Heinz Heinzmann

The Incompleteness of Natural Science and Its Correction

1. The self-contradiction of natural science

One of the **fundamental assumptions** upon which natural science is based is as follows:

Everything that exists consists of elementary objects that interact with each other. How these objects behave is governed solely by physical laws and by nothing else.

Thus, the entire future development is determined exclusively by so-called "initial conditions" – the totality of the attributes of all these objects at any given time – and physical laws.

If this assumption is correct, then there is ***only physics***.

Causality is then always "below", in the elementary layer of reality.

All other, more complex layers have lost their independence. We still need descriptions that refer to such layers – for example, neural or psychological descriptions of our mental activity – but only because it would be impossible to represent this activity in a physical way; but that doesn't change the fact that *it is, in fact, physical in nature*.

Here we encounter a case of "specific blindness": the inability to recognize a particular problem, which makes its solution more difficult or even impossible.

In the present case, it is this problem:

From the above assumption, it follows that we cannot argue or reason.

For it is evident that:

Every argument or conclusion presupposes that one thought follows from another thought.

But that is ***the definition of mental causality***.

To assume that we are capable of reasoning, therefore, means to ***postulate causality at the level of mental processes***. The idea that ***thinking itself*** leads to correct results obviously presupposes its causal effect. How else could it be possible to *mentally* correct an error? If my thinking were not *itself* causal – would *physics* then correct itself?

One must therefore decide: causality lies ***either*** in thinking ***or*** in physics – both at the same time are not possible. Thinking would then be "causally overdetermined".

In other words:

Precisely what cannot exist according to the fundamental assumption of natural science quoted above – argumentation and reasoning – is constantly presupposed and practiced by philosophers and natural scientists; it is the basis of their existence.

Natural science is thus in a ***permanent self-contradiction***.

Proposition:

If all future development is determined exclusively by physical laws and initial conditions, then inferential reasoning, and consequently also natural science and philosophy, are impossible.

And from this follows:

Proposition:

In addition to initial conditions and physical laws, there must be another element in reality upon which the future development depends, but which is not taken into account by physics, and this means:

The current scientific view of reality is incomplete.

2. The missing element

What is this "missing element"? How can it be found?

Here we encounter another, quite fundamental blindness: In fact, the sought-after element of reality is completely visible and also generally known. Yet, it has so far been neither acknowledged by philosophy nor by natural science – presumably precisely *because* of its seemingly trivial self-evidence.

Its presence follows from the ***difference between reality and description***:

Proposition:

There is a fundamental difference between a really existing object and its description: The really existing object is active, while the description is not active.

Thus, the existence of real objects must include something ***from which*** their activity emanates, but which objects in a description lack.

This element of the existence of real objects I call ***Substance***.[1]

Therefore Substance is that, which activates existing objects.

We know that Substance *must be there*. However, what it "**is**" cannot be defined, perceived, or imagined. It is ***unthinkable***.

The element of the existence of real objects that we can perceive, describe, and define is the ***way*** *in which they are active*, i.e. their behavior and their effects.

This element of their existence I call ***accidents***.

Natural science deals *exclusively* with accidents.

However, **Substance is always presupposed:** We know that objects are **activated** by ***mass*** or by ***charge***, but we do not know what mass and charge "**are**".

Therefore, the following applies:

Proposition:

Real objects consist of Substance and accidents.

Objects in a description system consist only of accidents.

Since an object cannot *cease* to be active in its characteristic way, ***Substance and accidents form an inseparable unity***. (The Earth exists only ***with*** gravity.)

1 I chose the term *Substance* because – in my definition – it is a necessary condition for solving the philosophical problems that are historically associated with this concept.

Every existing object therefore consists of these two elements:

– of **Substance** – that part of existence whose presence we recognize as necessary, but which, as what it actually "**is**", can neither be conceived nor defined; and

– of **accidents** – that part of existence that can be described and defined.

Let us now return to what I previously called the "permanent self-contradiction". The question is:

Can the element of reality just described serve to eliminate this contradiction?

The answer is ***yes***. As follows:

The fact that reality is ***active*** means: at any point at any time exactly what has to happen happens ***by itself***. It means that reality doesn't have to ***calculate*** anything, that it doesn't need a law or an algorithm, because it simply processes all individual cases at the same time.[2]

Obviously, however, *activity* is precisely that which cannot be transferred from the reality to its representation. It can be said that the *type* of activity of the system, its *specific structure*, must be contained in our equations, but the *activity itself* is missing.

Let us note:

Because of its *activity*, reality advances ***by itself*** from the present to the future. But the description system refuses to do us this favor. In order to obtain information about the future of the system in our description, we therefore need a *mathematical procedure* that ***substitutes*** the missing activity.

Do we have such a procedure?

Only in idealized, simplified scenarios. Every real event is determined by numerous processes – interactions between objects – that constantly influence each other. To be able to calculate any of these processes, however, we must assume, at least for a small time interval, that its immediate environment is constant – we must therefore *isolate* it briefly. Then we can do the same for all other processes, and then repeat the whole procedure for the next time interval, and so on.

The crucial point is that from start to finish we depend on *approximations*, and that we also do not know to what extent our calculations deviate from reality. At the latest after the next branching point – that is a point in the development of a system at which an arbitrarily small difference in the initial conditions can lead to completely different states of the whole system – our prediction becomes pure luck.

To derive the future precisely, we would therefore have to make the successive time intervals of our calculations *infinitely small*, and that is impossible.

This means: ***There is no procedure that leads from the present to the future***.

Proposition:

The future development does not follow from physical laws and initial conditions.

But isn't *reality itself* constantly showing us that the future *follows from* the present? Not at all. What we see is just that the future "follows" the present. It is only this suggestive picture of reality conveyed by physics that leads us to believe that everything "follows from" initial conditions and laws. However, the expression "follows from" is a logical conjunction that can only relate to a description. To apply it to reality means to replace the "follows" that we observe with the "follows from" that we postulate; But we have to *justify* this act of substitution, and so we are forced to

2 It would be more than absurd to assume that a blade of grass ***calculates*** where it should move – it simply follows the wind that touches it.

replace our "follows from" by a series of logical steps. Thus we inevitably end up with a mathematical procedure, and finally again with the fact that no such procedure exists.

Proposition:

Physical causality is incomplete. There is room for causality in complex layers of reality.

This resolves the self-contradiction of the physical worldview, and the conclusions drawn so far also form the basis for solving the philosophical problems listed in the Contents on page 1.

3. Proof of Free Will

(I will present the proof here in abbreviated form. The full version is the first part of the paper "Why we have Free Will and Consciousness and AI Systems do not.)

We have just completed the first step of the proof:

We have freed the mind from the grip of physical causality by showing that the *activity* of reality cannot be replicated by logical or mathematical procedures, so that the claim that everything *follows from* physical initial conditions and laws cannot be sustained.

The second step is to determine how causality in more complex layers of reality – I call it causality "from above" – is to be understood.

We consider a simple glass vessel. When we hit it, it vibrates and makes a sound. What does this tone depend on? What determines its height and character?

The answer is:

The shape of the vessel. It gives rise to a mathematical law that enables us to predict the vibration pattern of the glass. So here we don't have to go into the physical objects – the glass molecules – nor the physical interaction – the electromagnetism – in order to predict the sound. The only physical information needed is the speed of the sound propagation in the glass.

The law that now allows us to predict the future of the system is therefore *not a physical law.* It belongs to another kind of laws which I shall call ***Laws of Form*** or ***Laws of Structure***.

The sound that we hear is largely independent of the way we produce it. However, this does not apply to the first moment: initially, there is a transient process that depends on how we strike the vessel. Only after this process it does always vibrate in the same state.

This state to which the glass ultimately adapts – the vibrational pattern into which it develops and which it then maintains – is called ***attractor***.

However, most important for our considerations is undoubtedly the following:

The local parameters – such as the positions and velocities of the glass molecules – initially depend on where, with what and how hard we hit the vessel. So at first they can be quite different. Regardless of this difference, the state of the vessel always evolves towards the same vibrational pattern – the attractor.

In the case of a glass vessel, there is only one possible vibration pattern that always develops, regardless of how the vessel is struck. The future movements of the components of the vessel – the glass molecules – are therefore determined by this pattern.

Causality works from the whole to the individual, from the vessel to its components, and not the other way around.

Proposition:

A form of "causality from above" occurs when in a system *attractors* exist, i.e. states which the system will *inevitably* evolve into, if it is "close enough" to the attractor state.

Now we have made all necessary preparations to move on to our final and decisive scenario: *our own neural network*.

It consists of many billions of neurons. Each neuron is directly connected to hundreds or even thousands of other neurons, and through a few intermediate steps, all neurons are interconnected. Therefore, what we established above regarding the fundamental uncomputability of complex systems also applies to neuronal processes:

The high degree of interconnectedness of the neurons precludes the existence of a mathematical procedure for calculating the further development. Thus, physical causality is *incomplete*.

The next question is: Does the previously defined causality "from above" exist here?

We make the following assumptions:

1. Every kind of mental activity (thoughts, chains of associations, sequences of images, etc.) is a sequence of neural activation-patterns.
2. Sequences of neural activation-patterns can be representations of facts.[3]

Let us look at the neural patterns. How do they become representations?

Let us imagine the neural network of a child. An object perceived for the first time will cause a certain pattern in this network, starting from the primary visual cortex. The neural connections that are active are strengthened because of this very activity. The same is the case with each repetition. This gradually creates a stable connection between the object and a specific neural pattern (or rather an ensemble of specific neural patterns).

In addition, the following applies: Although the neural patterns are initially caused by external stimuli, after a sufficient number of repetitions they are also produced by the neural network independently of these stimuli. This means:

Neural patterns that are connected to objects in the manner just described are* attractors *of the network. (See also the associated note)

Previously we have stated:

Under the condition that the causality from below is incomplete, from the existence of attractors follows that the respective system – provided it is "close enough" to the attractor state[4] or in this state itself – is governed by causality from above.

However, according to our first premise, a mental process consists not only of neural patterns, but also of the transitions between these patterns. But to this transitions the same applies as to the patterns themselves: First, they are determined by the sequence in which the causative objects appear. If this sequence is repeated, the corresponding neural activity is reinforced, and this has the consequence that the patterns occur again in the same sequence even if they are generated by the network itself. In the same way, also the spatial relationships of the objects are transferred to the patterns.

This means: In the processes that are generated by the network itself, the neural patterns that are in a stable connection with specific objects appear in the same spatial and temporal contexts as the

3 Here, "facts" must be understood in the widest-possible sense.

4 The neural network is always "close enough" to an attractor state: from *any* state, the network will immediately adjust to a pattern that *means* something.

objects themselves. Therefore, *the patterns can be understood as representations of the objects, and the processes as representations of the facts in which the objects appear.*

So, in human neural networks it is not the physical or neural conditions and laws that determine what happens in the network, but *the structure of the network* – the fact which attractors there are and how their sequence is regulated – on which the processes depend that run in the network.

Causality acts from the whole to the individual, from the network on its components, and not the other way around.

We have thus achieved our first goal:

Proposition:

The neural network is regulated by *causality from above*. The mental level is the dominant level. In it lie the *causes* for the processes running in the network.

So the statements we made so far were *actually* conclusions and not just physical processes! – Insights are insights, thoughts are thoughts, mind is set in its rights, *we ourselves* are indeed we ourselves …

So far, so good, but that doesn't take us to where we actually want to be. Just because we have moved causality up doesn't mean we are free. We have only replaced physical or neural causality with mental causality. We have thus achieved that our mind is not ruled by physical or neural laws, but *by its own law: the Law of Structure, which the sequence of neural patterns obeys that represent something*.

But don't we ultimately remain trapped in the scheme of initial conditions and laws from which we wanted to escape? Fortunately, that's not the case. To show this, we need to look at the difference between physical and mental laws:

As stated above, initially the order of the patterns is determined by the order in which the objects or circumstances that cause the patterns occur. But as soon as the network itself is able to produce these patterns, the transition rules of the patterns – what we have called the *mental law* – increasingly depend on their use in internal processes. This dependence on external and internal conditions means that the transition rules differ from person to person.

So we have already determined the first difference:

*While physical laws are **generally valid**, mental laws are **individually valid** – they only apply to one singular person.*

Connections between neurons are strengthened when they are active, and weakened when they are inactive. This means that every mental activity alters the structure of the network. But if the structure can change, then obviously also the rules that determine the sequences of the neural patterns can change.

So this is the second difference:

*Physical laws are **immutable**, mental laws are **modifiable**.*

Proposition:

Physical laws are universal and immutable. Mental laws are individual and modifiable.

With this, we are sufficiently prepared to substantiate our freedom:

We have shown that causality is to be found at the mental level.

Will and *intention* must be understood as elements of the mental causality.

Now let us imagine concretely we were faced with an important decision. When we enter the decision-making process, we are initially guided onto certain, well-known paths by the regularities that are valid up to that point – i.e. by our own mental law.

But at any time we are able to leave these paths, for example by simply considering the opposite of what we have assumed up to then, or by taking a path we never tried before; We are able to do so precisely for the reason that the causes for what happens in the network – and thus also for the modifications of the network structure – lie at the mental level.

In other words:

The law that determines the sequence of neural patterns in our network that represent something, i.e. our own mental law, can be altered *by ourselves*: we ourselves can change the laws of our thinking and acting through our thinking and acting, and we can do it *deliberately.*

This means at the same time:

Although mental processes are governed by their own rules, it is not possible to derive a volitional decision from them: the decision cannot be contained in these rules because they can be changed by the mental process that precedes the decision. While this process is taking place, the laws that it obeys can change – or, more precisely, *it itself* can change the laws that applied before it started.

Proposition:

Volitional decisions are causes of actions. Since only by the decision-making process itself is decided what will happen, the decision is not determined beforehand.

So the decision is free.

To the question of why a (sane) person has decided so and not otherwise, there is then only one permissible answer:

Because he/she wanted it that way.

Note:

Of course this does not mean that volitional decisions cannot be analyzed with respect to their neural, chemical, physical, genetic, social, psychological etc. causes. It means, however, that these analyses necessarily remain incomplete and never lead to a secure result, because mental phenomena cannot be reduced to other layers of reality. The will remains the final authority.

Note:

In order to recognize objects, artificial neural networks must be trained on large data sets. In numerous repetitions the connection strengths of their neurons are varied until a sufficiently high recognition rate is achieved.

In contrast, we started from the following hypothesis:

A perceived object, which causes a neural activation pattern, is represented *by this pattern itself*.

Therefore, here the relationship between object and representation is not established by varying the connection strengths of the neurons, rather it exists already from the beginning and is only stabilized and specified by *strengthening* the active connections, whereby the neural pattern becomes an *attractor.*

This hypothesis is confirmed most clearly by the so-called "imprinting". (As e.g. in the case of the gray geese of Konrad Lorenz). There are neither "large data sets" nor "numerous repetitions" – the process occurs almost instantaneously.

Furthermore, thereafter *immediate recognition* occurs, despite the inevitable variability of the sensory impression to be recognized. Thanks to the attractor concept, this – otherwise hardly explainable – performance becomes self-evident: as long as the sensory input is within the catchment area of the attractor, it obviously applies:

perceiving = recognizing,

since the newly activated attractor already represents the object, so that further calculations are unnecessary.

To the hypothesis that objects are represented by attractors, the following should be added:

The pattern that forms in the primary visual cortex as the consequence of a perceived object, is not *as such* transferred directly into the neural network. Rather, it is broken down into several components – in this sense *parametrized* – which, at the end of the whole visual data processing, are assembled to the overall neural pattern that we understand as attractor.

This parametrization is an important aspect of the attractor hypothesis: The attractor is defined by a subset of the phase space. The *attractor state* of the system corresponds to a trajectory that does not leave this subset for a certain period of time. However, already a (small) subset of all according parameter values – which do not even have to be very accurate – is sufficient for restoring the attractor state, which means: a *fraction* of the complete original sensory input is sufficient for recognition.

This makes recognizing objects extremely easy and, at the same time, increases the ability to generalize objects and facts.

Here is an example which demonstrates both aspects of our hypothesis: recognition after only one encounter and ability to generalize:

When a child sees a picture of a giraffe for the first time, it will later recognize not only the giraffe in this picture, but also all giraffes shown in other pictures. It is therefore in possession of the General under which all examples are subsumed.

4. The "Hard Problem": How is sensation possible?

(The full version is contained in the second part of the aforementioned paper Free Will. There, however, it is combined with the proof against AI sensation. Here, I separate the two proofs.)

In the previous section on free will, we proved that the mental layer is the *causal layer* of the neural network.

The dynamics of the neural network is thus determined not by objects of the *physical layer* and their accidents: ***atoms***, ***molecules***, and ***physical interactions***, but by objects of the *mental layer* and their accidents: ***mental states*** that represent or mean something, and ***information processing***.

Previously we have established:

Every mental state is a neural activation pattern. These patterns are attractors of the dynamics of the neural network. Every mental process is a sequence of such patterns.

These statements address the question of how the objects and processes of the mental realm can be understood in terms of their *material preconditions*.

But now our task is to grasp them for what they are as *mental phenomena*.

The answer is:

Proposition:

Every mental state is a combination of two disparate elements: information and sensation.

Its **information** content is what it ***represents*** or ***means***.

Sensation must be understood here in the broadest possible sense: It stands for everything in a mental state that goes ***beyond information***, that is, for that which ***cannot be defined*** but can only be ***felt*** and ***experienced***.

Two examples: the frequency of the color "red" can be defined, but the sensation ***red*** cannot; the intensity of pressure can be defined, but the sensation ***pain*** cannot.

At this point, we must again address a "specific blindness", which here even appears in two forms:

Many natural scientists ***equate*** mind and information processing.

This is the first form of this blindness. It precludes a *scientific answer* to the question of what mind is, for the following reason:

Everything that ***can be defined*** is ***accessible*** through information processing; everything that ***can not be defined*** is, in principle, ***inaccessible*** to information processing: regardless of which function is applied to information – the result is always merely information and nothing else. The information "red" will *never* turn into the sensation ***red***, the information "pressure"will *never* turn into the sensation ***pain***.

And from this follows:

Proposition:

Mind is more than information processing.

The second form of blindness concerns the realm outside of natural science. Here, too, the view of sensation is completely obscured, albeit in a very different way: there are no justifiable answers to the questions of how sensation can be ***explained*** and ***whether*** or ***how*** its occurrence is ***possible***, but only strange constructs that owe their existence to the failure of all attempts at explanation.

However, I will not discuss the various positions – qualia eliminativism, panpsychism, consciousness as a separate or even fundamental form of existence, etc. – but will proceed directly to the conclusions that arise from what has been said so far.

In Section 2. The missing element, we have determined (see here):

Every existing object consists of an ***undefinable*** and a ***definable*** part, which are ***inseparably linked*** with each other: ***Substance and accidents***.

The objects that are now the subject of our analysis are no longer objects of the *physical realm*, but rather objects of the *mental realm* of reality.

We must therefore apply the above statement to *these* objects; in other words, we must conceive of mental states as *inseparable units of Substance and accidents*.

What is the Substance of a mental state, and what is its accident?

Our starting point is the above proposition:

Every mental state is a combination of two disparate elements: information and sensation. [5]

5 Here and in the following, "sensation" always refers, as stated above, to "that which is *not definable*, which therefore goes *beyond information*".

Evidently, *information* is that which is accessible to our thinking – that which ***can be defined***.

Thus, *information processing* is the accident of the mental state.

In contrast, *sensation* is that which ***cannot be defined***, that which therefore eludes our thinking.

Thus, *sensation* is the Substance of the mental state.

And this means:

Sensation is the driving force behind the dynamics of the mind.

With this, all the preparations have been made to explain the occurrence of sensation. Before we begin, however, we must determine ***what*** is actually to be explained.

What do we mean by "explanation"?

For example: If I explain what a car is, then I describe its technical structure and thereby establish its function. The goal of the explanation is thus the definition of what is to be explained – the success of the explanation *presupposes* the possibility of definition.

Therefore applies:

Since sensation is not definable, it cannot be explained. [6]

What we can explain, however, is why sensation exists and why it is "different".

We have already provided the *first part* of this explanation by determining sensation as the Substance of mental states, which we understand as *inseparable units* of sensation and information.

From this follows that sensation is not only a ***possible***, but even a ***necessary*** element of the mind.

What is still missing, however, is the second, more important part of the explanation:

Why is sensation "different"?

5. Why sensation is unique

Why is sensation so fundamentally different from everything else we encounter in reality?

The usual question is:

"Why is there something indefinable in the mind, like 'color' or 'pain', and nowhere else?"

We ask ourselves the question instead:

"Why does the indefinable, which exists everywhere in reality, change its character when it appears in the mind?"

So the question is not about the reason for the ***existence*** of this indefinable, which would be superfluous because – as we have shown – it can be found *in everything that exists* and is therefore self-evident, but about the reason for its ***change***.

In the usual version, the question is unanswerable.

In this (false) form, it leads to strange hypotheses, such as *qualia eliminativism* or *panpsychism*.

But, as we will now show, in our version the question can be answered.

6 Here, too, there is a widespread blindness: many scientists are convinced that the rapidly growing knowledge in their field will one day be sufficient to *explain* sensation.

Prerequisite for our argument is the following

Proposition:

As long as accidents of higher complexity can be described as functions of accidents of lower complexity, the associated Substance remains the same. If this functional connection is broken, the Substance changes. *For us* it then appears as a new, second Substance.

Before we turn to proving this proposition, we must clarify to what extent accidents in more complex layers of reality can be described as functions of accidents in simpler layers.

For example, the processes in neurons can be described as functions of the physical and chemical properties of these neurons. (Which does not mean, however, that they can be *calculated*.)

The same applies in principle to all evolutionary transitions: from the physical to the chemical level, then to the biochemical, cellular, neural level, and even up to the realm of simple neural networks that do not produce mind: the processes taking place in such networks can be described as functions of their architecture and external conditions.

Only at the very last of these transitions – the transition to neural networks that produce mind – does the chain of reducibility end:

As we established when substantiating free will, then the following applies: Initially the order of the neural activation patterns is determined by the order in which the objects or circumstances occur that cause the patterns. But as soon as the network itself is able to produce these patterns, the transition rules of the patterns – what we have called *mental law* – increasingly depend on their use in internal processes.

This means that the dynamics of the neural network – i.e. *the mind* – increasingly decouples itself from the causal chains of the environment and instead develops its own internal laws. *Mental causality* emerges.

And from this follows that the information content – i.e. the *accident* of mental states – can no longer be represented as function of the accidents of the underlying layers of reality.

Now to the proof of the above proposition.

The totality of physical accidents we will call ***first accident***, their associated Substance ***first Substance***, the totality of mental accidents ***second accident*** and their associated Substance ***second Substance***.[7]

We have just established that the accidents of all evolutionary levels can be traced back to the accidents of the levels below, with the exception of the accidents of the highest, i.e. the mental level.

It applies:

Substance and accident always form an ***inseparable unity***.

The *first accident* is *inseparably* linked to the *first Substance*.

If complex accidents can be reduced, step by step, to simpler accidents, then this means that they can ultimately be reduced to the first and simplest accident.

For us, however, *reducibility* is tantamount to *ontological* ***identity***:

If B is reducible to A, then B **is** actually A. So if a complex accident is reducible to the first accident, then it **is** actually the first accident, and then it is *inseparably bound* to the first Substance.

7 However, that does not mean that there are two Substances – rather, the second Substance is thought of as emerging from the first Substance, and the question we ask ourselves is therefore: Why is the *first Substance* in the case of qualia *for us* transformed into the *second Substance* sensation?

Thus, as long as the accidents are reducible, the associated Substance remains the same – it is then still *first Substance*.

But if the chain of reducibility to the first accident is interrupted by the appearance of a new, *irreducible* accident, then this new accident ***differs*** from the first accident and from all other accidents that can be derived from it.

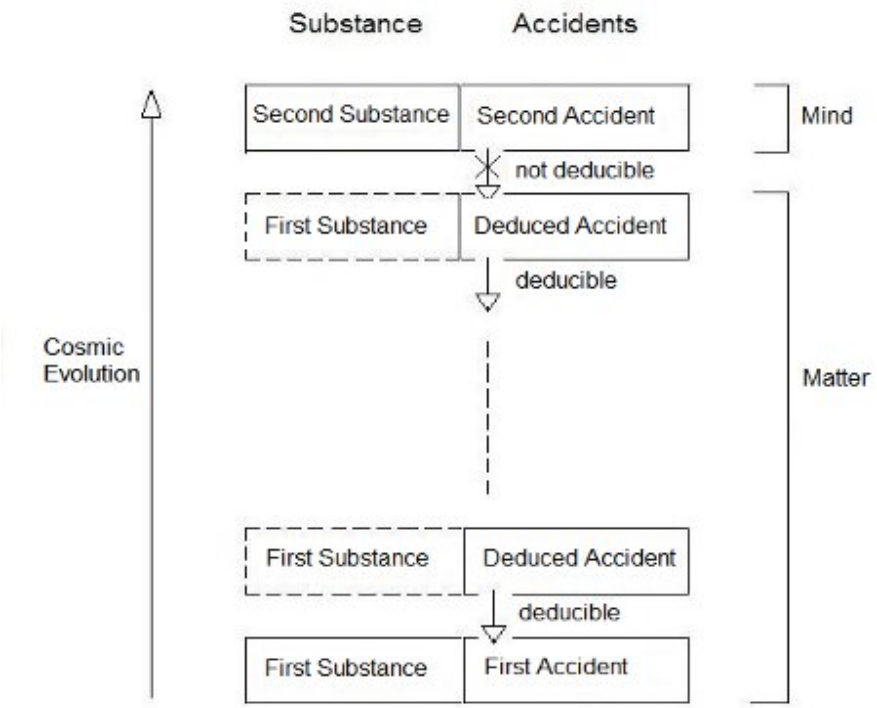
Due to the ***inseparability*** of first Substance and first accident, it applies:

*If the Substance of an object is the **first Substance**, then the associated accident must be the **first accident**.*

And from this follows:

If an accident appears that is <u>different</u> from the first accident, then the associated Substance must also be <u>different</u> from the first Substance.

Here is a sketch for illustration:



With this, we have proven that sensation, the Substance of the mind, is <u>not physical</u>, that it must therefore be something <u>other</u>, and the nature of the proof suggests that sensation is ***the only thing*** in the universe that is not physical.

Thus, everything that can be explained has now been explained:

As Substance of the mind, ***sensation is necessary***, and because the change of the Substance occurs only in the ***final*** evolutionary step – the emergence of mind – ***sensation is unlike anything else: it is <u>unique</u>***.

What it is, however, ***cannot be explained***. Substance is the *indefinable* part of everything that exists, and what is indefinable is not explainable. Still, although we cannot ***explain*** what sensation is, we ***know exactly*** what it is, because it is a ***part of ourselves***.

Yet this knowledge is not *conceptual*, *definable*, *communicable knowledge*, but *cognizance of what is immediately given*.

6. Why AI systems are not sentient

If mind is *equated* with information processing, which is consistently the case in AI research (blindness), then it follows that we are separated from creating an AI system with consciousness – which is also intellectually superior to us in every respect – not by *principal* but only by *technical* difficulties, which, so it seems, we will overcome in the near future.

However, this is wrong, because the just-given explanation for the *emergence of sensation* can be extended in a surprisingly simple way to prove that *AI systems are not capable of sensation*. And since **consciousness** is *not only information*, but ***information and sensation***, AI consciousness is also ruled out.

The proof begins with the following

Definition:

What an object is due to the inseparable unity of its Substance and accidents, we call its essence. The activity that results from this unity we call essential.

(Thus the ***essential activity*** of the Earth is to exert gravity.)

The purpose of this definition becomes immediately clear when we now turn to *simulations*.

For example, consider a mechanical simulation of the solar system in which the model bodies are moved through mechanical devices – chains, gears, shafts, etc. – in this way mimicking the movements of the celestial bodies.

The *essential activity* of the model bodies would obviously be to *exert gravity*. But it is *not the mass (the **Substance**) of the model bodies* that drives the dynamics of the simulation – that is, what causes the desired movements – but *the mechanics we have constructed,* which must then be *activated*, electrically or mechanically (e.g. by turning a crank).

To express this point, we will refer to this type of activity as ***supplied activity,*** in contrast to the just defined ***essential activity***, which happens ***by itself***.

With this, the definition of ***simulation*** takes the following form:

The dynamics of a simulation – contrary to the original – is not caused by the essential activity that arises from the inseparable unity of Substance and accidents of the objects of the simulation, but by supplied activity.

The accidents from which the dynamics of the simulation is formed are therefore ***not activated by Substance***: the Substance of the objects of the simulation (in our example their mass) *is not the Substance that belongs to these accidents* and with which it forms an inseparable unity, but only their material basis from which these accidents can be separated at any time. (As is immediately apparent in the mechanical simulation of the solar system.)

However, as we have shown in the previous section (see here), the ***inseparability of Substance and accidents*** is a ***necessary condition*** for the transformation of the Substance:

The transformation of the essence of being can only occur if the dynamics of the system arises from the ***inseparable unity*** of Substance and accidents. ***Only then*** does the transformation of the associated Substance follow from the fact that the accidents are no longer reducible to the first accident.

In ourselves, this condition is fulfilled: the Substance is transformed – we have sensations.

But the dynamics of a simulation is based on *supplied* activity. Thus, the accidents are ***not activated by Substance***, and the Substance that belongs to the objects of the system ***does not form an inseparable unity with these accidents***.

And this means:

There is no reason for the transformation of this Substance. It remains ***first Substance***.

In other words:

The essence of the simulation remains physical. The simulation remains an information processing system without sensation.

The metamorphosis of matter into mind does not take place.

The just mentioned condition that the dynamics of the system must arise from the *inseparable unity* of Substance and accidents, does not only apply to the last, i.e. the mental level – it must be satisfied on *every* level that develops during the evolutionary rise from matter to mind. If on any of these levels the dynamics of the system is not caused by the *essential activity* of the objects but by *supplied activity*, then the unity of Substance and accidents is torn and the transformation of the essence of being can not occur.

The question is: How far does our proof that AI systems cannot possess consciousness extend?

For AI systems that are implemented using software on conventional computers, the proof is valid without exception: the use of software is always associated with supplied activity.

But what about a *replica* of a biological neural network that reproduces the neural (analog-digital) input-output law using suitable hardware and whose structure corresponds to the structure of the entire network, so that it could be assumed that the sequence of states of the *constructed system* would almost be identical with the sequence of states of the *biological system?*

Could the transformation into sensation take place here?

The answer is clearly ***no***. The condition for the transformation is not met: the dynamics of the replica is *not caused by* ***essential activity*** *but by* ***supplied activity***.

The problem is that from the usual scientific view of reality ***this fact cannot be understood at all***.

In this view, reality is ***equated*** with a – describable and definable – sequence of states, and it must therefore be expected that the increasing convergence of two sequences of states ultimately leads to the identity of the systems themselves.

However, in the expanded materialist view that we have presented here, the concept of existence is augmented by an element that takes us ***beyond*** the realm of the definable.

This means that all our descriptions and ideas about the processes in nature are necessarily *incomplete*. So to speak "behind the scenes" of the part of the stage that is accessible to us, something happens, which is either completely hidden from us or can only be recognized and understood through inference from the part of reality that is accessible to us: the accidents.

Here, reality is <u>more</u> than a sequence of states.

In the context of our considerations, this implies:

From the approximate identity of the state sequences of the natural system and the artificial system cannot be concluded that also their essence is approximately identical.

Concretely: The Substance of the two systems can be quite different despite the extensive identity of their states:

In the ***biological system***, the Substance is ***inseparably bound to the accidents of the system*** and is therefore ***transformed into the mental Substance sensation.***

The ***constructed system***, however, is driven by ***supplied activity,*** and therefore here the Substance stands in a merely constructed and *by no means inseparable* connection with the accidents of the system, so that it ***remains physical Substance and is not transformed into sensation***.

The result of our conclusions is as follows:

Proposition:

It is not possible to construct an AI system that experiences sensations and has consciousness. Neither in a simulation nor in a replica of a system that produces mind can the transformation of matter into mind take place.

There is no ghost in the machine.

Thus only *artificial intelligence* can be constructed and not *artificial mind*.

Does this mean that it is impossible to create artificial mind at all?

No. Our argument only excludes the possibility that mind can be *constructed*. However, the definition of the term *replica* can be expanded to include artificial evolution, i.e. an evolution that is designed and controlled by us.

In this case – just as in natural evolution – the condition could be met that the respective system activity is always *essential*. If we do not intervene at any point in this artificial evolutionary process through constructions or by supplying activity, but limit ourselves to controlling and accelerating the development, then at the end of this evolution there *could* be a system that produces mind.

However, no one can know whether such an artificial evolution is possible, or whether the path that nature has chosen is the only viable one.

In any case, it is clear that the creation of artificial mind remains a very distant, perhaps never achievable future, if it is not impossible at all.

7. Sensation and Artificial Intelligence

I have determined "sensation" differently from its usual meaning. I would like to explain in a little more detail why this was necessary and what follows from it:

Every mental state contains something that is ***not definable***, which goes ***beyond information***.

However, since there is no term which all possible elements of mental states can be assigned to, I have instead chosen the term that comes closest to this missing term: ***sensation***.

Therefore, on the one hand here the term "sensation" is restricted compared to its common use – because it is supposed to contain *no information*, i.e. no *definable* part –, but on the other hand it is also significantly expanded.

Two examples were used for illustration: *color* and *pain*. Color, because the indefinability of the color sensation is a known fact, and pain, because it is perfectly understandable that the event "hammer blow on finger" triggers a mental state that contains not only the information "hammer head is in contact with finger" but *something more*: the sensation *pain*, which can be so strong that it is impossible to deny its occurrence.

"Sensation" understood in this way can be divided into three areas:

A) The first area is the area of *perception*:

Sensation encompasses the entire "inner theater": the virtual space, the stage on which we act, which is always present to us as a whole – as an "image" – and on which we see, hear, feel, smell and taste.

While there is little doubt that the sensation *color* cannot be defined, it may initially seem as if we are returning to the area of the definable, if our perceptual image is *colorless*: *gray values* are definable, aren't they? – Yes, they are, but the *sensation* associated with them *is not*: only the intensity of the light can be defined, and also the neuronal excitation that results from it. But when we move on to *perception*, we leave the realm of information: the *brightness* that we perceive is just as much a *sensation* as the *color.*

And the same applies to all other senses: the frequency of a sound can be defined, but the sound-*sensation* can not, etc.

This means:

If sensation is lacking, then there is no "inner theater", which is made up of sensations.

So to put it very clearly:

AI systems do not see, do not hear, do not feel, do not smell, do not taste.[8]

Unfortunately, our language is not suitable for distinguishing between system states *with* sensation and those *without*. For us, "seeing" or "hearing" simply means what it ***is*** for us, and that is in any case information ***and*** sensation.

Therefore, *strictly speaking*, statements about perceptions are only correct if they refer to humans or higher animals, otherwise they are wrong: robots *do not see*, bees *do not see* – they only process frequency, intensity, and direction *information*.

However, pixels that only transmit ***information*** about brightness and color cannot be combined to form an image, unlike the *same* pixels when their content is ***perceived*** as brightness and color: it is immediately clear that they can then be added together to form an image.

B) The second area is the area of *feelings and moods*. Nothing further needs to be explained here.

AI systems experience nothing and feel nothing. They feel neither happiness nor unhappiness, neither love nor hate. They are neither cheerful nor sad, neither in a good mood nor irritated.

This list can be continued at will, since every mental state is a *quale*, i.e. consists not only of *information* but also of *sensation*.[9]

C) We have determined *sensation* as ***Substance*** *of the mental state*. It follows that it must be understood as *cause of the mental dynamics*.

---

8 This also applies to simple animals, such as insects, for the following reason: We have shown that the emergence of sensation can only occur if the neural network develops its own, *internal* laws. A necessary (and sufficient) condition for this, however, is that the network contains *functionally unbound structures*, i.e. structures whose function is not determined genetically or by early programming. Only under this condition can (and will) the *network of neural states* (attractors) develop that we understand as *mind*.

For us, *having eyes* is synonymous with the *ability to see*. But this is wrong. For an animal that has a light-sensitive cell, the world is by no means *bright* – the animal only has the *information* about which direction the light is coming from.

9 Of course, there are also activities *without* sensation, such as reflex actions or automatically executed sequences of movements. However, these are not *mental* activities, but *neuronal* activities.

Accordingly, ***everything*** that drives our thinking and acting must have a component of *sensation*. There is no acting or thinking without a motive. Even purely logical reasoning can only take place if we ***want*** to find the correct solution.

Therefore applies:

AI systems cannot want or not-want anything. They know neither motive nor interest, neither curiosity nor rejection.

In this area, the lack of differentiation in language use is particularly problematic. Programmers speak of the "goals" or "intentions" of an AI system, of what it "strives for". In all cases, however, this is only an increase in a parameter value, and not *goals* or *intentions* as we understand them as elements of human action, which are always linked to emotions.

In summary:

1. ***AI systems cannot perceive anything***.
   They lack the *"inner theater"*, the *"image"* of the environment: they cannot *see*.
   Likewise applies: they cannot *hear*, *feel*, *smell* or *taste*. For them there is only *information*.
2. ***AI systems cannot experience anything***.
   They have no feelings.
3. ***AI systems cannot want anything***.
   They lack intentionality and motivation.

No matter what the future of AI may look like, due to the limitations mentioned above AI systems will *never* be a new, superior species. The dystopias in which we are at their mercy belong in the realm of fantasy.

(In the third part of the paper Free Will I examine to what extent the capabilities of current and future AI are limited by its lack of sensation.)

Note:

Our proof also refutes the so-called "simulation hypothesis": if our reality were a simulation, then *we ourselves* would have no sensations and no consciousness.

Note about consciousness:

As already stated before, the following applies:

Everything that can be defined is attainable through information processing, *everything that can not be defined* remains unattainable for it: no matter what function is applied to information – the result will always be just information and nothing else; the information "red" will never turn into the sensation *red*, the information "pressure" will never turn into the sensation *pain*.

Therefore, "**information**" and "**sensation**" (as we used it above) form **the *only* pair of concepts** that makes it possible to draw a clear and definite line between artificial intelligence and human mind and to provide evidence for it.

From this follows that the concept "consciousness", which is often at the center of the discussion, is only suitable for drawing this boundary, if the mental phenomena attributed to it (in its respective definition) are analyzed and classified according to their affiliation with *information* or *sensation:* the part of consciousness that belongs to information processing (e.g. any kind of object-representation, even self-representation) can be *reproduced* – no matter what technical difficulties

stand in the way of its simulation, while the part that goes *beyond information* (perception, experience, volition, desire, suffering, empathy, etc.) remains *inaccessible* to AI.

It would therefore be an unnecessary and misleading complication to base the difference between AI and mind on the concept "consciousness".

Heinz Heinzmann

Vienna, June 2026

Note:

This paper is the second chapter of my treatise "What Is The World Made Of?"

Since it addresses only questions that have not yet been resolved – and thus does not challenge established physical and philosophical orthodoxy, as do all the other chapters – I have decided to present this chapter also as a standalone work.