit? In: De Preester / Knockaert (2005), 127–152.

Müller, Monika C. M. / Schmidt-Rhaesa, Andreas (2003): Reconstruction of the Muscle System in *Antygomonas* sp. (Kinorhyncha, Cyclorhagida) by Means of Phalloidin Labeling and cLSM. In: Journal of Morphology, vol. 256, 103–110.

Nagel, Thomas (1974): What is it like to be a bat? In: The Philosophical Review, vol. 83, 435–50.

Nebelsick, Monika (1993): Introvert, mouth cone, and nervous system of Echinoderes capitatus (Kinorhyncha, Cyclorhagida) and implications for the phylogenetic relationships of Kinorhyncha. In: Zoomorphology, vol. 113, 211–232.

O'Connor, Timothy / Wong, Hong Yu (2005): The Metaphysics of Emergence. In: Nous, vol. 39, 658–678.

Putnam, Hilary (2000): Brains in a Vat. In: Bernecker, Sven (Ed.): Knowledge. Readings in Contemporary Epistemology. Oxford: Oxford Univ. Press.

Quitterer, Josef / Runggaldier, Edmund (Eds.): Der neue Naturalismus, eine Herausforderung an das christliche Menschenbild. Stuttgart, Berlin, Köln: Kohlhammer.

Silberstein, Michael / McGeever, John (1999): The Search for Ontological Emergence. In: The Philosophical Quarterly, vol. 49, 182–200.

Stamenov, Maxim I. (2005): Body schema, body image, and mirror neurons. In: De Preester / Knockaert (2005), 21–43.

Stephan, Achim (1999): Emergenz. Von der Unvorhersagbarkeit zur Selbstorganisation. Dresden, Munich: Dresden Univ. Press.

Van Gulick, Robert (2001): Reduction, Emergence and Other Recent Options on the Mind/Body Problem, a Philosophic Overview. In: Journal of Consciousness Studies, Vol. 8, No. 9–10, 1–34.