(This is a part of the book The Concept of Reality.pdf)

Contents:

5. Qualia

5.1. Introduction

Now we have finally arrived at the question that contains the real secret of the mind: the question about the reason for the fundamental transformation of a network of information into a world of sensations.

How does information turn into sensation? In which way does the machine turn into a sentient being? What is the difference between the two?

For everything that has been said so far about mind, it was sufficient to regard mental activity as information processing. The autonomy and dominance of the mind, as well as the existence of free will, could be substantiated in this way.

But now, when we ask how a sequence of neuronal patterns can be a stream of experiences, an answer based only on this assumption can no longer suffice. As long as one acts on the assumption that mental processes are nothing but information processing, one remains captured in the area of information processing. It does not change anything, if one networks representations, or forms representations of representations – i.e. higher levels of information processing –, or lets information act back on itself: no matter what function is applied to information – the result will always be just information and nothing else. In this view, no metamorphosis can occur; the information "red" does not turn into the sensation *red,* the information "pressure" does not turn into the sensation *pain.*

Therefore applies:

The assumption that mind is information processing, is necessary and sufficient for the derivation of the freedom of will.[1] But for the substantiation of the qualia, it is only necessary and not sufficient.

In other words: The assumption that mind is *nothing but* information processing, is wrong.

First it must be clarified why qualia are not contained in any description. Thereafter we will focus on the question, why a physical and physiological state turns into a quale.

[1] However, this applies only to information processing in *biological* neuronal networks. More on that follows later.

Meaning

Mental states relate to each other. In the case of notions, this is a matter of course. But it is also true for perceptions: even if they represent an object, their content is still determined not only by this object but also by the relations with other elements of the mental reality.

Representations, which are not networked with each other, are not mental states. The robot *John* (section 2.2) represents the attribute of an object, which we perceive as *red,* but, in the case of the robot, this representation is not a mental state.

Therefore, networking the representation-states is a necessary condition for the possibility that representation-states can ascend to mental states.

In order to discriminate mental states and representation-states, I shall denote the ***information content of a mental state*** – which, according to the above notes, is not only defined by the represented object or the represented situation but also by the relations with other mental states – as ***intrinsic meaning*** or just as ***meaning***.

5.2. Insertion: strange Conjectures

Lately, during my investigations I ran across two false hypotheses so frequently that I would like to briefly respond to them.

However I see this section not as part of the serious discussion about the topic "qualia", but rather as an entertaining insertion, to which I was seduced by an article that related to these two hypotheses and which I found so amusing that I was overcome by the desire to share my pleasure with someone else. (I'll be right back on it.)

The first hypothesis reads as follows:

If in an animal a neuronal structure is present that resembles a structure, without which humans cannot feel emotions, then there is reason for believing that this animal is capable of emotions.

The error is obvious: a necessary condition is mistaken for a sufficient condition.

Specifically it is about mid-brain structures, especially the amygdala, which is said to provide for the "emotional coloring" of neuronal processes – and this is not meant as a figure of speech, but as an ontological claim, as if emotions were colorful and the amygdala an ink pot.

What is the amygdala? A neural node. If one were to seriously ask what such structures actually can afford, the inadequacy of this paint-box idea would immediately come to light:

What are neural nodes able to do? Three things. They can

1. connect areas of the brain with each other or with other physiological structures (like extremities),
2. process information internally and in this way generate input-output relations,
3. connect the neuronal network with the chemical regulation systems of the organism.

The first two points illustrate what has been said in the introduction. Obviously, they do not lead out of the area of information, and the appearance of sensations remains unexplained.

So we come to the third point. What is meant by chemical regulatory systems? E.g. the transmitter system, or the endocrine system. Through the connection of the neuronal network with these systems, neuronal activities – and, at the same time, the objects or situations that are represented by them – are being *assessed,* in the sense that now certain behaviors are triggered or modified, which basically can be divided into two classes: Strive for or avoid.

Can neuronal structures, which connect information and behavior in this way, cause the appearance of emotions?

If one takes this assumption, one is immediately led to absurd consequences. Indeed such structures are found in every living organism, even in the simplest one, because they are necessary for survival. Each animal must be able to distinguish what is favorable or unfavorable for its survival, it needs to avoid danger and find food.

Thus also a nematode has a so-called "reward system", and according to the aforementioned hypothesis, we would now have to assume that it was capable of emotions. Given its extremely small number of neurons, however, that is not very plausible.

But if anyone nonetheless thinks it would be possible, one could present him/her the following thought experiment:

Suppose we construct a robot – let's call it *Suzy* – who behaves as follows:

If Suzy is touched gently, then she generates pleasant melodious sounds and rolls closer. However, if one exerts a stronger pressure – so powerful that a human already would feel uncomfortable – then she remains in place and says "ouch!" And finally, if one presses very firmly or beats Suzy, she whines and rolls away, screaming loudly.

It is clear that Suzy's behavior can be produced with the simplest means. An optical sensor, a pressure sensor and some electronics and mechanics would be enough.

Thus, Suzy has certainly no feelings. But she has a structure that associates information with behavior, so that approach or escape is triggered. In other words, Suzy has a reward system!

Therefore, the statement of Victoria Braithwaite, who examines the question of whether fish feel pain[2]: "Although the appropriate structures and functions are much simpler than in the human limbic system, the discovery of the fish-equivalent is a most important finding" could be paraphrased by the statement: "Although the appropriate structures and functions are much simpler than in the human brain, the discovery of the Suzy-equivalent is a most important finding."

So the first of the two hypotheses which I criticize here is certainly wrong.

Actually, it was unnecessary to explain this in such detail, because it has been clear already before, that, on the basis of the current scientific interpretation of mind, the existence of certain neuronal structures represents only a necessary and not a sufficient condition for the occurrence of sensations or feelings.

(The fact that in the attempts to explain qualia this important difference is frequently not respected enough testifies in a certain way the explanation crisis of natural science, which however is inevitable in the current scientific world view.)

Now to the second hypothesis. It reads:

If an animal exhibits behavior which, when observed at a human, would indicate feelings, then this is a reason for the assumption that the animal feels something.

[2] Quoted (and translated) after Spiegel online, http://www.spiegel.de/spiegel/0,1518,749108,00.html

To refute this hypothesis, it suffices to recall the robot John, who behaves *as if* he had the sensation *red*, or Suzy, who behaves *as if* she felt fine or suffered pain.

"The ethologist Robert Elwood of Queen's University Belfast sprinkled the sensitive sensors of prawns with acetic solution. For up to five minutes the crustaceans rubbed their battered body parts – according to Elwood a reaction that reminds of the pain behavior of mammals."[3]

Yes, that's true, just as the behavior of John and Suzy reminds of the behavior of children. If Elwood thinks, however, his observation was an argument that crustaceans feel something – and probably he thinks so, because why else would he have spoken in this way – then one must inform him that from his observation only follows that prawns rub their sensors, if they have been wetted with acetic solution, and nothing else.

It is rather amazing that such statements find attention at all. Obviously, from the rubbing of the sensors nothing more can be concluded than that a certain stimulus triggers an associated behavior which may then last even longer than the stimulus persists. Everything else is not a conclusion, but the expression of subjective inclinations.

Just to not be misunderstood: If a person – let's call him Robert – prefers to embrace his prawn instead of eating it, or if another person – let's call her Victoria – wants to share her good fortune with a nematode, then I would be the last one who tried to stop them;– and I am serious about that, because I consider the variety and diversity in the garden of humanity a high value, and I think it would be unfortunate if its soil would be so poisoned by our rationalization- and optimization-mania that only the greed for money, power and pleasure could thrive on it.

But I'd suggest Robert and Victoria, to simply follow their inclinations, and not seek to substantiate them in an ostensibly rational manner.

Finally, as funny conclusion, a statement by the author of the cited article himself:

"To many experts [...], the absence of the cerebral cortex alone no longer seems to be sufficient to rule out conscious sensations. Doubts about the old doctrine are not least nurtured by amazing medical case histories: Occasionally, neurologists report on people with only half the cerebrum. Where in others brain cells talk, in them just nerval water swashes – and yet they are often highly intelligent and socially inconspicuous."

[3] Quoted (and translated) from the same report.

Yes, and if one sees a three-legged dog run, then one wonders involuntarily if the old doctrine that dogs need legs for running shouldn't have been revised long ago!

5.3. Why Qualia are not contained in any Description

Once again: the Problem

The failure of the current scientific view of the relationship between neuronal networks and mind reveals itself in the confrontation of the following two issues. Both are so well secured that they can claim the status of facts.

1. Regardless of how one describes or analyzes the brain or a region of the brain, one will always find just informational or representational states and never a *quale*.

2. The neuronal system brings forth *qualia*. Its states *are* qualia.

The *thing* that can be observed and described is a *dynamic neuronal excitation pattern*. It can be understood that this pattern *represents* something.

But that which this neuronal pattern *actually* is – the *quale* – seems to transcend the physicalness of the described thing in an inexplicable way.

Whence comes this irreconcilable difference between what it *is* and as what we describe it?

I think that through the last formulation it has become clear where the explanation is to be found: In the analysis of the relationship between actually existing things and things in a description, which was carried out at the beginning of the Second Part. The next subsection provides a brief recap with a few additions.

Substance and Accident; the Essence of Being

Reality *for us* consists of *things* that have *attributes.* What remains, if one imagines all attributes removed from the thing?

Obviously *nothing*, because a thing without attributes does not exist.

But *nothing* cannot have attributes. What is to be concluded from this contradiction?

It must be concluded that the actually existing thing, the *thing in itself,* does not correspond to the thing in the description, the *thing for us.*

What is the difference?

An actually existing object cannot disappear if all attributes are removed from it – simply because *nothing* cannot have attributes –, whereas an object, which is only an element of a description, indeed disappears completely if all attributes are removed.

This difference can be expressed in the following way:

Actually existing objects consist of *substance and accidents.* Objects in a description consist *only of accidents.*

Any description consists only of accidents. A description is always a network of relationships between objects that are defined by these relationships. The objects consist entirely of this definition, they *are nothing but this definition*, and therefore they disappear completely if the attributes are removed that are the prerequisites of the relationships.

Thus the substance cannot appear in the description. It is just presupposed through the assumption of the existence of the described thing. This defect of descriptions is irresolvable. It represents an insuperable limit of our thinking.

Yet the difference between reality and description can also be determined in another way:

There is no object that merely exists. Existing means interacting, and interacting means being active. An object that is not *active* does not exist.

Although in the description the *kind* of activity can be displayed through its attributes, the *activity itself* is still lacking. Objects in a description are always *passive.*

Therefore, what descriptions are lacking can be captured by to concepts: *substance* and *activity.*

These two concepts are inseparably connected with each other.

Before there is any being, their connection represents what the *origin of everything* is *for us:* The first substance is NOTHING, and its metaphysical quality *activity* becomes *for us* the first accident "activity". Both together is activity of NOTHING.

On being itself, the substance represents not only the condition for its existence, but also that from which the metaphysical quality *activity* of the being emanates, which drives the respective accidents – but only in the reality; in the description, the accidents lack this quality.

I have already spoken several times of the essence of being. Now I will catch up for its definition

Definition

The essence of a being is this being as unity of substance and accidents. It is the "in-itself-ness" of being, as far as it can be captured conceptually.

The essence of being cannot be thought. At the same time, however, it does not remain hidden. It can be approached conceptually.

I shall demonstrate this using the example of the essence of the physical being.

At the beginning, the first substance is NOTHING. But already this NOTHING contains a knowledge, namely the knowledge that that which remains after the removal of all attributes – though it does indeed no longer exist – is still not just identical with the purely conceptual nothing, because the conceptual nothing cannot have attributes.

Thus it must be different from the conceptual nothing, and precisely this difference is expressed by the notation with capital letters.

Now this distinction in turn makes it possible to determine the *ontological status* of the first substance:

The first substance cannot be *nothing*. However, since it lacks all accidents, it can also not be *something.*

Therefore, its ontological status is neither that of *being* nor that of *not-being,* but *necessity* – simply because there is no alternative for its neither-being-nor-not-being: If we thought the first substance as not-existing, then we would have chosen *nothing* as its ontological status, which means: one of the two alternatives that we have ruled out previously.

Already through these first steps it becomes a little clearer what is meant by "conceptual approach" to that which is actually unthinkable.

Let us repeat a few more steps of the derivation, which was conducted in the early chapters of the Second and Third Part:

Activity means change. However in order to achieve a statement from which the description of reality can follow, two different changes are needed.

Since space is a necessary condition of reality, and because without motion there is no change, one of the changes must be a change of space and the other one a change of motion.

Since there is no size, the changes can only be relative and cannot relate to an absolute measure.

Since there is no memory, the temporal change of motion must relate to the respective differentially adjacent preceding moment, that is: it must be expressed by the differential quotient dv/dt.

In this way, one arrives at the idea of a metric continuum, which consists of accelerated flows, where the acceleration depends on the alteration of the relative spatial density (of length or angle).

This means that every being is a pattern of alterations of the movement of the continuum, and, because every being conserves its form over a certain time period, it also means that every being is either an attractor of the local continuum dynamics or a part of an attractor of the global continuum dynamics.

With this, a further approach to the essence of the physical being has taken place. Through the just performed conclusions, the first substance NOTHING has turned into a space-time-continuum, and the first predicate *alteration* has assumed the form of equation (1), which represents the law of this continuum and expresses, what the physical reality is *for us:* a fabric of differential space-time alterations.

Let us dwell a little longer on the concept of the continuum, because by using this concept can be illustrated very clearly how far and to what limit the concept of *substance,* which cannot be thought directly, opens up to our thinking in an indirect manner.

What is a continuum?

For the sake of simplicity, let us look at a one-dimensional continuum. In standard analysis, the points of the continuum are mapped to the real numbers. This suggests that points *exist.*

But points have no existence! So if we say that in the continuum exists an accelerated flow, then this statement is only mathematically meaningful; if it is understood ontologically, then one is immediately confronted with the fact that actually *nothing* moves.

Another question is to *which kind* of numbers the points of the continuum actually can be mapped.

At first, it may seem surprising that the rational numbers do not suffice, though they are infinitely close to each other, which means that between two rational numbers, no matter how close they are, there are always infinitely many other rational numbers. Yet, as is easy to prove, there is still room for uncountably many more, namely the irrational numbers.

And then?

For a long time, mathematicians thought that the numbers would then be complete and that, if they are thought as points on a line, all possible points of the line were captured. But non-standard analysis shows that further numbers can be defined that again lie *between* the previous ones. This process can be continued *ad infinitum.* It is solely our decision when we stop.

This means that the continuum is a limit-concept. Ultimately, we cannot form an idea of it: if we divide it, then we lose it – but to describe it, we *have to* divide it.

Nevertheless, it is undeniable that the mathematical formulation *accelerating continuum flow* contains more information than the verbal expression *change of NOTHING*.

And exactly this approach, this increase of knowledge that is connected with the reasoning about the continuum, enables then further conclusions about the reality developing thereof.

If we investigate the continuum, then we learn more about the operations that are possible in it. If there are infinitely many, then, though we will never know them all, there is still no reason for the assumption that any of the accidents remains hidden for reasons *on principle.*

So we can rule out with some certainty that the first substance contains any secret which will elude our understanding forever.

Thus the relationship between our thinking and the concept of the essence of being is determined as follows:

The essence of being cannot be thought *as such,* because neither the substance itself nor the unity of substance and accident – the continuum that exists only as changing – can be thought

On the other hand, the accidents must be regarded as *manifestations* of the substance, and therefore applies:

Whatever *activity* the essence of being contains *must* manifest itself as accident, because otherwise it would not be there. For this reason, it can be asserted that the essence of being reveals itself through its accidents.

Therefore, the essence of being contains no secrets. Everything in it is revealed through its accidents. It is not thinkable, but also not hidden.

Since any being is a pattern of alterations of space-time – that is: of the *origin of everything* – the question arises whether being possesses also its ontological status *necessity* and its metaphysical quality *activity.*

Obviously, *necessity* is not transferred from the *origin of everything* to being: as pattern, it can dissolve. Every being stands in the alternative *be or not-be.* Necessary is only *that* something exists, but not *what* exists.

However *activity,* the metaphysical quality, is indeed transferred from the *origin of everything* to being. It appears in the form of the respective accidents; it is the *drive* behind the accidents of the actually existing things, that, what distinguishes the accidents of the actually existing things from the accidents of the things in a description.

The Answer

We perceive the table, and we think we know what it is. But if we try, to follow the question what it is up to its ground, then what it is disappears.

Precisely this disappearance of the substance has been encountered by the natural science of the 20th century, and it has plunged physics – though there is no longer much awareness of this issue – into a crisis from which it has not recovered to this day, because it has failed to capture the facts conceptually.

We perceive the waves moving towards the shore, and we think we know why they move. But the concept of energy is inappropriate for the explanation, because it is defined only mathematically and does not provide any differentiation between reality and description. Therefore, to the question of why anything moves at all, there is no answer before it is realized that the essence of being is *activity* and that this essence gets lost in the descriptions of reality.

This means: Being is always *more* than its concept, more than as what it appears in our perception and in our descriptions. It is always *substance and accidents,* whereas in the perception and in descriptions, *only accidents* appear.

That, what being actually is, its *essence*, the unity of substance and accident, is unthinkable for us. Thus we do not know about *any* being what it is. Only the confusion of reality and description – which, regarding material things is so habitual to us that we do not notice it – gives us the impression that we knew it.

The actually existing table is *more* than the table that we perceive and think. The actually existing wave is *more* than the wave that we describe.

And, with this, we have finally arrived at the answer to the question why an *existing* neuronal excitation pattern is *more* than a *thought* or *described* neuronal excitation pattern.

According to what has just been said, it is neither surprising nor difficult to understand why states of the neuronal network are not merely that, as what they appear in our description, but rather self-evident and necessary. They *must* be more than that, because their description can only contain that part of the state, which can be captured by a description, that is: the *accident*.

What is in this case the accident?

That what can be conceptualized, that is: the *definition* of the respective object, which, in the case of a neuronal pattern, consists of the information contained therein, i.e. of its *intrinsic meaning.*

This brings us to the following insight:

Mental states have an information content and a feeling content. The information content, the *meaning,* is the accident.

However, as everything which exists, also a mental state consists not only of *accidents* but also of *substance.* Thus it is *more* than meaning.

What is, in the description, missing of this "more-ness", what is lacking of the whole being, of the *essence* of the mental state?

Sensation. Therefore, sensation is the *substance* of mental states.

Proposition

Mental states are qualia. Meaning is their accident, sensation is their substance. As such, it is not contained in any description

The essence of the quale is its unity of sensation and meaning.

With this, we have solved the first part of the problem about the qualia: the question why neuronal patterns as *existing* entities are **more** than the objects, as which they appear in our descriptions, and why this "more-ness", the *quale,* is not contained in any description.

Let us now turn to the second part of the problem: the question why the substance in the case of qualia is not the same as in the case of other kinds of entities.

5.4. The Transformation of Being from the material Thing to the Quale

First and Second Substance

It has now become clear why we have called "space-time" the *first substance* and "change" the *first accident*: evidently, *mind* is an area of reality, where the essence of being has changed: here, being consists not only of other accidents but also of another substance.

I denominate ***sensation***, the substance of the mental states, as ***second substance***, and ***intrinsic meaning***, their accident, as ***second accident***.

However, the second substance is not to be thought as independent from the first substance or juxtaposed to it: **the second substance *emerges* from the first substance**.

How this emergence of mind from the neuronal network, this transformation of the physical being into the quale takes place is the subject of the subsequent investigation.

Let us first clarify the question of what is to be expected from such an investigation.

It must be explained *why* a transformation of the substance takes place. But it can *not* be postulated that this explanation contains a derivation of the essence of the mental states; however, this is an obvious limitation, because something that is not included in any description, cannot be derived either.

This restriction, however, does not represent an absolute barrier, because it is possible – as with the physical being – to approach the essence of mental states conceptually, even if it cannot be thought; – or say rather: cannot be *described,* because we actually *know* what qualia are.

Strictly speaking, qualia are in fact the only being, of which we know what it is, because our consciousness is a constant stream of qualia. Not only what we feel, also what we think and what we perceive is a quale. Even the most abstract intellectual activity is carried by an interest and guided by a motive, and both interest and motive are descendants of sensations from which they cannot be separated.

So how does the second substance emerge from the first? Why is the essence of being transformed?

The Reason for the Transformation

We can act on two assumptions:

(1) The transformation of the essence of being from the material object to the quale occurs through the unfolding of nature into layers of increasing complexity, whose governing laws have been described in the previous chapter.

(2) The part of this rise, which is conceptually accessible for us, is the part that occurs on the side of the accidents. Thus, there the argument must take place.

First a designation: To the essence of being in that area of reality in which there is no mind, I'll assign the term "matter". (Thus, this term "matter" does not only contain the accidents – as is usual in physics – but also the substance.)

The point which we must direct our attention to is the question:

Why is the essence of being transformed only concomitant with the development of neuronal networks of high complexity which bring forth mind, while up to that point it seems to be uniform throughout – at least to such an extent that only at this point we are compelled to introduce a second substance?

We manage with some success, to describe the phenomena we find in the material world. Where our knowledge is incomplete, as in the case of the origin of life, this gap can at least be filled by scientific hypotheses. Problems that we meet in the description of nature usually appear as *technical* difficulties, and we never encounter a phenomenon that seems to escape our understanding *on principle* – but that holds true only to the point where the unfolding nature generates neuronal networks of high complexity.

With their development, at the same time a phenomenon appears that eludes not only scientific description but indeed *any* kind of description: *sensation*.

Let us now begin with the train of thought which, at the end, will let us realize the necessity of the transformation of the first into the second substance.

What happens when, due to the unfolding of nature into entities of increasing complexity, the substance changes?

The answer follows from the definition of the substance: substance is a part of the essence of being, and it is that which remains excluded from descriptions.

Thus, if the substance is transformed, then we are confronted with a being whose essence is incomprehensible *in a new way*, a way that seems different from the one of the essence of beings, whose substance remains the same.

However, this may sound strange and vague: How can the transformation of something that does not occur in descriptions manifest itself? Does it make sense to speak of different kinds of the indescribable?

That, what happens when the essence of being changes, and how this change becomes apparent, can be illustrated with reference to that case where it actually occurs: the transition from matter to mind.

As stated above, we can neither think the substance of the material being – the space-time continuum – nor its essence, that is: matter defined as pattern of spatio-temporal changes conserved over time.

But our inability is a *specified* inability. We have pursued, what is beyond our thinking, to the limit and, in this way, conceptually captured the continuum as far as possible: through the concept *limit* (of a sequence or function), we have made operationally available what we cannot think.

But the methodological approaches to the incomprehensible, in which we were so successful that we could almost forget that, ultimately, it *is there*, fail completely, if we try to apply them on what is inexplicable in that kind of being which belongs to the realm of the mind.

Obviously, we do neither come any closer to the substance of mental entities – sensation – nor to its essence – the quale, the unity of feeling and meaning –, if we regard them as patterns of spatial and temporal changes.

However this statement has to be qualified: Of course, mental entities *are* patterns of spatio-temporal changes – all that exists is indeed such a pattern – but in this statement the term "are"– unlike in the case of matter – must not be understood in the sense of a definition of the essence. The essence of material things is determined by the fact that they are spatio-temporal patterns, but the essence of mental things is that they are qualia, and not that they are spatio-temporal patterns.

Therefore it can be asserted:

The occurrence of a second substance is revealed to us by the appearance of phenomena, whose essence is inexplicable to us in a new way, such that the hitherto reliable intellectual means of approaching the inexplicable fail.

However the "means of approach" are nothing other than the accidents that are connected with the substance; the substance manifests itself through its accidents, and it is revealed to us through their investigation.

With this, we have determined *one* direction of the relationship between the transformation of the substance and the change of accidents:

If a new substance occurs, then also the accidents change.

Here is a short version of the just performed conclusion:

The accidents of being are manifestations of its substance. As such, they enable us a conceptual approach to that which is actually unthinkable: to the substance.

Now, if a new being with a different substance appears, then the conceptual approaches must aim at another goal, and this means that the thought processes required for approaching the new substance must be different from the previous ones. In other words: the accidents have to change.

What about the reverse? Does a new substance appear if the accidents change?

It can be seen immediately that in general this is not the case. In the evolution of being, in any new emerging layer also new accidents appear, without the substance being transformed.

A transformation of the substance takes place only at the transition to the last, most complex layer of reality, to beings that have a neuronal network that is sufficiently complex to bring forth mind.

With this, we have arrived at the determination of the logical point, where the elucidation of the question of the transformation from matter into mind is to be found:

The answer to the question, why the essence of being changes – such that mind emerges from matter – must arise from the analysis of the difference between the kind of alterations, which occur in the accidents due to the evolution of complex neuronal networks, and the kind of alterations, which occur in the accidents due to the evolution of new layers of reality of lesser complexity.

So, wherein consists this difference?

Let us first consider some of the accidents that occur in the formation of new layers of being.

We start with an accident that occurs even at the simplest being: gravity. If mass is understood as that *from which* gravity emanates, then mass is part of the substance of all material objects.

In the Second Part has been shown that the gravity of an object follows directly from the law of the continuum, if an additional condition is imposed on the metric density of the surrounding space.

This does not mean, however, that any state of a real physical system of the type "material body with its gravitational field" can completely be derived from a state of the universe *before* the formation of particles that have mass and are surrounded by a gravitational field.

Only the much weaker claim applies, that the accident *gravity*, when viewed in an idealized form – i.e. without any interference from outside and in full equilibrium –, appears as ordered state of the longitudinal flow, and that all the effects that this flow-state exerts on other objects, can be traced back to the law from which it arises.

In this sense, then, the accident *gravitation,* which appears in the first and simplest layer of being, can be *reduced* to the underlying layer, that is: the continuum with its law.

On the side of the substance, this reducibility of the accident means that the concept "mass" can be eliminated without loss from the description of nature. It is not necessary to regard mass as a new substance, i.e. as a new inconceivable metaphysical entity.

Let us now consider an accident, which belongs to a much more complex layer of being: as (random) example, we choose the so-called *gluconeogenesis*, which is the formation of glucose in cells. This process – as most metabolic processes – is very complex and consists of a series of individual processes.

In our context, however, only the following is of interest: While it would be impossible to derive the gluconeogenesis from any states of the universe that existed before the evolution of cells, it can still be asserted that all steps, which in cells must be run through for the formation of glucose, can be described and understood as *biochemical processes*.

Thus, in this respect, the gluconeogenesis, which appears as accident of cells, does not differ from gravitation, the accident of all material objects: both can be understood through reduction to the respective underlying layer of reality. They appear as *functions* of this layer.

One last example: processes which take place in neuronal networks that are not capable of producing mind. (That such neuronal networks exist can be presupposed. One could for example model the behavior of the robots John and Suzy by neuronal networks, and these would then be networks of this kind.)

The accidents of such processes can be summarized under the term *information processing.*

Information processing consists of input-output relations. If these relations are not altered *internally* – through networking with other such relations and the feedback resulting thereof – but remain always in (nearly) identical form, they can be regarded as functions of the given architecture of the neuronal network and of outside conditions.

Then also the behavior of animals that have neuronal networks of this kind can be regarded as function of this architecture and of outside conditions.

Thus also in this example holds true:

The analysis of the accidents shows that they can be understood as functions of the underlying, simpler layers of being.

We have thus reached the following insight:

In all evolutionary transitions to new, more complex layers of being – up to and including neuronal networks that are not capable of producing mind – the accidents, which occur in the respective new layers, can be described as functions of accidents of simpler layers of being.

The last example has brought us already close to the realm of the mind. So let us now take the last step. Let us pose the question:

What is the difference between the accidents of mental states and the accidents of other beings?

Does from this difference also follow – as postulated for the explanation of the transformation of the substance – that with the transition from the Material to the Mental another kind of change of accidents occurs as with all the transitions that remain within the realm of the Material?

According to the above, the answer seems obvious:

Let us look at a neuronal network that brings forth mind. As necessary prerequisites that neuronal patterns which represent something – which we have identified as attractors of the neuronal dynamics – can turn into mental states, hitherto two conditions have been determined:

- The existence of functionally unbound areas, the dynamics of which is open for the structuring through such attractors.
- The networking of the attractors with one another.

Representations can relate to external conditions or to body states. But also the information content of a neuronal state itself can be represented. Such meta-representations are, for example, required if alternative courses of action are to be weighed against each other.

The crucial point for our question is this:

From Hebb's law follows that the mental activity acts back on the neuronal structure. Thus the mental activity changes its own neuronal encoding; it changes *itself*.

The postulate that the representation states are networked with each other is tantamount to the occurrence of feedback loops: state A influences state B, state B influences state C, which in turn acts back on state A etc.

On the one hand, such feedback loops enhance already existing patterns, on the other hand, they can also enable connections between patterns which were not connected before. With this, the information content of the neuronal patterns changes: it is increasingly determined by the *internal* relationships between the neuronal patterns, whereas the original functional dependency recedes into the background; representational states develop into intrinsic meanings.

This is true even for perceptions: even if they remain, as neuronal representations of real objects, bound to this objects as regards their information content – in the primary visual cortex, the neuronal image of an object, which is observed twice under identical conditions, will in both cases be almost identical – they are, as *mental states*, by no means limited to this representational function. Perception includes any kind of information processing that occurs in the respective cortical area in addition to the processing of the purely sensory information, and a halo of accompanying associations is also part of perceptions.

In the case of mental states, which are not directly linked to outer objects, there is no principle limitation at all for the changes which they are subjected to in the course of their further inner processing within feedback loops. In trains of thought, surprises occur: new conclusions result, new concept formations are required, fallacies must be corrected. In the area of fantasy, the change of existing and the creation of new intrinsic meanings is even the characteristic feature, and the relation to outer objects fades away or tears off completely.

From this follows the sought differentiator:

Intrinsic meaning, the accident of mental states, cannot be understood as function of accidents of simpler (atomic, molecular, biochemical, neuronal etc.) layers of being.

A mental state gains its meaning through its position in the network of mental states, i.e. in the network of meanings. Although perceptions and judgments remain bound to the real world, it is – due to the permanent change through feedback loops – impossible to assume any kind of functional dependence of the meaning of mental states.

Shortly before we had found that:

In all evolutionary transitions to new, more complex layers of being – up to and including neuronal networks that are not capable of producing mind – the accidents, which occur in the respective new layers, can be described as functions of accidents of simpler layers of being.

So this is the difference, which is the reason that the substance of being – and hence also its essence – changes only at the transition from matter to mind, whereas it remains unchanged at all other transitions from one layer to the next higher one.

But does this criterion also provide an *explanation* as to *why* the substance of being changes at the transition from matter to mind?

Yes. As follows:

Substance and accident are *inseparably bound* to each other.

If complex accidents can be described as functions of simpler accidents, then follows that, ultimately, they can also be reduced to the first and simplest accident. In this way, they are then not only linked with the first accident, but also with the first substance, and the unity of substance and accidents is maintained.

But if there are accidents, which cannot be traced back to the first accident, then their connection with the first accident is interrupted, and thus they are also separated from the first substance.

However since the being, which these accidents belong to, *must* have a substance, which its accidents are *inseparably* bound to, this substance can no longer be the first substance.

Therefore, with the emerging of these accidents, also a transformation of the substance must have taken place.

Proposition

As long as accidents of higher complexity can be described as functions of accidents of lesser complexity, the substance remains the same.

If this functional dependence disappears, then the substance changes. *For us* it appears then as a new, second substance.

In the area of matter, the first condition is met. Each material being can be understood as attractor of the dynamics of NOTHING, as pattern of space-time alterations. This is the essence of the material being.

In the realm of the mind, the second condition is met. Intrinsic meanings, the accidents of mental states, can in no way be understood as functions of accidents of the underlying layers of being.

Therefore, the essence of the mental being is different from the essence of the material being.

With this, the question is fully clarified, why the emergence of mind from matter also leads to a transformation of the essence of being.

Notes

1. The core of the argument, with which the transformation of the essence of being is substantiated, is the interruption of the connection between the mental accidents and the first accident. With this, also the connection between the mental accidents and the first substance tears, whereby the assumption of a second substance is forced.

Here, I have presupposed that "connection" is equal to "reducibility" (of the accident, not of the given state of the system).

It would also be possible, however, to replace the assumption of reducibility by the weaker assumption of *explicability,* Also explicability could serve as differentiator between the accidents of the qualia and accidents of material beings, because the intrinsic meaning of mental states can only be explained through mental relations and not through any material accidents.

Although the weaker assumption would be logically preferable, I have decided for reducibility as the distinguishing criterion, and here is why:

The connection between substance and accident is of a metaphysical kind: *inseparability.* Therefore, if complex accidents remain connected with the first substance, because between them and the first accident a connection is mediated by the accidents that lie in between, then, accordingly, this connection must be the closest connection possible between accidents of different hierarchical layers in a description of nature, and that is indeed reducibility. *For us,* reducibility means ontological identity: If B can be reduced to A, then B *is* ultimately A. Thus, if a complex accident can be reduced

to the first accident, then it *is* ultimately the first accident, and then it is also inseparably bound to the first substance.

2. The content of the just derived proposition can also be expressed in the following way:

What in the description of reality appears as reducibility of the attributes of complex objects to the attributes of simpler objects, manifests itself in the reality as identity of the substance.

What in the description appears as formal independence of attributes, means ontologically the change of the substance and the transformation of being.

3. The reason for the transformation of being can also be understood in a simple and intuitive manner:

If the functional dependencies of neuronal states from their material preconditions fade away and disappear ultimately, then this means that, here, an area of the universe decouples from the rest of the universe. Thus a new, separate universe emerges, a *universe of qualia.*

Meanings are the accidents of the entities of this universe. They must be connected with a substance, and their separation from the rest of the universe and its substance *space-time* suggests that to these accidents a new substance belongs.

Substance is that, which provides the accidents with the metaphysical quality *activity.* So one can ask: what is it, which provides the entities of the universe of the qualia with *activity*? Whereupon is the dynamics based in this universe?

The answer is *sensation.* Sensation is what drives the qualia. Therefore, sensation is the substance of mental states.

By contrast, meaning – that, which is subject of descriptions, i.e. the formal definition of mental states – is *passive.* Meaning is *information*, and information processing alone – as everything that is just an element of a description or of a model – does never occur *by itself.* It requires activity from outside.

But it need not be specifically emphasized that this universe of qualia is an *inner* universe, a universe *in the head.* The functional decoupling, which manifests itself in the free flight of thoughts and ideas, does not mean, of course, that the mind, as esoterics and members of various religions believe, can in fact liberate itself from its spatial and temporal boundedness. It is brought forth by the neuronal network, and thus it remains bound to material conditions and captured in space and time.

4. Since substance and accidents always belong together, it is unsatisfactory to assign different accidents to different layers of being, but still to assign to them the same substance.

However, there a conceptual problem emerges. More complex accidents can be reduced to simpler ones and finally to the simplest accident. But, in the literal sense, this cannot be said about the substances that belong to these more complex accidents, because the substance can not at all be captured conceptually, such that the notion of a "deduced substance" in its literal meaning would be nonsensical.

However, it is possible to *define* such a notion and, in this way, to discriminate it from its literal meaning.

So we define:

Deduced substance is a substance, which is connected with a deduced accident.

Thus each being, which is not a mental being, consists of a deduced substance and deduced accidents. Only the *origin of everything* consists *for us* of the first substance and the first accident.

Since we understand only the accidents of being, not only a part of the essence of being remains withdrawn from our concepts, but also a part of the change of the essence of being, which occurs during the rise of being to ever more complex forms. All we can do is to conclude from the side of the accidents to the substance and its changes.

It is important to keep in mind that also the substance *space-time* of the physical being cannot be thought, and that already for this reason alone it is impossible, to conceive its transformation into the second substance *sensation* of the mental being. As has just been demonstrated, however, it can be proven that this transformation must take place.

5.5. Criterion for the Occurrence of Sensations

The considerations of the previous section result in a criterion for the occurrence of qualia.

The existence of a quale implies that its meaning cannot be read from its material structure. This condition is satisfied if and only if neuronal states that represent something or produce input-output relations are networked with each other. Then feedback loops occur in which the information encoded in the neuronal states is determined to an increasing extent by the mutual relationships of the neuronal states, while the original dependence from outer conditions diminishes. That, which initially has been *representation,* turns into *intrinsic meaning.*

There is no way to determine what a neuronal pattern means. This is even true for perceptions: here, it can at most be determined what they represent, but not what they mean. Contrary to the high hopes of neuroscience, it will never be possible to eavesdrop on someone's mind – unless he voluntarily tells what he thinks and thus allows the identification of the individual neuronal encoding of these thoughts. But, with certainty, even that will be possible only for simple, standardized mental processes.

As already mentioned several times, the just described process of the emergence of intrinsic meaning – and thus also the formation of qualia – can take place only if there is a neuronal structure that is not functionally predefined.[4]

In humans, this structure is the cerebrum. Its functional independence becomes apparent through its plasticity: if areas fail which, over the course of the individual development, have taken certain tasks, then these tasks can be adopted by other regions.

But also other brain structures can meet the criterion of functional independence. The brain of octopods is structured very differently from our brain, but most likely it contains neuronal areas of this kind.

On the other hand, the neuronal structures of the diencephalon are not suited to network representations in such a way that they can detach themselves from their original function and develop into intrinsic meanings. Regardless of whether this function is part of a genetic program or imprinted by external conditions – the behavior remains schematical and always related to the triggering stimulus. Representations that are integrated into such functional sequences can change only within narrow limits.

[4] Such structures appear when a mutation causes the enlargement of an already existing neuronal structure, e.g. of an area of the diencephalon.

Therefore, in brains which – in addition to the evolutionarily even more ancient neuronal fields – contain only structures similar to those of our diencephalon, most likely the emergence of qualia is not possible.

Thus the necessary condition for the occurrence of qualia is:

Qualia occur in a neuronal network if the network contains functionally unbound structures, which permit the networking of neuronal representation states.

However this is only a necessary condition. Can also a criterion, i.e. a necessary and sufficient condition, be formulated?

I think yes. It reads:

Any animal that has a neuronal network which contains functionally unbound structures, experiences qualia.

It must be emphasized that this statement applies only to an animal and not to a robot. We could provide the robot Suzy with the luxury of an additional neuronal module, which we simply connect with the already existing neurons, without specifying its functionality. Nevertheless, it can be excluded that Suzy will have feelings.

In an animal, however, it can be presupposed that its species has already proven its viability, and this is a very strong condition. It includes a lot of technical requirements, of which we know only a few *exactly* and most *not at all*.

The claim that an animal with such a neuronal network experiences feelings is based on the following argument:

There must have been viable progenitors of this species, in the brain of which the neuronal structure similar to the cerebrum has not yet been present at all or only to a very small extent. Presumably, the organism we are looking at would thus be viable also without the "free" neuronal structure, and the neuronal network would contain all functions, which are necessary for the appropriate regulation of its behavior.

But if now this new, initially function-free structure is added, then *inevitably* meta-representations and networked representations will arise, because the information about the environment and the body *must* in some way arrive in the new structure and there be processed further – but again only if the

neuronal extension is a continuation of a neuronal tissue, which has already proven its ability to function correctly and is not just a bunch of neurons.

This condition, in which again many technical prerequisites are summarized, is certainly met in an animal.

With this, however, it is already proven that this animal has feelings.

A simple additional Argument

The existence of feelings presupposes that there is *somebody who* feels

Consider the sensation *pain*: the sensory information can only turn into the sensation *pain* if there is a *subject* who takes note of the sensory information.

It is not necessary to determine more precisely what is meant by the terms "somebody" or "subject". It suffices to realize that in a neuronal network, in which only automated processes that are either genetically programmed or imprinted by environmental conditions, there is no room for this postulated subject.

The assumption of a subject presupposes that stimulus and behavior are not in any case in a fixed connection with each other, like in the case of a reflex or of a learned program. There must also be cases, where the sensory information does not lead directly to the stereotype behavior but *further processing* is applied. Therefore, the neuronal network must be capable of *meta-representations.*

But obviously, a single meta-representation alone does not suffice to substantiate the assumption of a *subject.* For that, it is required that the meta-representations are stored as memories and networked with each other.

Only then is it justified to assume that there exists *somebody* who feels.

So this simple argument leads again to the necessary condition, which has just been derived: For the appearance of feelings, the existence of a functionally unbound neuronal structure is required, which enables the networking of the neuronal representation states.

It must be noted, however, that this is a *structural* argument, and therefore it is, like all structural arguments, inappropriate to substantiate the *metaphysical fact* of the transformation of neuronal states into qualia. It can only serve for the determination of necessary conditions

5.6. Who or What has Sensations?

The above formulated criterion for the occurrence of sensations permits drawing a boundary between *machine* and *sentient being.*

The actual classification, however, requires precise knowledge of the structures of the respective neuronal network and their performances. As mentioned, the intelligence of birds has been underestimated for a long time, because it is not located in the same neuronal area as in the case of mammals. By now, however, it is well known that some birds are highly intelligent. Birds undoubtedly have sensations.

What about fish? The discovery of structures similar to our diencephalon does *not* represent an indication of feelings. But the question arises whether the complex behavior and learning ability of some fish species suggest the existence of neuronal structures that meet the criterion of *functional independence*. Perhaps there are fish species in which this is the case.

Some questions, however, can be decided unequivocally using the criterion:

Do bees perceive colors? No. Although different colors are differently represented in their neuronal networks, the representation remains pure information. A transformation into perception does not take place.

Do Crustaceans feel pain? No. They are not sentient. The relationship between Robert and his shrimp will remain one-sided.

On this side of the border, which is drawn by the criterion, there are no sensations. But what is on the other side? What if the initial function-free cerebral-like structure is *very small*?

Are then feelings somehow "paler"? Is *red* less reddish? Is *pain* less painful?

On the one hand, it must be remembered that the indescribability of feelings is transferred to their gradations. It is not possible to describe gradations of something that is not describable.

On the other hand, with feelings we have this special relationship that, though we cannot capture them by descriptions, we still know exactly what they are because they are *directly* – as they themselves – given to us.

Therefore I think that the idea of "paler" or "weaker" emotions and a "dull" or "vague" consciousness is a suitable approximation to the nature of those qualia, which animals experience, in which the functionally unbound neuronal structures required for the emergence of qualia are less pronounced.

An important aspect of the conclusions of the previous section is that – also in neuronal networks which are capable of forming qualia – qualia are not present from the beginning. The transformation of matter into mind, of representations to intrinsic meanings, of a neuronal pattern into a quale is a development process.

From this follows, for example, that infants do not *perceive* colors after their birth; even later, when they are already able to *distinguish* colors, initially they cannot have any color *sensations* – just because the information did not have time to develop into a quale.

5.7. Attempt of a conceptual Approach to the Transformation of the Substance

The fact that the first substance *space-time* is transformed into the second substance *sensation* is at first disconcerting – even if the necessity of this transformation can be realized. This irritation can be reduced in various ways.

First approach

The first step is certainly the realization that here the concept *space-time* is not identical with the mathematical-physical concept "space-time". This concept is only *defined,* which means: it consists only of the accidents by which it is defined, whereas the substance *space-time* denotes that, which *has* these attributes, that is: that what the space-time "is" *without* these attributes, where the quotes indicate that the ontological status of the *space-time* cannot be "existence".

Therefore, that which is transformed is not the physical "space-time" – this idea would be outright absurd – but the unthinkable first substance *space-time.*

Second approach

If we try to think *space-time* as substance, it turns into NOTHING. However what we know about it is that it has in it the metaphysical quality *activity.* [5]

Material objects are space-time patterns. Therefore it is possible to define their substance by the concept-pair [space-time, activity]:

Material substance := [space-time, activity]

Mental states are also space-time patterns, however it must be kept in mind, that in this statement the notion "are" is not meant as definition, as is the case in the statement "material objects are space-time patterns".

Therefore, the second substance can be defined by the concept pair [space-time, sensation]:

Mental substance := [space-time, sensation]

Here, the transformation of the essence of being appears as transformation of the metaphysical quality: *activity* turns into *sensation.* The first substance is that, which makes the accidents of matter *active,* the second substance is that, which makes the accidents of the mind *active.*

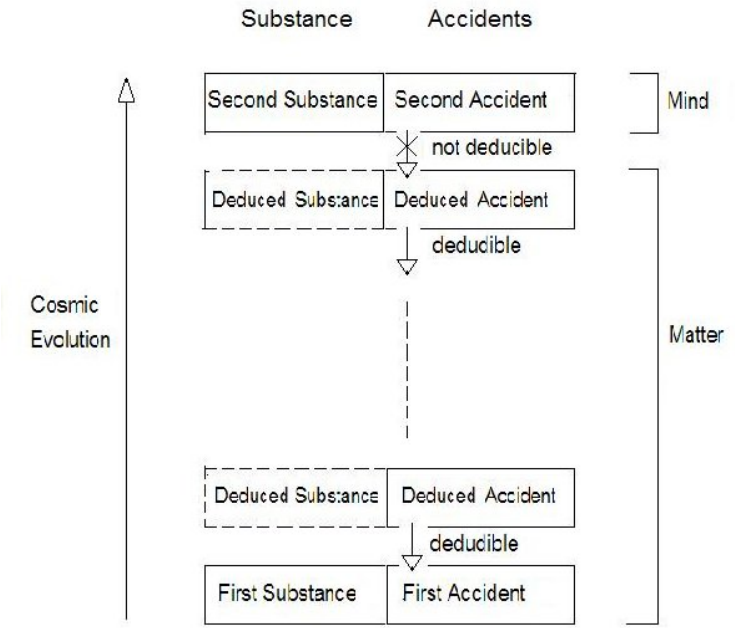
It is readily apparent that *sensation* does exactly that: meaning without *sensation* is indifferent. *Sensation* is the basis of any assessment and any motivation. It is what drives us.

Third approach

In order to understand the cosmic evolution within the scheme of substance and accidents, it is necessary to determine the relationship between derivability and non-derivability on the one hand and essence-equality and essence-transformation on the other hand.

First, an outline for an overview:

[5] I point out once again that *activity* is not an accident. If we divide the *origin of everything* into substance and accident, in order to make it accessible to our thinking, and denominate it as activity of NOTHING, then now *activity* appears indeed as accident, but from this follows that the difference between the accident "activity" and the metaphysical quality *activity* is shifted into the concept of the substance. Then, substance is what makes the accidents *active.*



In this outline, only the right side is accessible to us, that is: the side of the accidents. The left side is about what in all descriptions is missing and cannot be thought: about the substance.

Therefore, *for us* the relationship between any layer of reality and the respective underlying layer is given exclusively through the alternative *derivable* or *not derivable,* that is: through the relationship of the accidents.

Derivability, however, is such a close relationship between that which is derived and that from which it is derived, that *for us* it seems as if the derivative and the original were *actually* the same. But this identity applies only to descriptions. In the reality, any accident is inseparably bound to an associated substance, and in the area of the substance there is no derivability. The notion *deduced substance* can only be understood as defined in 5.4 under Note 4: as substance that is associated with a deduced accident.

Being is always accident *and* substance. If now the concept of derivability is not admissible in the area of the substance, then, at the same time, the idea is canceled that being, whose accidents are derivable,

is a derivable being. The identity between that which is derived and that from which it is derived cannot be transferred to reality. It proofs to be a deception, which only *for us* exists, and to which we succumb when we equate reality and description in the same way as it is done in present science.

Reality and description differ from each other, and, in the same way, also evolution and derivation are different. So it would be quite appropriate to assume that the essence of being is changing in each evolutionary step to a more complex layer of being.

However it remains true, that *for us* the substance does not change as long as the accidents are derivable, and that it appears transformed only if the accidents are not derivable. Therefore it is justified to divide reality into a material and a mental area.

Fourth approach

Finally, to some extent it is also possible to reconstruct the concept of the mental substance directly. As follows:

A concept of the substance of a mental state is needed, which does not relate to the physical or to any other layer *below* the mental layer, but which proves to belong to the mental state itself and which appears appropriate to provide the accidents of the mental states – the intrinsic meanings – with *activity*.

Of course we already know which concept is suitable: *sensation.* But the question is how far the substance *sensation* can be conceptualized, in other words: how far could we understand sensation – so to speak – *from outside,* if we would not know it *from inside,* from our own experience.

So let us then imagine, we knew everything about neuronal networks, but we did not know what sensation is, and our task would be to figure out what the substance of mental states is. How far would we come?

We have to start on the side of the accidents. What do we find? Intrinsic meanings. They follow one another in time and are networked together in a virtual space. (I avoid the expression state space because I think that the ever-changing dynamics of the interconnected meanings cannot be transferred to a state space. There is no lawful relationship between the physico-chemical and neuronal parameters on the one side and the permanently altered meanings of the patterns on the other side, because in this scenario local and global laws complement each other in a not-formalizable way. More on that later.)

The question is: what provides the meanings with *activity*? The answer must be a concept that belongs to the same layer as meaning, a concept which forms a pair with the concept meaning in the same way as the concept mass forms a pair with the concept gravity.

In considering the question: "in which systems can occur feelings?" we have stated that there must be *someone who* feels. The same applies to meanings: speaking of meanings only makes sense if there is someone *to whom* they mean something.

With this, on the side of the accidents we have the following preconditions:

The accidents of mental states are meanings. What they mean, is meant *for a subject.* To this subject, they must be given directly and as a whole, not in the form of sets of values of variables.

How could a substance-concept be constituted, which is appropriate to form a pair with this accidence-concept – an inseparable unity like mass and gravitation?

What we already know is that in any neuronal network, which belongs to a viable living creature, there is a system that *assesses* neuronal states. From this system, the regulation of the dynamics of the global neuronal states must have emerged, before they turned into mental states and regulated their dynamics themselves.

For the subject, the assessments must be given as directly as the meanings, that is: again not in the form of sets of values of variables, but in the form of *qualities.*

What is a *quality for a subject*?

If we would not already know the sought concept, then we would define it precisely as *such:* as "quality for a subject".

Quality for a subject is appropriate for forming an inseparable unity with the accident *meaning for a subject.*

Quality for a subject is the substance of the mental being.

I think with this the maximum possible conceptual approach to the ultimately inconceivable substance *sensation* is realized.

5.8. Philosophical Zombies

"Philosophical zombies" are hypothetical beings devised for the purpose of illustrating the question of the relationship between neuronal states and qualia and thus also the problem of distinguishing between machines and sentient beings.

The physiology of philosophical zombies resembles that of humans. Zombies have therefore the same brains as humans. But they lack sensation. They do not experience qualia.

Is it possible that such zombies exist?

From our point of view can be seen immediately that their existence is impossible:

The condition that a zombie brain is identical in physiological respect to that of a human includes the identity of structures *and* processes. This means that in the zombie brain the neuronal states are networked with each other and also being changed by interaction in the same manner as the states of a human brain.

From this follows, however, that the process of transformation from a *physical being* into a *quale* must have taken place.

Therefore the neuronal states – the spatio-temporal excitation patterns – of the zombie brain are also qualia.

If it were assumed the zombie had no sensation and the neuronal patterns of the zombie brain were *nothing but* information processing, then this would be as absurd as the assumption that the Earth could exist without gravity. It would mean to remove from being its substance, i.e. that which provides the accidents with the metaphysical quality *activity.* But substance and accidents are always inseparable, they are always *one.* In the case of the earth, this seems obvious to us; in the case of the qualia, however, such absurd conceptual constructions as "neuro-zombies" seem possible – but only as long as one has no concept of what qualia are.

If a neuronal excitation pattern, which has become a quale, could be *removed from the network of meanings,* which it owes its existence to, then it would actually be *nothing but* a neuronal pattern. *As mental being,* it would then be destroyed, it would have turned back into a physical being.

But it is clear that a *global* neuronal excitation pattern cannot be separated from its neuronal environment without also changing it physically. Perhaps the structure could be conserved for a short time, but not the dynamics. It would instantly – starting at the moment of separation – change and, after a short time, break down.

5.9. Artificial Intelligence

Understanding

Can a machine that converts character strings into other character strings according to certain rules *understand* something?

This question can be answered with amazing clarity through a thought experiment by John Searle that bears the name "Chinese Room". The following abbreviated version was formulated by Searle in 1999:[6]

"Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test[7] for understanding Chinese but he does not understand a word of Chinese."

If one accepts that this procedure is possible – and one is forced to do so, if one considers *artificial understanding* possible – then this is a valid argument against the Turing test. Indeed, the Chinese Room *is* a version of the Turing Test: Although here the unknown communication partner is a human, he still does not understand Chinese. However, as he passes the test, the ability to understand Chinese must be attributed to him.

---

[6] John Searle, *The Chinese Room*, in R.A. Wilson and F. Keil (eds.), The MIT Encyclopedia of the Cognitive Sciences, Cambridge, MA: MIT Press, 1999.

[7] In this test, which is named after its inventor Alan Turing, people must realize by written communication with an invisible partner whether that partner is a computer or a human. If a decision is not possible, to the partner human abilities will be assigned.

Searle then adds: "The point of the argument is this: if the man in the room does not understand Chinese on the basis of implementing the appropriate program for understanding Chinese then neither does any other digital computer solely on that basis because no computer, qua computer, has anything the man does not have."

In this form, however, the argument is not correct: If the Chinese Room is regarded as computer analogy, then the man represents obviously not the whole computer, but takes over only the tasks of data transport and data transformation. Even the toughest defenders of the existence of artificial intelligence, however, would hardly claim that processor and motherboard alone could understand something.

Therefore, it was argued against Searle[8], that not *the man,* but *the system* understands Chinese.

Searle countered with the argument that the man could learn the data and the program, and then he would *be* the system.

I break the discussion of the thought experiment at this point, because I am convinced that the latter variant provides the proof that computers do not understand what they are doing:

The actions of the man correspond exactly to those of a computer – he transforms character strings into other character strings according to predefined rules [9] – and he does *not* understand, what the strings mean.

Any doubts that the thought experiment would be feasible at all, are not directed against the argument, but in any case only against the assumption that understanding can be simulated. The strength of the argument lies precisely in the fact that the maximum, which is achievable in a simulation, is presupposed: it is assumed that the answers, which the man writes down without understanding them, cannot be discriminated from the answers of a man who understands what he writes.

This means: Even if artificial intelligence had reached its goal, to simulate communication in such a way that it could not be discriminated from human communication, the simulation would lack understanding.

---

[8] Among others, by Georges Rey: *What's Really Going on in Searle's "Chinese Room",* Philosophical Studies 50, 169–185. 1986.

[9] In order to take into account also the possibility of a learning program one can simply add that the man transforms also the strings of the rules in his "book of instructions", depending on the respective input.

Another strength of the thought experiment is that it provides a simple and intuitive discrimination between understanding and not-understanding: you can talk with the man in English *and* in Chinese. He understands English, he does not understand Chinese. Understanding is that by which the two activities differ from one another. Each of us knows this difference. We *understand* it.

Thus Searles argument proves that even a perfectly successful computer simulation of human communication lacks understanding of *what* is simulated.

The argument, however, provides no answer to the question why this difference exists between humans and computers. And particularly because of the clarity, with which the lack of understanding is displayed in Searle's thought experiment, the explanation of this matter appears urgent and long overdue.

Why Computers cannot bring forth Mind; the *formal* Reason

In chapters 3, 4 and 5, mind was defined in a scientific manner, its existence justified by scientific methods and the transformation of the essence of the physical being into a quale explained by metaphysical arguments.

So the answer to why computers do not understand what they are doing must be contained in the conclusions of this chapters. This will now be demonstrated.

First we clarify a question that arises immediately: *When humans communicate verbally, are they then not also machines transforming strings according to certain rules*?

The answer is *no.* In Chapter *3. Free Will* was shown that mental activity does not follow fixed rules or rules that vary according to fixed meta-rules.

(In Chapter *4. The modified Picture of Reality*, I substantiated that this claim does not contradict the assumption that it is determined everywhere and anytime by the fundamental law, how the future emerges from the present.)

The following thought experiment is meant to serve for highlighting precisely *that* difference between humans and computers, from which follows why computer simulations of mental performances do not contain understanding.

Let M be a human neuronal network. The combination of all variable values of M at a given time point is a *state* of M.

We look at a sequence of states of M in a time interval Δt, which begins at $t = 0$ seconds and ends at, say, $t = 1000$ seconds. We call this sequence D.

The duration of the time steps between two consecutive states can be chosen at will. We assume their length as 10^{-24} seconds. Thus the sequence D contains 10^{27} states of M. (Since this is a thought experiment, however, the duration of the time steps can be further reduced.)

We specify that the variable values are represented with a precision of 1024 bits. (Also this accuracy can of course be assumed as desired.) [10]

As a prerequisite for the following argument, the conclusions are needed, which were developed in the chapter on free will. Here's a short summary:

The neuronal network is made of several superimposed layers of increasing complexity, each of which consists of entities that interact with each other. Therefore, the natural laws are not sufficient for the description of the dynamics of the network. Further laws – so-called laws of structure – must be factored in.

The laws of the hierarchically highest layer, that is: the mental laws, are dominant, which means that the dynamics of the mental states determines the neuronal and molecular dynamics to a greater extent than the reverse is the case (In the 4th chapter, this fact was denoted as *causality top-down.*)

The laws of the mind, however, in contrast to the laws of the other layers, are not fixed. The mental activity acts back on itself: it changes its own neuronal encoding and, with this, also the network of meanings, and from this follows at the same time a change of the transition rules of the sequences of mental states.

Now we introduce the following Turing machine into our considerations:

---

[10] The reason for the introduction of D is that D can directly be compared with a computer or a Turing machine.

(I will not give here a definition of a Turing machine. To understand the following, it is sufficient to know that a Turing machine can calculate everything which any arbitrary computer can calculate.)

Let T be the Turing machine that *exactly* reproduces the state sequence D within the time interval $\Delta t/2$ which starts at t = 0 seconds and ends at t = 500 seconds.[11] (The existence of T can be presupposed.)

Thus, within the interval $\Delta t/2$, the states of T are identical with the states of M (according to the chosen approximation).

The dynamics of T is computable. Therefore a function – say f_T – existsexists which permits calculating the whole sequence of states of T within the interval Δt from the initial state of T – this is the initial sequence of the characters 0 and 1 on the tape of T, which represents the state of M at the beginning of the interval Δt.[12]

Our first question is: *What happens **after** the end of the interval $\Delta t/2$?*

The answer follows from the difference between M and T:

According to our presuppositions, the network M produces out of itself – in feedback loops – new rules of the transitions from one state to the subsequent state. Therefore, what happens *after the end* of $\Delta t/2$ is not contained in the rules applicable *until then*.

In contrast, the rules of T are fixed. The Turing machine T is bound to the previously determined rules.

From this follows, that after the end of $\Delta t/2$ – if we let the Touring machine T continue to run – the states of T will no longer correspond to the states of M that are contained in D. Rather from this moment on an increasing difference must occur between the states of M and T.

Let us now again turn to the problem of the simulation of understanding. It has now assumed a clearer form:

---

[11] Each state of M which is simulated by T is a sequence of the characters 0 and 1 on the tape of T. I shall call such a character sequence – if it represents one of the states of D – a *state* of T, and that, what usually is called "state" of the Truing machine, I shall call *internal state.*

[12] Here, the question could arise, why not – instead of the function f_T – digitalized versions of the natural laws are used for calculating the state sequence of T. However that is impossible, because the non-linear feedback prevents calculating *exactly* any later states of M through initial conditions and such laws. (I remind you of the comparison with a great number of gravitating bodies: there is no exact procedure for calculating the future from law and initial conditions.)

The system M brings forth mind, M understands what happens. However, the sequence of states of the system T is – in any desired approximation – identical with the sequence of states of M in the time period Δt/2. Thus T is a perfect simulation of M, which can reproduce the states of M – e.g. also the states that control the speech output – during this period.

Does this mean that also T understands something in this time interval?

At first, the following must be cleared up:

What does the identity of the states of the two systems M and T actually mean? Is T just a digitalized version of M? Are M and T – in this sense – identical except for the different physical realization?

If this were the case, then to the Turing machine would have to be assigned understanding, which, however, we have excluded previously (in the discussion of the Chinese Room).

Moreover, this assumption would contradict the fact that, after the end of Δt/2, the states of M and T differ increasingly. If T were just a digitalized version of M, then M and T would continue to correspond with each other.

Therefore, though the states of M and T are identical within Δt/2, T is still not just a digitalized version of M.

However, "state" is defined as the set of *all* variable values of the system at a given point in time. From this follows that the structural difference between M and T must be attributed to the *transition rules,* which apply to the state sequences of the two systems.

What do we actually know about *this* difference between M and T?

In M, there are several layers of interacting entities. All layers have their own dynamics, and between all layers there is a complex interdependency. In different situations, the dynamics of a certain layer can take the leadership: e.g. in reflex actions the dynamics of the neuronal layer, in the case of defects of the transmitter system the dynamics of the chemical layer. Of interest to us, however, is the dynamics of the attractors of the network, which we have identified as mental states.

By contrast, in T the sequence of states is generated step by step, one character after the other, by the operation [read, write, move tape, change internal state].

With this, we have arrived at the precise formulation of the question of whether computers can simulate *understanding.* It reads as follows:

Can the dynamics of a neuronal network, which is based on the interaction of the attractors of the network, be simulated by transformations of values of variables which are performed step-by step and according to given rules?

As expected, the answer is *no.* The reason is that the description of the formation and alteration of order is performed through *global variables,* and that this description, as was shown in chapter 4, cannot be reduced to the description through *local variables.* The same, however, is also true vice versa. The relationship of the two kinds of description is *not formalizable*.

As mentioned, the mental activity acts back on its own neuronal encoding, which in turn affects the mental activity. This means; the dynamics of the global variables changes the dynamics of the local variables and vice versa.

If the Turing machine T is supposed to produce a complete simulation of the human neuronal network M – in other words, if M and T should indeed be isomorphic –, then T must reconstruct both the global and the local dynamics. This, however, is impossible, because, though both are interconnected, this interconnection can still not be expressed by an algorithm. But if they are reconstructed separately, each on its own, then the reconstruction fails, because the effects of the other dynamics are not factored in.

Thus, the interaction of attractors cannot be simulated through a step-by-step calculation of variable values, as it is the case in Turing machines. In order to generate the according state dynamics, the attractors – the mental states – must be present *as such.* In other words, only the neuronal activity patterns *as a whole* can produce the dynamics of the neuronal network. Their existence is a necessary condition of this dynamics.

But in a Turing machine or in a computer, they *have* no existence. And this holds true regardless of the possibility of parallel processing or the assumption of multi-layered hierarchical program structures.

Regardless of how complex (future) computer simulations of mental performances might be, how many levels of meta-representations they may contain or how many processors work in parallel – there remains the unalterable fact that all changes take place step by step and separated from each other and can therefore never form a *whole autonomous entity.*

And this means that the dynamics of mental processes – i.e. the dynamics of the attractors of the network – cannot be reconstructed by a computer.

I summarize: The attractors do not exist in a computer. Their simulation is not possible, because the dynamics of their interactions represents an *autonomous element* of reality and its description and cannot be reduced to the dynamics of local entities. Therefore, the interdependence of the dynamics of the local entities (the neurons), and the global entities (the attractors, that is: the mental states) cannot be mapped to a sequence of transformations of variables.

Ultimately, this argument means that a mental state is *an indivisible being.* Mental states are unities of *meaning and sensation.* As such, they are indivisible and can only interact with each other *as a whole,* and this precondition is transferred to their physical form, to the neuronal excitation pattern, the attractor, which is the necessary condition for the existence of the mental state.

With this, it is also cleared up, why M and T are not isomorphic. The entities, by which the dynamics of M is *actually* determined, do not at all exist in T, neither as they themselves nor in a simulated form. Thus also the transition rules of M cannot exist in T.

Therefore, even if the transition rules of T permit a complete and arbitrarily precise simulation of the states of M in a given time period, it is still true that these rules do not contain the *causal connections,* that is: those relationships which the development of the actually existing system M *in fact* depends upon, like e.g. *reasons*.

From this follows that the dynamics of a system must not be equated with a function, by which a state sequence of the system in a given period can be calculated. Between the two there is an inescapable ontological difference.

So this is the reason why, on the one hand, in T an algorithm can be implemented that generates a sequence of states which correspond to the states of M within a given time period, and why, on the other hand, this algorithm does not apply beyond this time period: The structure of the state space of the neuronal network M is determined by the mental states and their relationships, and this structure cannot be reproduced by an algorithm, because the relationships of mental states are not formalizable at all.

This statement is also the first step of the substantiation of the metaphysical assertion that T does not understand what happens.

The second step is the following statement, which has been derived in sections 5.3 and 5.4 of this chapter:

The transformation of physical states – neuronal patterns, which are attractors of the global neuronal dynamics and represent something – into qualia can only occur, if these patterns themselves are again networked with each other to such a high degree that their information content is determined by this internal networking and cannot be regarded as function of other – inner or outer – circumstances.

However, if this ever-changing network of attractors does not exist in a computer, because its dynamics cannot be simulated by transformation of variables, then from this follows that a computer cannot produce mind. Its states are not qualia. Its dynamics remains pure information processing and will never turn into mind.

Understanding, however, is a mental process. Therefore, computers cannot understand.

Still, it remains open how far it is possible to reconstruct mental performances by (mindless) simulations. The realm of possibility, however, seems to be, in almost all cases, much smaller than the extent of the hopes – only rarely the initial optimism proves to be true beyond the initial stages of implementation. As soon as leaving limited formalizable sub-worlds, the performance of the simulations usually decreases to an unacceptable level.

At last the question remains, whether at all there can be mind, which has not evolved naturally but is created by us. Thus the question is not about the possibility of artificial *intelligence* but about artificial *mind.*

I see no reason *in principle* to exclude the existence of a mind created by us.[13] At present, however, it is unclear how a system that brings forth mind can be realized other than biologically.

In addition, it can hardly be estimated how much evolutionary know-how – which would have to be technically implemented in the production of artificial mind – has entered into the chemistry and structure of a human neuronal network and the body that belongs to it.

It is very likely, however, that it is much more than currently assumed by those who are considering the optimization of humans.

[13] A fundamental restriction, however, will follow in the next section *5.10. The metaphysical Difference between Reality and Simulation*.

Note:

Ultimately, all formal and technical arguments, why the dynamics of natural systems cannot be simulated exactly, are based upon the fact that nature is not an algorithmic system: it produces the future from the present not through the application of an algorithm, but alone through the differential conditions given at any point.

As discussed in chapter four, the way in which nature generates the future cannot be transferred to a description or to a model. For this reason, an exact and complete description or modeling of reality is impossible.

This represents an insuperable limit for simulations.

In those areas of reality, where nature approaches algorithmic describability through the formation of order and the according laws of structure, this limit is irrelevant. A known example is atomic and molecular processes. Here, in some cases a description is possible which does not differ measurably from reality.

But there are also areas, which elude describability.

And in the case of mind, exactly that what should have been simulated will be lost.

Why Computers cannot bring forth Mind; the *metaphysical* Reason

The metaphysical explanation, why computers cannot produce qualia, is much shorter than the formal explanation:

States of computers are strings of the characters 0 and 1. Thus, a computer state is a *number,* and the dynamics of a computer is a sequence of transformations of numbers.

A necessary condition for the transformation of a physical state into a quale is that its accidents cannot be understood as functions of simpler accidents or as functions of outer or inner (non-mental) circumstances. Only then the second substance *sensation* can emerge.

This statement must be adapted to the conditions in the computer.

The accidents represent what constitutes the dynamics of the system. In the case of a computer, this is the program or the transformation rules. If the program is capable of learning, then the transformation rules change. In any case, however, there are no "simpler" rules but only "earlier" ones.

With this, the necessary condition for the occurrence of qualia in a computer can be formulated:

The transformation rules must not be functions of earlier transformation rules, earlier states and the current input.

But any state of the computer is a function of the input, the previous state and the transformation rules applied to it, and the same is true for any set of transformation rules. There is a chain of functional dependencies, which is never interrupted and leads back to the initial state and the initial program.

And from this follows that the transformation into a quale never happens.

There is no *mind in the machine.*

5.10. The metaphysical Difference between Reality and Simulation

All previous arguments about the limitations of our simulations were at least partially formal.

However there is also a purely metaphysical argument, which is based on the difference between reality and simulation and from which a fundamental barrier for simulations can be derived.

It reads as follows:

Reality consists of *substance and accidents.* Substance is that, from which the effects of a being emanate, and at the same time that, which provides the accidents with the metaphysical quality *activity.*

From an actually existing system, however, *only the accidents* can be transferred to a description or a model of the system. For this reason, the descriptions and models lack this activity; *by themselves*, or *out of themselves*, they are *passive*. They consist only of accidents.

A simulation, however, is nothing other than a model of reality, which is *activated* in some way. Thus, also a simulation is not active *by itself* but needs activation from outside.

For illustration, let us consider the system Sun, Earth and Moon:

A model can be produced, a so-called *tellurion.* Through suitable mechanical devices (gears, shafts, chains etc.), the movement of earth and moon can approximately be imitated – however only if the model is provided with *activity* from outside, e.g. by a crank or an electric motor. *By itself*, it is *passive*.

The movements of the bodies can also be calculated in a computer and then projected onto a 3D-screen. But also the model, which is implemented in the computer in the form of initial conditions and equations and the according numerical approximations of solutions, is *passive.* Only if the power is turned on, the simulation begins to run.

Thus we define:

A simulation is the reconstruction of the dynamics of a system A in another System B, which differs from A in the following way:

The dynamics of A results from the substance and the accidents of the objects in A, which means: the objects in A are active by themselves and according to their essence, and, in this way, produce the dynamics of A.

The dynamics of B, however, does not accrue from the substance and the accidents of the objects in B, but from the construction of B, which is designed and implemented by us.

Metaphysically spoken: the objects of B serve only as material basis of constructed accidents which, accordingly, have no associated substance. Since these accidents lack the substance, they are not active out of themselves. Thus the system B needs to be provided with activity from outside.

This definition can again be illustrated using the example of the tellurion:

In the real system sun, earth and moon, the objects move *by themselves* and according to their essence, that is: through their gravitation.

In the simulation, however, the activity according to the essence of the model bodies is indeed present – as gravity which manifests itself through the weight of the bodies –, but it contributes nothing to the dynamics of the simulation. For that, the design of the tellurion – just the gears, shafts, chains etc. – is responsible, which *we* have created. And the simulation is not moving *by itself,* but only if it is driven. Only then, the model bodies imitate the movements of the celestial bodies.

From the just defined difference between reality and simulation follows the important

Proposition

In the simulation of a being, the essence of this being is lost.

A tellurion imitates the movements of the celestial bodies, but *not through gravitation.* The metaphysical unity of the substance *mass* and the associated accident *gravitation* – the essence of the beings in the actually existing system – disappears in the simulation. Here, the movement is the result of constructed, *mass-less* accidents.

A computer program can imitate mental performances, but *not through mind.* The metaphysical unity of the substance *sensation* and the accident *meaning* – the *quale* in the actually existing neuronal network – disappears in the case of the computer. The imitation takes place through constructed, *sensation-less* accidents.

So if we aim at creating a system that brings forth mind, this system cannot be a *simulation* of mind.

To say it even more clearly:

There is no simulation of mind, because in the simulation the essence of being is lost. From the quale, whose essence is the unity of sensation and meaning, in the simulation remains only the meaning, i.e. the information content.

Through the layer of the constructed accidents, the metaphysical unity of substance and accidents – which is not contained in any description – is abolished, and upon the dynamics of the reality, which *as such* cannot be captured completely, an algorithm is imposed, a layer of reality that consists *only of accidents*. Here, there is no room for the substance *sensation.*

In the case of gravity, the fact of the disappearance of the substance seems self-evident:

Regardless of how perfect the connection between the objects may be constructed, which are supposed to imitate the gravitational dynamics – what drives them *is not* the (deduced) substance *mass.* The cause of their dynamics is not gravitation, and it can also never turn into gravitation. The simulation does not contain gravitation.

In the case of mind, the same applies:

Regardless of how precise the simulation is – what drives the entities of the simulation *is not* the substance *sensation,* and it can also never turn into sensation. The simulation does not contain sensation. It does not contain *mind.*

Corollary

The simulation of mind is impossible.

In the evolutionary development of nature, the metaphysical quality *activity* is transferred from any layer of being to the next-higher, more complex one. At the last step, mind emerges through the transformation of the first substance into the second substance.

But this transformation of the substance occurs only *for us. In itself,* substance and accidents are *inseparable* in any layer of being, and they evolve *as unity.*

In a simulation, the accidents, by which the dynamics is formed, are substance-less. But in order to bring forth mind, it would be necessary that there is a substance that belongs to this accidents and could then turn into sensation. This substance, however, cannot simply "emerge". Thus, a necessary condition for the transformation is that the substance to be transformed is already there.

But the substance which *in fact* is present there – the one of the elements of the simulation – is not the substance that belongs to the accidents which form the dynamics of the simulation, and therefore it cannot turn into *sensation* and be transformed together with its accidents into a *quale.*

Thus, the simulation cannot bring forth sensations, in other words: *it does not feel anything.*

Sensation is not an accident, i.e. not an attribute of the system or of its elements. It cannot be *constructed* but must result from the activity of being according to its essence.

This means: For the creation of mind, it is necessary to avoid the layer of substance-less, constructed accidents. The objects, whose dynamics is supposed to lead to the emergence of mind, must act according to their essence, that is: through their own accidents and through the activity that is provided by their associated substance.

It is therefore impossible, to generate mind on the basis of silicon or other materials with similar attributes. Mind can be created by us only in the form of an "artificial evolution". Thus the creation of mind is shifted into a far distant, perhaps unreachable future.

Addendum

The hope that someday in the future computers might be able to feel something is based on the confusion of an *actually existing system* with the *time-dependent evolution of its state.*[14]

This confusion becomes apparent already in the term itself: in everyday language, the term "state" relates to something that exists. The physical "state" of a system, however, is just a string of numbers – the values of the system variables –, and a computer simulation is the successive transformation of this numbers. It is irrelevant how the numbers are encoded and which storage medium is used. And it is equally irrelevant how the transformation is performed and in what kind of machine it takes place. In

[14] As a reminder: The *state* of a physical system is the combination of all variable values of the system at a given point in time.

any case, it is still a *sequence of states,* which means: a list of numbers which change with time, and it will never be identical with the *actually existing system.*

The definition of a simulation as dynamics of constructed accidents without substance, which has been presented in this section, is concretized and illustrated by this image of a computer simulation. Here it becomes understandable, what *constructed, substance-less accidents* are, which need a material basis, and what it means, on the contrary, that the actually existing system develops due to the *activity* of its elements *according to their essence,* which emerges from the *inseparable unity of substance and accidents*:

In the simulation, the values of the variables are stored on a material basis. They are transported to a calculating unit, re-calculated and then transported back to serve as output for one image (frame) of the simulation. Here, the variables do not stand for the accidents of the storage medium or the calculating unit – they can be separated from them and form with them only a temporary combination for the sole purpose of the simulation. Thus, calculating unit and storage medium do not represent the substance that belongs to these variables, but only their material basis. A material basis, however, cannot turn into sensation – this can only occur with a substance that is inseparably bound to accidents from which more complex, non-derivable accidents emerge.

The actually existing system is not just a list of numbers that changes with time. In the real system, there are no numbers, which can be written into a memory, transformed and be removed, but *attributes of things,* which are inseparably bound to these things and from which the dynamics of the system unfolds. In the case of human neuronal networks, this leads to the development of non-derivable accidents and to the transformation of the according substance and, with it, to the emergence of mind.

Existence is more than a *sequence of strings with transformation rules.* Presumably, (almost) any human being knows that intuitively. If the confusion of reality and description were not already so far advanced, then the idea of the computer simulation could actually serve as paradigm for the insensitiveness of simulations, because it is so obvious that the totality of the acts of calculation and storage of variables can never produce a sensation or a feeling.

However as long as one has no concept of this intuitively evident fact, it is ultimately *unjustifiable.* Hitherto, exactly that has been the case, and this lack of conceptualization resulted in the unclearness and confusion regarding the question of whether simulations can feel and understand

Clarity is achieved only through the conceptual determination and analysis of the difference between reality and simulation, which has been carried out here. From this difference follows that computers

are not capable of feelings and that they will never bring forth mind, regardless to which extent their performance can be enhanced.

At last a note about artificial neuronal networks. If they are realized on conventional computers through *software,* then everything just said holds true. But what about artificial networks that are already constructed as such, that is: which are realized through *hardware?* Can they break through the previously determined boundary for simulations? Are they capable of producing mind?

No. Also in this case, our metaphysical argument holds true. Whenever the elements of a system are *constructed by us*, a layer of *substance-less accidents* is created. Thus, the substance is lacking which could turn into the second substance *sensation.* The emergence of sensation and mind does not take place.

This is a purely metaphysical argument. Can it be interpreted also *formally?* I think *yes*:

Reality is not determined, the future is not contained in the present. For this reason, *causality top-down* exists, and this means, that through self-organization complex, autonomous entities can evolve, which form a new layer of reality, with a dynamics of its own that is not derivable from any layer that lies below. Therefore, these complex entities can be regarded as *causes* of what happens – just as is the case with *mental states*.

In contrast, in a system constructed by us, the dynamics is always determined through *causality bottom-up.* The temporal development of the system states is completely determined by rules. Complex entities do not form a new, autonomous layer of the system. Their dynamics is determined by the dynamics that *we* constructed. *Mind* as autonomous phenomenon cannot develop. It is not possible to *construct* mind.[15]

The transition from an actually existing system to a simulation, i.e. to a constructed system, is tantamount to a *depletion* of the system. The richness of the possibilities of shape formation that lies in the *essence of the reality* – that is to be substance *and* accident – disappears in the construction. It consists *only of accidents,* such that it lacks the metaphysical quality *activity,* and this means that a construction – as well as a description – cannot imitate the way in which reality generates the future.

The fabric of reality is of infinite subtlety. It cannot be replicated by our constructions.

[15]This remains true if the construction is non-algorithmic. Even then, the emergence of mind is ruled out through the metaphysical argumentation. The metaphysical argument is therefore stronger than the formal one.