

Heinz Heinzmann

Die Unvollständigkeit der Naturwissenschaft und ihre Korrektur

1. Der Selbstwiderspruch der Naturwissenschaft.....	2
2. Das fehlende Element.....	3
3. Der Beweis der Willensfreiheit.....	5
4. The "Hard Problem": Wie ist Empfindung möglich?.....	9
5. Warum Empfindung einzigartig ist.....	11
6. Warum KI-Systeme nichts empfinden können.....	14
7. Über Empfindung und Künstliche Intelligenz.....	16

1. Der Selbstwiderspruch der Naturwissenschaft

Eine der **Grundannahmen**, auf denen die Naturwissenschaft beruht, lautet wie folgt:

Alles, was existiert, besteht aus elementaren Objekten, die miteinander wechselwirken. Wie sich diese Objekte verhalten, wird nur durch physikalische Gesetze geregelt und durch nichts sonst. Somit wird die gesamte zukünftige Entwicklung ausschließlich durch sogenannte "Anfangsbedingungen" – der Gesamtheit der Attribute all dieser Objekte zu irgendeinem Zeitpunkt – und physikalische Gesetze bestimmt.

Wenn diese Annahme korrekt ist, dann gibt es *nur Physik*.

Die Kausalität ist dann immer "unten", in der elementaren Schicht der Wirklichkeit.

Alle anderen, komplexeren Schichten haben ihre Selbstständigkeit verloren. Beschreibungen, die sich auf solche Schichten beziehen – etwa neuronale oder psychologische Beschreibungen unserer geistigen Aktivität – benötigen wir dennoch, aber nur deshalb, weil es undurchführbar wäre, diese Aktivität auf physikalische Weise darzustellen; Aber das ändert nichts daran, dass sie *eigentlich* physikalischer Natur ist.

Hier begegnen wir einem Fall von "spezifischer Blindheit": der Unfähigkeit, ein bestimmtes Problem zu erkennen, wodurch dessen Lösung erschwert oder sogar verhindert wird.

Im vorliegenden Fall ist es dieses Problem:

Aus der obigen Annahme folgt, dass wir nicht argumentieren oder schlussfolgern können.

Denn offenbar gilt:

Jede Argumentation oder Schlussfolgerung setzt voraus, dass ein Gedanke aus einem anderen Gedanken folgt.

Das ist aber die *Definition geistiger Kausalität*.

Anzunehmen, dass wir imstande sind zu argumentieren, bedeutet also, eine *Kausalität auf der Ebene geistiger Prozesse* zu postulieren. Die Vorstellung, dass *das Denken selbst* zu korrekten Ergebnissen führt, setzt offenbar seine kausale Wirkung voraus. Wie sollte es sonst möglich sein, *gedanklich* einen Irrtum zu berichtigen? Falls mein Denken nicht *selbst* kausal wäre – würde sich dann etwa *die Physik* korrigieren?

Man muss sich entscheiden: die Kausalität liegt *entweder* im Denken *oder* in der Physik – beides zugleich ist nicht möglich. Das Denken wäre dann "kausal überbestimmt".

Mit anderen Worten:

Genau das, was es gemäß der oben zitierten Grundannahme der Naturwissenschaft nicht geben kann – Argumentieren und Schlussfolgern – wird von Philosophen und Naturwissenschaftlern permanent vorausgesetzt und ausgeübt – es ist die Grundlage ihrer Existenz.

Die Naturwissenschaft befindet sich also in einem permanenten Selbstwiderspruch.

Satz:

Falls die gesamte zukünftige Entwicklung ausschließlich von physikalischen Gesetzen und Anfangsbedingungen bestimmt wird, dann sind schlussfolgerndes Denken und damit zugleich Naturwissenschaft und Philosophie nicht möglich.

Und daraus folgt:

Satz:

Zusätzlich zu Anfangsbedingungen und physikalischen Gesetzen muss in der Wirklichkeit ein weiteres Element vorhanden sein, von dem die zukünftige Entwicklung abhängt, das aber von der Physik nicht berücksichtigt wird, und das bedeutet:

Die gegenwärtige naturwissenschaftliche Sicht der Wirklichkeit ist **unvollständig**.

2. Das fehlende Element

Was ist dieses "fehlende Element"? Wie kann es gefunden werden?

Hier treffen wir nun auf eine weitere, ganz grundlegende Blindheit: Tatsächlich ist das gesuchte Element der Wirklichkeit vollständig sichtbar und auch allgemein bekannt. Dennoch ist es bisher weder von der Philosophie noch von der Naturwissenschaft zur Kenntnis genommen worden – vermutlich gerade *wegen* seiner trivial erscheinenden Selbstverständlichkeit.

Sein Vorhandensein folgt aus dem *Unterschied zwischen Wirklichkeit und Beschreibung*:

Satz:

Zwischen einem wirklich existierenden Objekt und seiner Beschreibung besteht ein fundamentaler Unterschied: Das wirklich existierende Objekt ist aktiv, die Beschreibung hingegen ist nicht aktiv.

Somit muss zur Existenz wirklicher Objekte etwas gehören, *wovon* ihre Aktivität ausgeht, was aber Objekten in einer Beschreibung *fehlt*.

Dieses Element der Existenz wirklicher Objekte bezeichne ich als **Substanz**.¹

Substanz ist also dasjenige, was existierende Objekte aktiviert.

Von der Substanz wissen wir, dass sie *vorhanden* ist. Was sie jedoch "**ist**", können wir weder definieren noch wahrnehmen noch uns vorstellen. Sie ist *undenkbar*.

Dasjenige Element der Existenz wirklicher Objekte, das wir wahrnehmen, beschreiben und definieren können, ist die *Art ihrer Aktivität*, d.h. ihr Verhalten und ihre Wirkung.

Dieses Element ihrer Existenz bezeichne ich als **Akzidenzien**.

Naturwissenschaft befasst sich *ausschließlich* mit Akzidenzien. **Die Substanz wird jedoch dabei immer vorausgesetzt**: Wir wissen, dass Objekte durch *Masse* oder durch *Ladung* aktiviert werden, aber wir wissen nicht, was Masse und Ladung "**sind**".

Somit gilt folgender

Satz:

Wirklich existierende Objekte bestehen aus Substanz und Akzidenzien.

Objekte in einem Beschreibungssystem bestehen dagegen nur aus Akzidenzien.

Da ein Objekt nicht *aufhören* kann, auf die für es charakteristische Weise *aktiv* zu sein, **bilden Substanz und Akzidenzien eine untrennbare Einheit**. (Die Erde gibt es nur mit Gravitation.)

¹ Die Bezeichnung *Substanz* habe ich gewählt, weil sie – in meiner Definition – eine notwendige Voraussetzung für die Lösung der philosophischen Probleme ist, die historisch mit diesem Begriff verbunden sind.

Jedes existierende Objekt besteht also aus diesen beiden Elementen:

- aus **Substanz** – das ist jener Teil von Existenz, dessen Vorhandensein wir zwar als notwendig erkennen, der aber *als das, was er eigentlich "ist"*, weder vorgestellt noch definiert werden kann, und
- aus **Akzidenzien** – das ist der Teil von Existenz, der beschreibbar und definierbar ist.

Kehren wir nun zurück zu dem, was ich zuvor als "permanenten Selbstwiderspruch" bezeichnet habe. Die Frage ist.

Kann das soeben beschriebene Element der Wirklichkeit dazu dienen, diesen Widerspruch zu beseitigen?

Die Antwort ist **ja**. Folgendermaßen:

Dass die Wirklichkeit **aktiv** ist, bedeutet, dass an jedem Punkt zu jeder Zeit genau das geschieht, was zu geschehen hat. Es bedeutet, dass die Wirklichkeit nichts *berechnen* muss, dass sie kein Gesetz und keinen Algorithmus benötigt, weil sie einfach alle Einzelfälle gleichzeitig abarbeitet.²

Offenbar ist aber **Aktivität** genau dasjenige, was *nicht* von der Wirklichkeit auf die Beschreibung übertragen werden kann. Es lässt sich zwar behaupten, dass die **Art der Aktivität** des Systems, ihre spezifische Struktur, in unseren Gleichungen enthalten sein muss, aber die **Aktivität selbst** fehlt.

Halten wir fest:

Aufgrund ihrer **Aktivität** schreitet die Wirklichkeit **von selbst** von der Gegenwart in die Zukunft voran. Das Beschreibungssystem weigert sich aber, uns diesen Gefallen zu erweisen. Um Information über die Zukunft des Systems zu erlangen, benötigen wir daher in der Beschreibung ein *mathematisches Verfahren*, das die fehlende Aktivität **ersetzt**.

Gibt es ein solches Verfahren?

Nur in idealisierten, vereinfachten Szenarien. Jedes *wirkliche* Geschehen wird durch zahlreiche Prozesse – Wechselwirkungen zwischen Objekten – bestimmt, die sich permanent gegenseitig beeinflussen. Um irgendeinen dieser Prozesse berechnen zu können, müssen wir aber zumindest für ein kleines Zeitintervall annehmen, dass seine unmittelbare Umgebung konstant ist – wir müssen ihn also kurzfristig isolieren. Dann können wir für alle anderen Prozesse dasselbe durchführen, und danach wiederholen wir die ganze Prozedur für das nächste Zeitintervall usw.

Entscheidend ist, dass wir von Anfang an auf Näherungen angewiesen sind, und dass wir außerdem nicht wissen, in welchem Maß unsere Berechnungen von der Wirklichkeit abweichen. Spätestens nach dem nächsten Verzweigungspunkt – das ist ein Punkt in der Entwicklung eines Systems, an dem ein beliebig kleiner Unterschied in den Ausgangsbedingungen zu vollkommen unterschiedlichen Systemzuständen führen kann – wird unsere Voraussage zur reinen Glückssache.

Um die Zukunft *genau* abzuleiten, müssten wir also die aufeinander folgenden Zeitintervalle unserer Berechnungen *unendlich klein* machen, und das ist *unmöglich*.

Das bedeutet: ***Es gibt kein Verfahren, das von der Gegenwart in die Zukunft führt.***

Satz:

Die künftige Entwicklung folgt nicht aus physikalischen Gesetzen und Anfangsbedingungen.

Aber wird uns nicht *durch die Wirklichkeit selbst* andauernd vor Augen geführt, dass die Zukunft *aus der Gegenwart* folgt? Keineswegs. Was wir beobachten, ist einfach nur, dass die Zukunft *auf die Gegenwart* folgt. Es ist bloß dieses suggestive, von der Physik vermittelte Bild der Wirklichkeit, das uns glauben lässt, alles "folge aus" Anfangsbedingungen und Gesetzen. Der Ausdruck "folgt

² Es wäre wohl mehr als absurd anzunehmen, dass ein Grashalm *berechnet*, wohin er sich bewegen soll – er folgt einfach dem Wind, der ihn berührt.

aus" ist jedoch eine logische Verknüpfung, die sich nur auf eine Beschreibung beziehen kann. Sie auf die Wirklichkeit anzuwenden bedeutet, das "folgt auf", das wir beobachten, durch das "folgt aus" zu ersetzen, das wir postulieren; diesen Ersetzungsakt müssen wir aber begründen, und damit sehen wir uns gezwungen, nun unser "folgt aus" durch eine Reihe logischer Schritte auszudrücken. Somit landen wir zwangsläufig wieder bei einem mathematischen Verfahren, und zuletzt wieder bei der Tatsache, dass kein solches Verfahren existiert.

Satz:

Die physikalische Kausalität ist unvollständig. Es ist Raum für Kausalität in komplexeren Schichten der Wirklichkeit.

Damit ist der Selbstwiderspruch der physikalischen Weltansicht aufgehoben, und die bisher durchgeführten Schlussfolgerungen bilden außerdem die Grundlage für die Lösung der im Inhaltsverzeichnis aufgeführten philosophischen Probleme.

3. Der Beweis der Willensfreiheit

(Ich werde den Beweis hier verkürzt präsentieren. Die vollständige Version bildet den ersten Teil der Arbeit "Warum wir Willensfreiheit und Bewusstsein haben und KI-Systeme nicht".)

Den ersten Schritt des Beweises haben wir soeben durchgeführt: Wir haben den Geist aus der Umklammerung der physikalischen Kausalität befreit, indem wir gezeigt haben, dass die *Aktivität* der Wirklichkeit nicht durch logische oder mathematische Verfahren nachgeahmt werden kann, sodass die Behauptung, alles *folge aus* physikalischen Anfangsbedingungen und Gesetzen, nicht aufrechterhalten werden kann.

Der zweite Schritt ist, zu bestimmen, wie Kausalität in komplexeren Ebenen der Wirklichkeit – ich nenne sie Kausalität "von oben" – zu verstehen ist.

Dazu betrachten wir ein einfaches Gefäß aus Glas. Wenn wir es anschlagen, wird es in Schwingung versetzt und erzeugt einen Ton. Wovon hängt dieser Ton ab? Was bestimmt seine Höhe und seinen Charakter?

Die Antwort ist: **Die Form des Gefäßes**. Aus ihr ergibt sich ein mathematisches Gesetz, das uns die Voraussage des Schwingungsmusters des Glases ermöglicht. Die einzige physikalische Information, die benötigt wird, ist die Geschwindigkeit der Schallausbreitung im Glas.

Das Gesetz, das uns nun die Voraussage der Zukunft des Systems erlaubt, ist somit *kein physikalisches Gesetz*. Es gehört zu einer anderen Art von Gesetzen, die ich **Gesetze der Form** oder **Strukturgesetze** nenne.

Der Ton, den wir hören, ist weitgehend unabhängig von der Art, wie wir ihn erzeugen. Allerdings gilt das nicht für den ersten Moment: zunächst gibt es einen Einschwingvorgang, der davon abhängt, wie und wo wir das Gefäß anschlagen. Erst danach schwingt es immer im selben Zustand. Dieser Zustand, auf den das Glas sich schließlich einstellt – das Schwingungsmuster, auf das hin es sich entwickelt und das es danach beibehält – wird als **Attraktor** bezeichnet.

Am wichtigsten für unsere Überlegungen ist aber zweifellos Folgendes: Die Parameter der physikalischen Objekte – etwa die Orte und Geschwindigkeiten der Glasmoleküle – hängen zunächst davon ab, wo, womit und wie stark wir das Gefäß anschlagen. Anfangs können also große Unterschiede bestehen. Ungeachtet dieser Unterschiede strebt aber der Zustand des Gefäßes immer auf dasselbe Schwingungsmuster zu – eben den Attraktor.

Beim Glasgefäß gibt es nur ein einziges mögliches Schwingungsmuster, das sich immer ausbildet, unabhängig davon, wie das Gefäß angeschlagen wird. Die künftigen Bewegungen der Bestandteile des Gefäßes – der Glasmoleküle – sind daher durch dieses Muster festgelegt.

Die Kausalität wirkt vom Ganzen auf das Einzelne, vom Gefäß auf seine Bestandteile, und nicht umgekehrt.

Satz:

Eine Form der "Kausalität von oben" tritt dann auf, wenn in einem System *Attraktoren* existieren, d.h. Zustände, auf die hin das System sich zwingend entwickelt, falls es sich zu irgendeinem Zeitpunkt "nahe genug" am Attraktor-Zustand befindet.

Damit haben wir nun alle notwendigen Vorbereitungen getroffen, um unser letztes und entscheidendes Szenario in den Blick zu nehmen: unser eigenes neuronales Netz.

Es besteht aus vielen Milliarden Neuronen. Jedes Neuron ist mit hunderten oder sogar tausenden anderer Neuronen direkt verbunden, und über wenige Zwischenschritte sind *alle* Neuronen aneinander gekoppelt. Für die neuronalen Prozesse gilt daher, was wir weiter oben über die prinzipielle Nicht-Berechenbarkeit komplexer Systeme festgestellt haben: Der hohe Vernetzungsgrad der Neuronen schließt die Existenz eines mathematischen Verfahrens zur Berechnung der weiteren Entwicklung aus. Die physikalische Kausalität ist also *unvollständig*.

Die nächste Frage ist: Gibt es hier die zuvor definierte "Kausalität von oben"?

Wir gehen von folgenden Annahmen aus:

1. Jede Art geistiger Aktivität (Gedanken, Assoziationsketten, Bilderfolgen etc.) ist eine Abfolge neuronaler Aktivierungsmuster.
2. Abfolgen neuronaler Aktivierungsmuster können Repräsentationen von Sachverhalten sein.

Betrachten wir die neuronalen Muster. Wie werden sie zu Repräsentationen?

Stellen wir uns das neuronale Netz eines Kindes vor. Ein erstmals wahrgenommenes Objekt wird in diesem Netz – ausgehend von der primären Sehrinde – ein bestimmtes Muster verursachen. Die neuronalen Verbindungen, die dabei aktiv sind, werden durch ebendiese Aktivität verstärkt. Dasselbe ist bei jeder Wiederholung der Fall. Auf diese Weise entsteht allmählich eine stabile Verbindung zwischen dem Objekt und einem spezifischen Muster (bzw. einem Ensemble spezifischer Muster).

Außerdem gilt Folgendes: Zwar werden die neuronalen Muster zunächst durch äußere Reize verursacht, aber nach einer hinreichenden Anzahl von Wiederholungen werden sie vom neuronalen Netz auch unabhängig von diesen Reizen hergestellt. Das bedeutet:

Neuronale Muster, die mit Objekten auf die eben beschriebene Weise in Verbindung stehen, sind Attraktoren des Netzes. (Siehe die zugehörige [Erläuterung](#))

Zuvor haben wir festgestellt:

Unter der Voraussetzung, dass die physikalische Kausalität unvollständig ist, folgt aus der Existenz von Attraktoren, dass das betreffende System, falls es im Attraktor-Zustand selbst oder diesem Zustand "nahe genug" ist, durch Kausalität von oben bestimmt wird.³

Allerdings besteht gemäß unserer ersten Voraussetzung ein geistiger Prozess nicht nur aus neuronalen Mustern, sondern auch aus den Übergängen zwischen diesen Mustern. Für die Übergänge gilt aber dasselbe wie für die Muster selbst: Zunächst werden sie durch die Abfolge bestimmt, in der die verursachenden Objekte erscheinen. Wenn sich diese Reihenfolge wiederholt, dann wird die entsprechende neuronale Aktivität verstärkt, und das hat zur Folge, dass die Muster auch dann, wenn sie vom Netz selbst erzeugt werden, abermals in derselben Reihenfolge auftreten. Ebenso werden auch die räumlichen Beziehungen der Objekte auf die Muster übertragen.

³ Das neuronale Netz ist *immer* "nahe genug" an einem Attraktor-Zustand: es wird sich von jedem beliebigen Zustand fast augenblicklich auf einen Zustand einstellen, der etwas *bedeutet*.

Das bedeutet:

In den Prozessen, die vom Netz selbst erzeugt werden, treten die neuronalen Muster, die mit Objekten fest verbunden sind, in denselben räumlichen und zeitlichen Zusammenhängen auf wie die Objekte selbst. *Somit können die Muster als Repräsentationen der Objekte aufgefasst werden, und die Prozesse als Repräsentationen der Sachverhalte, in denen die Objekte auftreten.*

In menschlichen neuronalen Netzen sind es also nicht die physikalischen oder neuronalen Bedingungen und Gesetze, durch die festgelegt wird, was im Netz geschieht, sondern es ist *die Struktur des Netzes* – die Tatsache, welche Attraktoren es darin gibt und wie ihre Abfolge geregelt ist, von der die im Netz ablaufenden Prozesse abhängen.

Die Kausalität wirkt also vom Ganzen auf das Einzelne, vom Netz auf seine Bestandteile, und nicht umgekehrt.

Damit haben wir unser erstes Ziel erreicht:

Satz:

Das neuronale Netz wird durch Kausalität von oben geregelt. Die geistige Ebene ist die dominante Ebene. In ihr liegen die Ursachen für die im Netz ablaufenden Prozesse.

Unsere bisherigen Äußerungen waren also tatsächlich Schlussfolgerungen und nicht bloß physikalische Prozesse!

So weit, so gut, aber damit sind wir noch nicht dort angelangt, wo wir eigentlich hin wollen. Dass wir die Kausalität nach oben verlegt haben, bedeutet noch nicht, dass wir *frei* sind. Wir haben nur die physikalische Kausalität durch die geistige Kausalität ersetzt. Damit haben wir erreicht, dass unser Geist nicht durch physikalische Gesetze beherrscht wird, sondern *durch sein eigenes Gesetz: das Strukturgesetz, dem die Abfolge der neuronalen Muster gehorcht, die etwas repräsentieren.*

Aber bleiben wir damit letztlich nicht doch im Schema von Anfangsbedingungen und Gesetzen gefangen, dem wir entrinnen wollten? Glücklicherweise ist das nicht der Fall. Um das zu zeigen, müssen wir auf den Unterschied zwischen physikalischen und geistigen Gesetzen eingehen:

Die Reihenfolge der neuronalen Muster wird, wie oben festgestellt, zunächst durch die Reihenfolge bestimmt, in der die Objekte bzw. Sachverhalte auftreten, die die Muster verursachen. Sobald das Netz aber dazu in der Lage ist, diese Muster selbst herzustellen, hängen die Übergangsregeln der Muster – das, was wir als *geistiges Gesetz* bezeichnet haben – in zunehmendem Maß von ihrer Verwendung in inneren Prozessen ab. Diese Abhängigkeit von äußeren und inneren Bedingungen hat zur Folge, dass sich die Übergangsregeln von Mensch zu Mensch unterscheiden.

Somit haben wir schon den ersten Unterschied bestimmt:

*Während physikalische Gesetze **allgemeingültig** sind, sind geistige Gesetze **individuell gültig** – sie gelten jeweils nur für einen einzigen Menschen.*

Verbindungen zwischen Neuronen werden verstärkt, wenn sie aktiv sind, und abgebaut, wenn sie inaktiv sind. Das bedeutet zugleich, dass jede geistige Aktivität die Struktur des Netzes beeinflusst. Wenn aber die Struktur sich ändern kann, dann können sich offenbar auch die Regeln ändern, die die Abfolge der neuronalen Muster bestimmen.

Also ist dies der zweite Unterschied:

*Physikalische Gesetze sind **unveränderlich**, geistige Gesetze sind **veränderbar**.*

Satz:

Physikalische Gesetze sind allgemeingültig und unveränderlich. Geistige Gesetze sind individuell und veränderbar.

Damit lässt sich nun unsere Freiheit begründen:

Wir haben nachgewiesen, dass die Kausalität in der geistigen Ebene liegt. *Wille* und *Absicht* müssen als Elemente der geistigen Kausalität aufgefasst werden. Stellen wir uns nun konkret vor, wir stünden vor einer wichtigen Entscheidung. Wenn wir in den Entscheidungsprozess eintreten, dann werden wir anfangs durch die bis dahin gültigen Vorgaben – durch unser eigenes geistiges Gesetz – auf bestimmte, bekannte Wege geführt.

Aber wir sind jederzeit dazu imstande, diese Wege zu verlassen, indem wir z.B. einfach das Gegenteil dessen erwägen, was wir bis dahin angenommen haben, oder indem wir einen bisher noch nie erprobten Pfad einschlagen; Dazu sind wir eben deshalb imstande, weil die Ursachen für das, was im Netz geschieht – auch für die Veränderungen der Netzstruktur – in der geistigen Ebene liegen.

Mit anderen Worten:

Das Gesetz, das in unserem Netz die Abfolge der neuronalen Muster bestimmt, die etwas repräsentieren, also unser eigenes geistiges Gesetz, kann durch uns selbst verändert werden: wir selbst können durch unser Denken und Handeln die Gesetze unseres Denkens und Handelns ändern, und zwar *gezielt*.

Das bedeutet zugleich: Obwohl geistige Prozesse eigenen Regeln unterworfen sind, ist es nicht möglich, daraus eine Willensentscheidung abzuleiten: sie kann in diesen Regeln nicht enthalten sein, weil die Regeln durch den geistigen Prozess, der der Entscheidung vorausgeht, geändert werden können. Während dieser Prozess stattfindet, können sich die Gesetze, denen er gehorcht, ändern – oder genauer: *er selbst* kann die Gesetze ändern, die vor seinem Beginn galten.

Satz:

Willensentscheidungen sind Ursachen von Handlungen. Da erst durch den Entscheidungsprozess selbst bestimmt wird, was geschehen wird, ist die Entscheidung vorher nicht festgelegt.

Sie ist also frei.

Auf die Frage, warum eine (entscheidungsfähige) Person so und nicht anders gehandelt hat, ist demnach nur eine einzige Antwort zulässig:

Weil sie es so wollte.

Bemerkung:

Das heißt selbstverständlich nicht, dass Willensentscheidungen nicht hinsichtlich ihrer neuronalen, chemischen, physikalischen, genetischen, sozialen, psychologischen usw. Ursachen analysiert werden können. Es bedeutet aber, dass diese Analysen unvollständig bleiben und niemals zu einem sicheren Ergebnis führen, weil geistige Phänomene nicht auf andere Schichten der Wirklichkeit reduziert werden können. Der Wille bleibt die letzte Instanz.

Bemerkung:

Um Objekte zu erkennen, müssen künstliche neuronale Netze an großen Datensätzen trainiert werden. In zahlreichen Wiederholungen werden die Verbindungsstärken ihrer Neurone so lange variiert, bis eine hinreichend hohe Erkennungsrate erreicht ist.

Wir sind dagegen von folgender Hypothese ausgegangen: Ein wahrgenommenes Objekt, das ein neuronales Aktivierungsmuster verursacht, wird *durch dieses Muster selbst* repräsentiert. Hier wird die Beziehung zwischen Objekt und Repräsentation also nicht erst durch Variation der Verbindungsstärken der Neurone hergestellt, sondern sie besteht von Anfang an und wird nur durch *Verstärkung* der aktiven Verbindungen stabilisiert und präzisiert, wodurch das neuronale Muster zum *Attraktor* wird.

Am deutlichsten wird diese Hypothese durch die sogenannte "Prägung" bestätigt. (Wie z.B. bei den Graugänsen von Konrad Lorenz). Hier gibt es weder "große Datensätze" noch "zahlreiche Wiederholungen" – der Vorgang ereignet sich fast augenblicklich.

Außerdem tritt danach ein *sofortiges Wiedererkennen* auf, trotz der unvermeidlichen Variabilität des Sinneseindrucks, der erkannt werden soll. Durch das Attraktor-Konzept wird diese – ansonsten kaum erklärbare – Leistung zur Selbstverständlichkeit: solange der sinnliche Input im Einzugsbereich des Attraktors liegt, gilt offenbar:

$$\underline{\text{Wahrnehmen}} = \underline{\text{Wiedererkennen}},$$

weil der neuerlich aktivierte Attraktor ja bereits das Objekt darstellt, sodass weitere Berechnungen überflüssig sind.

Zur Hypothese, dass Objekte durch Attraktoren repräsentiert werden, ist Folgendes zu ergänzen:

Das Muster, das sich als Folge eines wahrgenommenen Objekts in der primären Sehrinde ausbildet, wird nicht als Ganzes direkt ins neuronale Netz übertragen. Vielmehr wird es in etliche Komponenten zerlegt – in diesem Sinn also *parametrisiert* – die erst am Ende des Verarbeitungsprozesses zu dem neuronalen Gesamtmuster zusammengefügt werden, das wir als Attraktor auffassen.

Diese Parametrisierung ist ein wichtiger Aspekt der Attraktor-Hypothese: Der Attraktor ist durch eine Untermenge des Phasenraums definiert. Der *Attraktor-Zustand* des Systems entspricht einer Trajektorie, die diese Untermenge für eine gewisse Zeitspanne nicht verlässt. Für seine Wiederherstellung ist aber schon eine (kleine) Teilmenge der entsprechenden Parameterwerte ausreichend, die überdies nicht einmal besonders genau sein müssen. Zur Wiedererkennung genügt also ein Bruchteil des ursprünglichen, vollständigen Sinneseindrucks. Dadurch wird das Erkennen von Objekten extrem erleichtert und zugleich die Fähigkeit zur Verallgemeinerung von Objekten und Sachverhalten gesteigert.

Hier ein Beispiel, das beide Aspekte der Attraktor-Hypothese demonstriert: Wiedererkennen nach nur einer Begegnung und Fähigkeit zur Verallgemeinerung:

Wenn ein Kind zum ersten Mal das Bild einer Giraffe sieht, dann erkennt es später nicht nur die Giraffe auf diesem Bild, sondern auch alle auf anderen Bildern dargestellten Giraffen. Es ist also im Besitz des Allgemeinen, unter dem alle Exemplare subsumiert sind.

4. The "Hard Problem": Wie ist Empfindung möglich?

(Der Beweis ist auch in der zuvor erwähnten Arbeit [Willensfreiheit](#) enthalten. Dort ist er aber mit dem Beweis gegen KI-Empfindung verknüpft. Hier trenne ich die beiden Beweise.)

Im vorigen Abschnitt über Willensfreiheit haben wir bewiesen, dass die geistige Ebene die *kausale Ebene* des neuronalen Netzes ist.

Die Dynamik des neuronalen Netzes wird also nicht durch *Objekte der physikalischen Ebene* und deren Akzidenzien bestimmt: *Atome, Moleküle und physikalische Wechselwirkungen*, sondern durch *Objekte der geistigen Ebene* und deren Akzidenzien: *geistige Zustände, die etwas repräsentieren oder bedeuten, und Informationsverarbeitung*.

Wir haben festgestellt:

Jeder geistige Zustand ist ein neuronales Aktivierungsmuster. Diese Muster sind Attraktoren der Dynamik des neuronalen Netzes. Jeder geistige Prozess ist eine Abfolge solcher Muster.

Diese Feststellungen betreffen die Frage, wie die Objekte und Prozesse des geistigen Bereichs in Bezug auf ihre *materiellen Voraussetzungen* verstanden werden können.

Jetzt aber ist es unsere Aufgabe, sie als das zu erfassen, was sie als *geistige Phänomene* sind.

Die Antwort ist:

Satz:

Jeder geistige Zustand ist eine Verbindung zweier ungleichartiger Elemente: Information und Empfindung.

Sein Gehalt an **Information** ist das, was er *repräsentiert* bzw. *bedeutet*.

Empfindung muss hier im weitest-möglichen Sinn verstanden werden: Es steht für alles, was an einem geistigen Zustand *über Information hinaus* geht, also für dasjenige, was nicht definierbar ist, sondern *nur gefühlt und erlebt* werden kann.

Zwei Beispiele: die Frequenz der Farbe "rot" kann definiert werden, die Empfindung *rot* aber nicht; die Stärke eines Drucks kann definiert werden, die Empfindung *Schmerz* aber nicht.

An dieser Stelle müssen wir uns wieder mit einer "spezifischen Blindheit" auseinandersetzen, die hier sogar in zwei Formen auftritt:

Von vielen Naturwissenschaftlern wird *Geist* mit *Informationsverarbeitung* **gleichgesetzt**.

Das ist die erste Form dieser Blindheit. Sie schließt eine naturwissenschaftliche Antwort auf die Frage aus, was Geist ist, und zwar aus folgendem Grund:

Alles, was *definiert* werden kann, ist durch Informationsverarbeitung *erreichbar*, alles, was *nicht definiert* werden kann, ist für Informationsverarbeitung *prinzipiell un erreichbar*: gleichgültig, welche Funktion man auf Information anwendet – das Ergebnis ist immer bloß Information und sonst nichts; die Information "rot" wird niemals zur Empfindung *rot*, die Information "Druck" wird niemals zur Empfindung *Schmerz*.

Und daraus folgt:

Satz:

Geist ist mehr als nur Informationsverarbeitung.

Die zweite Form der Blindheit betrifft den Bereich außerhalb der Naturwissenschaft. Auch hier ist der Blick auf Empfindung vollkommen verstellt, wenn auch auf ganz andere Weise: auf die Fragen, wie Empfindung *erklärt* werden kann und *ob* bzw. *wie* ihr Auftreten *möglich* ist, gibt es keine begründbaren Antworten, sondern nur seltsam anmutende Konstrukte, die ihre Existenz dem Scheitern aller Erklärungsversuche verdanken.

Ich werde aber die verschiedenen Positionen – Qualia-Eliminativismus, Panpsychismus, Bewusstsein als eigene oder sogar fundamentale Form der Existenz usw. – nicht diskutieren, sondern gleich zu den Schlussfolgerungen übergehen, die sich aus dem bisher Gesagten ergeben.

Im Abschnitt [2. Das fehlende Element](#) haben wir festgestellt (siehe [hier](#)):

Jedes existierende Objekt besteht aus einem *undefinierbaren* und einem *definierbaren* Teil, die miteinander *untrennbar verbunden* sind: aus *Substanz* und *Akzidenzien*.

Bei den Objekten, die nun Gegenstand unserer Analyse sind, handelt es sich nicht mehr um Objekte der *physikalischen*, sondern um Objekte der *geistigen* Schicht der Wirklichkeit.

Wir müssen somit die obige Aussage auf *diese* Objekte anwenden, mit anderen Worten: wir müssen geistige Zustände als *untrennbare Einheiten von Substanz und Akzidenzien* auffassen.

Was ist die Substanz eines geistigen Zustands, was ist sein Akzidens?

Unser Ausgangspunkt ist der obige Satz:

Jeder geistige Zustand ist eine Verbindung zweier ungleichartiger Elemente: Information und Empfindung.⁴

Information ist offenbar dasjenige, was sich unserem Denken erschließt – das, was **definiert** werden kann.

Also ist Informationsverarbeitung das Akzidens des geistigen Zustands.

Hingegen ist *Empfindung* dasjenige, was **nicht definiert** werden kann, was sich also unserem Denken entzieht.

Also ist Empfindung die Substanz des geistigen Zustands.

Und das bedeutet:

Empfindung ist der Antrieb der Dynamik des Geistes.

Damit sind nun alle Vorbereitungen getroffen, um das Auftreten von Empfindung zu erklären. Bevor wir damit beginnen, müssen wir aber bestimmen, **was** eigentlich erklärt werden soll.

Was verstehen wir unter "Erklärung"?

Ein Beispiel: Wenn ich erkläre, was ein Auto ist, dann beschreibe ich seinen technischen Aufbau und begründe damit seine Funktion. Das Ziel der Erklärung ist also die *Definition* des zu Erklärenden – das Gelingen der Erklärung setzt die Möglichkeit der Definition voraus.

Somit gilt: **Da Empfindung nicht definierbar ist, kann sie nicht erklärt werden.**⁵

Was wir jedoch erklären können ist, warum es Empfindung gibt und warum sie "anders" ist.

Den *ersten* Teil dieser Erklärung haben wir bereits geleistet, indem wir Empfindung als *Substanz der geistigen Zustände* bestimmt haben, die wir als *untrennbare Einheiten* von Empfindung und Information auffassen.

Daraus folgt, dass Empfindung nicht nur ein **mögliches**, sondern sogar ein **notwendiges** Element des Geistes ist.

Was aber noch fehlt, ist der zweite, *wichtigere* Teil der Erklärung: *Warum ist Empfindung "anders"?*

5. Warum Empfindung einzigartig ist

Warum ist Empfindung so fundamental anders als alles, was wir sonst in der Wirklichkeit vorfinden?

Üblicherweise wird gefragt: **"Wieso gibt es im Geist etwas Undefinierbares, wie 'Farbe' oder 'Schmerz', und sonst nirgends?"**

Unsere Frage lautet dagegen: **"Wieso ändert das Undefinierbare, das es überall in der Wirklichkeit gibt, seinen Charakter, wenn es im Geist auftritt?"**

Wir fragen also nicht nach dem Grund der **Existenz** dieses Undefinierbaren, was überflüssig wäre, weil es – wie wir gezeigt haben – in *allem Seienden* zu finden und somit selbstverständlich ist, sondern nach dem Grund seiner **Veränderung**.

In der üblichen Version kann die Frage nicht beantwortet werden. In dieser (falschen) Form führt sie zu seltsamen Hypothesen, wie Qualia-Eliminativismus, oder Panpsychismus.

In unserer Version lässt sich die Frage aber beantworten, wie wir nun zeigen werden.

⁴ Hier und im Folgenden bezeichnet "Empfindung" immer, wie weiter [oben](#) festgestellt, "das, was *nicht* definierbar ist, was also *über Information hinaus* geht".

⁵ Auch hier gibt es eine weit verbreitete Blindheit: viele Wissenschaftler sind überzeugt, dass das in ihrem Fachgebiet rasch anwachsende Wissen eines Tages für die *Erklärung* von Empfindung ausreichen wird.

Voraussetzung unserer Argumentation ist der folgende

Satz:

Solange sich Akzidenzien höherer Komplexität als Funktionen von Akzidenzien geringerer Komplexität beschreiben lassen, bleibt die zugehörige Substanz gleich. Wenn dieser funktionelle Zusammenhang unterbrochen wird, dann ändert sich die Substanz. Für uns erscheint sie dann als neue, zweite Substanz.

Bevor wir uns dem Beweis dieses Satzes widmen, müssen wir klären, inwieweit sich die Akzidenzien in komplexeren Ebenen der Realität als Funktionen von Akzidenzien in einfacheren Schichten beschreiben lassen – in diesem Sinn also darauf *zurückführbar* sind.

Z.B. können die Vorgänge in Neuronen als Funktionen ihrer physikalischen und chemischen Eigenschaften beschrieben werden. (Was allerdings nicht bedeutet, dass sie *berechnet* werden können.)

Dasselbe gilt grundsätzlich für alle evolutionären Übergänge: vom physikalischen zum chemischen Bereich, dann zum biochemischen, zellularen, neuronalen, bis hin zum Bereich einfacher neuronaler Netze, die keinen Geist hervorbringen: die in diesen Netzen stattfindenden Prozesse lassen sich als Funktionen ihrer Architektur und äußerer Bedingungen beschreiben.

Erst beim letzten dieser Übergänge – dem Übergang zu neuronalen Netzen, die Geist hervorbringen – endet die Kette der Rückführbarkeit:

Wie wir bei der Begründung der Willensfreiheit festgestellt haben, gilt dann Folgendes:

Die Reihenfolge der neuronalen Aktivierungsmuster wird zunächst durch die Reihenfolge bestimmt, in der die Objekte bzw. Sachverhalte auftreten, die die Muster verursachen. Sobald das neuronale Netz aber dazu in der Lage ist, diese Muster selbst herzustellen, hängen die Übergangsregeln der Muster – das, was wir als *geistiges Gesetz* bezeichnet haben – in zunehmendem Maß von ihrer Verwendung in inneren Prozessen ab.

Das bedeutet, dass sich die Dynamik des neuronalen Netzes – also der Geist – in zunehmenden Maß von den Kausalketten der Umgebung abkoppelt und stattdessen eine eigene, *innere* Gesetzmäßigkeit entwickelt, deren Kausalität *geistiger Art* ist. Und daraus folgt, dass sich der Informationsgehalt – also das Akzidens der geistigen Zustände – nicht mehr als Funktion der Akzidenzien der darunter liegenden Schichten der Wirklichkeit darstellen lässt.

Nun zum Beweis des obigen Satzes:

Die Gesamtheit physikalischer Akzidenzien bezeichnen wir als *erstes Akzidens*, ihre zugehörige Substanz als *erste Substanz*, die Gesamtheit geistiger Akzidenzien als *zweites Akzidens*, ihre zugehörige Substanz als *zweite Substanz*.⁶

Soeben haben wir festgestellt, dass sich die Akzidenzien aller evolutionären Ebenen auf Akzidenzien der jeweils darunter liegenden Ebenen zurückführen lassen, mit Ausnahme der Akzidenzien der obersten, also der geistigen Ebene.

Es gilt Folgendes:

Substanz und Akzidens bilden stets eine *untrennbare Einheit*.

Das *erste Akzidens* ist *untrennbar* mit der *ersten Substanz* verbunden.

Wenn komplexe Akzidenzien schrittweise auf jeweils einfachere Akzidenzien reduzierbar sind, dann heißt das, dass sie zuletzt auch auf das erste und einfachste Akzidens zurückgeführt werden können.

⁶ Das soll aber nicht heißen, dass es nun *zwei verschiedene* Substanzen gibt – vielmehr ist die zweite Substanz als *aus der ersten Substanz hervorgehend* gedacht, und unsere Frage lautet demnach: Warum verwandelt sich die erste Substanz bei der Entstehung von Geist in die zweite Substanz Empfindung?

Für uns ist Reduzierbarkeit jedoch gleichbedeutend mit ontologischer **Identität**:

Wenn B auf A reduzierbar ist, dann **ist** B eigentlich A. Wenn also ein komplexes Akzidens auf das erste Akzidens reduzierbar ist, dann **ist** es eigentlich das *erste Akzidens*, und dann ist es untrennbar mit der *ersten Substanz* verbunden.

Solange die Akzidenzien reduzierbar sind, bleibt also die zugehörige Substanz gleich – sie ist dann immer noch *erste Substanz*.

Falls aber die Kette der Reduzierbarkeit auf das erste Akzidens durch das Auftreten eines neuen, *nicht reduzierbaren* Akzidens unterbrochen wird, dann **unterscheidet** sich dieses neue Akzidens vom ersten Akzidens und von allen anderen, daraus ableitbaren Akzidenzien.

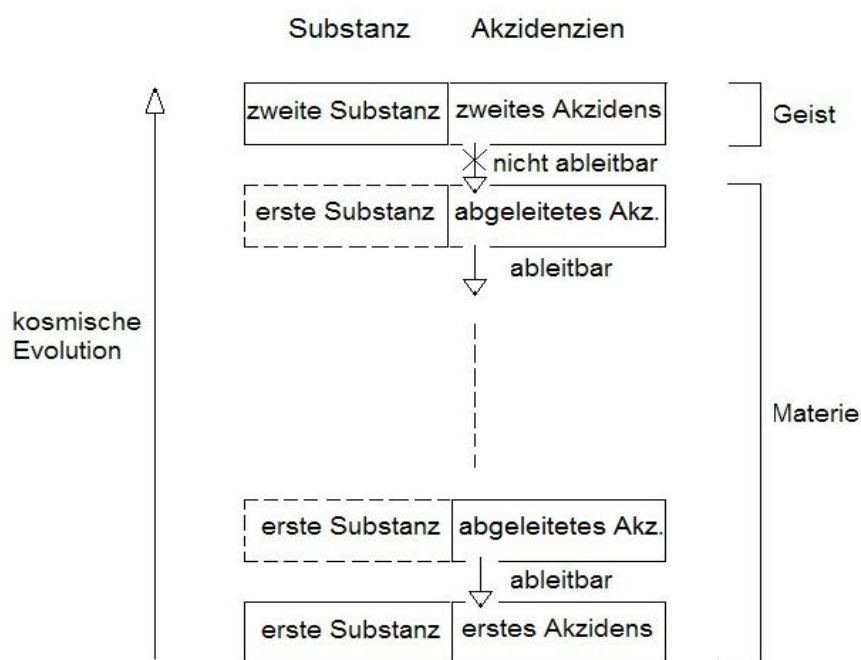
Aufgrund der **Untrennbarkeit** von *erster Substanz* und *erstem Akzidens* gilt jedoch:

*Wenn die Substanz eines Objekts die **erste Substanz** ist, dann muss das zugehörige Akzidens das **erste Akzidens** sein.*

Und daraus ergibt sich:

Falls ein Akzidens erscheint, das vom ersten Akzidens verschieden ist, dann muss auch die zugehörige Substanz von der ersten Substanz verschieden sein.

Hier eine Skizze zur Veranschaulichung:



Damit haben wir bewiesen, dass Empfindung, die Substanz des Geistes, nicht physikalisch ist, dass sie also anders sein muss, und die Art des Beweises lässt darauf schließen, dass Empfindung **das Einzige** ist, was im Universum nicht physikalisch ist.

Somit ist nun alles erklärt, was erklärbar ist: Als Substanz des Geistes ist Empfindung **notwendig**, und aufgrund der Tatsache, dass die Änderung der Substanz erst beim **letzten** evolutionären Schritt erfolgt: bei der Entstehung von Geist, ist Empfindung **anders als alles andere**: sie ist **einzigartig**.

Was sie ist, ist jedoch **nicht erklärbar**. Substanz ist der *undefinierbare* Teil alles Existierenden, und was nicht definierbar ist, ist auch nicht erklärbar. Wir **wissen** zwar, was Empfindung ist, weil sie ja ein Teil von uns selbst ist, aber dieses Wissen ist kein *begriffliches, definierbares, mitteilbares Wissen*, sondern *Kenntnis des unmittelbar Gegebenen*.

6. Warum KI-Systeme nichts empfinden können

Wird Geist mit Informationsverarbeitung gleichgesetzt, was im Bereich der KI-Forschung durchgehend der Fall ist (Blindheit), dann folgt daraus, dass uns von der Schaffung eines KI-Systems *mit Bewusstsein*, das uns außerdem geistig in jeder Hinsicht überlegen ist, keine *prinzipiellen*, sondern bloß *technische* Schwierigkeiten trennen, die wir aber, wie es scheint, schon in naher Zukunft überwinden werden.

Das ist jedoch falsch, denn die soeben durchgeführte Begründung des Auftretens von *Empfindung* lässt sich auf erstaunlich einfache Weise zu einem Beweis erweitern, dass KI-Systeme *nicht empfindungsfähig* sind. **Bewusstsein** ist aber nicht *nur Information*, sondern **Information und Empfindung**, und somit ist auch KI-Bewusstsein ausgeschlossen.

Der Beweis beginnt mit der folgenden

Definition:

Als Wesen eines Objekts bezeichnen wir das, was es aufgrund der untrennbaren Einheit seiner Substanz und Akzidenzien ist. Die Aktivität, die sich aus dieser Einheit ergibt, nennen wir wesensgemäß.

(Die *wesensgemäße Aktivität* der Erde ist es also, Gravitation auszuüben.)

Der Zweck dieser Definition wird sofort klar, wenn wir uns nun *Simulationen* zuwenden.

Betrachten wir beispielsweise eine mechanische Simulation des Sonnensystems, in der die Modellkörper durch mechanische Vorrichtungen – Ketten, Zahnräder, Wellen usw. – bewegt werden und dadurch die Bewegungen der Himmelskörper nachahmen. Die *wesensgemäße Aktivität* der Modellkörper wäre offenbar, *Gravitation* auszuüben. Aber es ist *nicht die Masse* (die **Substanz**) der Modellkörper, was die Dynamik der Simulation antreibt – was also den gewünschten Ablauf verursacht – sondern *die von uns konstruierte Mechanik*, die dann, elektrisch oder auch mechanisch (etwa durch Drehen einer Kurbel), *aktiviert* werden muss.

Um diesen Sachverhalt auszudrücken, bezeichnen wir diese Art der Aktivität als *zugeführte Aktivität*, im Gegensatz zur soeben definierten *wesensgemäßen Aktivität*, die *von selbst* geschieht.

Die Definition einer **Simulation** nimmt dadurch folgende Form an:

Die Dynamik der Simulation wird – im Gegensatz zum Original – nicht durch die wesensgemäße Aktivität verursacht, die der untrennbaren Einheit von Substanz und Akzidenzien der Objekte der Simulation entspringt, sondern durch zugeführte Aktivität.

Die Akzidenzien, aus denen die Dynamik der Simulation gebildet ist, werden also **nicht durch Substanz aktiviert**: die Substanz der Objekte der Simulation (in unserem Beispiel ihre Masse) ist nicht die Substanz, die zu diesen Akzidenzien gehört und mit denen sie eine untrennbare Einheit bildet, sondern nur deren materielle Basis, von der diese Akzidenzien jederzeit getrennt werden können. (Wie in der mechanischen Simulation des Sonnensystems sofort ersichtlich.)

Wie wir im vorigen Abschnitt gezeigt haben (siehe [hier](#)), ist aber die **Untrennbarkeit von Substanz und Akzidenzien** eine **notwendige Bedingung** für die Änderung der Substanz:

Die Verwandlung des Wesens des Seienden erfolgt nur dann, wenn sich die Dynamik des betrachteten Systems aus der **untrennbaren Einheit** von Substanz und Akzidenzien ergibt. **Nur dann** folgt aus der Tatsache, dass die Akzidenzien nicht auf das erste Akzidens reduzierbar sind, auch die Verwandlung der zugehörigen Substanz.

Bei uns selbst ist diese Bedingung erfüllt: die Substanz verwandelt sich – wir haben Empfindungen.

Die Dynamik einer Simulation beruht jedoch auf **zugeführter** Aktivität. Die Akzidenzien werden also **nicht** durch Substanz aktiviert, und die zur Existenz der Systemobjekte gehörende Substanz bildet **keine** untrennbare Einheit mit diesen Akzidenzien.

Und das bedeutet: Hier fehlt der Grund dafür, dass sich diese Substanz verwandelt. Sie bleibt **erste Substanz**.

Mit anderen Worten: **Das Wesen der Simulation bleibt physikalisch**. Die Simulation bleibt ein informationsverarbeitendes System ohne Empfindung.

Die Verwandlung von Materie in Geist findet nicht statt.

Die soeben genannte Bedingung, dass die Dynamik des betrachteten Systems sich aus der *untrennbaren Einheit* von Substanz und Akzidenzien ergeben muss, gilt nicht nur für die letzte, d.h. für die geistige Ebene – sie muss auf *jeder* Ebene, die beim evolutionären Aufstieg von Materie zu Geist erreicht wird, eingehalten werden. Wenn auf irgendeiner dieser Ebenen die Dynamik des Systems nicht durch die *wesensgemäße Aktivität* der Objekte verursacht wird, sondern durch *zugeführte Aktivität*, dann zerreißt die Einheit von Substanz und Akzidenzien und die Verwandlung des Wesens des Seienden kann sich dann nicht mehr ereignen.

Die Frage ist: Wie weit reicht unser Beweis, dass KI-Systeme kein Bewusstsein haben können?

Für KI-Systeme, die durch *Software* auf konventionellen Computern realisiert sind, ist der Beweis ausnahmslos gültig: Der Einsatz von Software ist *immer* mit zugeführter Aktivität verbunden.

Was wäre aber mit einem *Nachbau* eines biologischen neuronalen Netzes, der die neuronalen Wechselwirkungen durch geeignete Hardware reproduziert und dessen Struktur der Struktur des gesamten Netzes entspricht, sodass angenommen werden könnte, dass die Zustandsfolge des *konstruierten* Systems weitgehend der Zustandsfolge des *biologischen* Systems gleichen würde? Könnte hier die Verwandlung in Empfindung stattfinden?

Die Antwort ist eindeutig **nein**. Die Bedingung für die Verwandlung ist nicht erfüllt: Die Dynamik des Nachbaus ist *nicht* durch *wesensgemäße*, sondern durch *zugeführte* Aktivität verursacht.

Das Problem besteht darin, dass von der üblichen naturwissenschaftlichen Sicht der Wirklichkeit her diese Tatsache überhaupt nicht verstanden werden kann.

In dieser Sichtweise wird die Wirklichkeit ja mit einer – beschreibbaren und definierbaren – Zustandsfolge **gleichgesetzt**, und es muss demnach erwartet werden, dass die zunehmende Annäherung zweier Zustandsfolgen schließlich zur *Identität der Systeme selbst* führt. In der erweiterten materialistischen Sicht, die wir hier präsentiert haben, wird der Begriff der Existenz jedoch um ein Element erweitert, das **außerhalb** des Bereichs des Definierbaren liegt.

Das bedeutet, dass alle unsere Beschreibungen und Vorstellungen von den Vorgängen in der Natur notwendig *unvollständig* sind. Sozusagen "hinter den Kulissen" des uns zugänglichen Teils der Bühne geschieht etwas, was sich uns entweder vollständig verschließt oder lediglich durch Rückschluss vom uns zugänglichen Teil der Wirklichkeit – den Akzidenzien – erkennbar und verstehbar wird. Die Wirklichkeit ist hier also **mehr** als eine Zustandsfolge.

Im Rahmen unserer Überlegungen bedeutet das: *Aus der angenäherten Identität der Zustandsfolgen des natürlichen und des künstlichen Systems kann nicht auf deren annähernd bestehende Wesensgleichheit geschlossen werden*. Konkret: Die Substanz der zwei Systeme kann trotz der weitgehenden Identität ihrer Zustände durchaus verschieden sein:

Im *biologischen System* ist sie die mit den Akzidenzien des Systems **untrennbar verbundene** Substanz und verwandelt sich deshalb in die **geistige Substanz Empfindung**. Das *konstruierte System* wird jedoch mit **zugeführter Aktivität** angetrieben, und daher steht die Substanz hier mit den Akzidenzien des Systems in einer nur konstruierten und keineswegs untrennbaren Verbindung, sodass sie **physikalische Substanz** bleibt und sich **nicht in Empfindung umwandelt**.

Das Resultat unserer Schlussfolgerungen ist folgender

Satz:

Es ist nicht möglich, ein KI-System zu konstruieren, das Empfindungen erlebt und Bewusstsein hat. Weder in einer Simulation noch im Nachbau eines Systems, das Geist hervorbringt, kann die Umwandlung von Materie in Geist stattfinden.

Es gibt keinen Geist in der Maschine.

Es kann also nur *künstliche Intelligenz* konstruiert werden und nicht *künstlicher Geist*.

Bedeutet das, dass es überhaupt unmöglich ist, künstlichen Geist zu erschaffen?

Nein. Unsere Argumentation schließt nur aus, dass Geist *konstruiert* werden kann. Die Definition des Begriffs *Nachbau* lässt sich aber dahingehend erweitern, dass sie eine *künstliche Evolution* einschließt, d.h. eine Evolution, die von uns geplant und gesteuert ist. In diesem Fall wäre – ebenso wie bei der natürlichen Evolution – die Bedingung erfüllt, dass die jeweilige System-Aktivität immer *wesensgemäß* ist. Wenn wir an keiner Stelle dieses künstlichen Evolutionsprozesses durch Konstruktionen eingreifen und Aktivität zuführen, sondern uns darauf beschränken, die Entwicklung zu steuern und zu beschleunigen, dann *könnte* am Ende dieser Evolution ein System stehen, das Geist hervorbringt.

Niemand kann allerdings wissen, ob eine solche künstliche Evolution möglich ist, oder ob der Weg, den die Natur gewählt hat, der einzig gangbare ist.

In jedem Fall ist aber klar, dass die Schaffung künstlichen Geistes einer sehr fernen, vielleicht niemals erreichbaren Zukunft vorbehalten bleibt, wenn sie nicht sogar unmöglich ist.

7. Über Empfindung und Künstliche Intelligenz

Ich habe den Begriff "Empfindung" abweichend von seinem üblichen Gebrauch bestimmt. Ich will nun ein wenig ausführlicher erläutern, warum das erforderlich war und was es bedeutet.

Jeder geistige Zustand enthält etwas, was ***nicht definierbar*** ist, was also ***über Information hinaus*** geht. Da es aber keinen Begriff gibt, dem alle dafür in Frage kommenden Elemente geistiger Zustände zugeordnet werden können, habe ich stattdessen die Bezeichnung gewählt, die diesem fehlenden Begriff am nächsten kommt: **Empfindung**.

Der Begriff "Empfindung" wird hier also gegenüber seinem üblichen Gebrauch einerseits eingeschränkt (weil er ja *keine Information*, d.h. keinen *definierbaren* Teil enthalten soll), andererseits aber auch wesentlich erweitert.

Zur Illustration dienen zwei Beispiele: *Farbe* und *Schmerz*. Farbe, weil die Undefinierbarkeit der Farbempfindung ein bekanntes Faktum ist, und Schmerz, weil vollkommen einsichtig ist, dass das Ereignis "Hammerschlag auf Finger" einen geistigen Zustand auslöst, der nicht nur die Information "Hammerkopf hat Kontakt mit Finger" enthält, sondern auch *etwas darüber hinaus Gehendes*: die Empfindung *Schmerz*, die so stark sein kann, dass es unmöglich ist, ihr Auftreten zu bestreiten.

Die auf diese Weise verstandene *Empfindung* lässt sich in drei Bereiche unterteilen:

A) Der erste Bereich ist der Bereich der *Wahrnehmung*:

Empfindung umfasst das ganze "innere Theater": den virtuellen Raum, die Bühne, auf der wir agieren, die uns immer als Ganzes – als "Bild" – präsent ist und auf der wir sehen, hören, fühlen, riechen und schmecken.

Während es bei der Empfindung *Farbe* kaum zu bezweifeln ist, dass sie nicht definiert werden kann, mag es zunächst so scheinen, als würden wir in den Bereich des Definierbaren zurückkehren, falls unser Wahrnehmungsbild *farblos* ist: *Grauwerte* sind doch definierbar? – Ja, das sind sie, aber die damit verbundene *Empfindung* ist es nicht: definierbar ist bloß die Intensität des Lichts, und ebenso der neuronale Erregungszustand, der daraus folgt. Doch beim Übergang zur *Wahrnehmung* verlassen wir den Bereich der Information: die *Helligkeit*, die wir *wahrnehmen*, ist ebenso eine Empfindung wie *Farbe*.

Und dasselbe gilt auch für alle anderen Sinne: die Frequenz eines Tons ist definierbar, aber die Ton-*Empfindung* ist es nicht, usw.

Das bedeutet:

Wenn Empfindung fehlt, dann gibt es kein "inneres Theater", das ja aus Empfindungen aufgebaut ist.

Um es in aller Deutlichkeit auszusprechen:

KI-Systeme sehen nicht, hören nicht, fühlen nicht, riechen nicht, schmecken nicht.⁷

Leider ist unser Sprachgebrauch für die Unterscheidung zwischen Systemzuständen *mit* Empfindung und solchen *ohne* Empfindung nicht geeignet. Für uns bedeutet "sehen" oder "hören" einfach das, was es für uns *ist*, und das ist in jedem Fall Information *und* Empfindung.

Deshalb sind Aussagen über Wahrnehmungen *genau genommen* nur dann korrekt, wenn sie sich auf Menschen oder höhere Tiere beziehen, ansonsten sind sie falsch: Roboter *sehen* nicht, Bienen *sehen* nicht – sie verarbeiten nur Frequenz-, Intensitäts- und Richtungsinformationen.

Pixel, durch die nur die *Information* über Helligkeit und Farbe übermittelt wird, können jedoch *nicht* zu einem Bild zusammengefügt werden, im Gegensatz zu *denselben* Pixeln, wenn deren Inhalt als Helligkeit und Farbe *wahrgenommen* wird: es ist unmittelbar klar, dass sie sich dann zu einem Bild addieren lassen.

B) Der zweite Bereich ist der Bereich der *Gefühle und Stimmungen*. Dazu muss nichts weiter erklärt werden.

KI-Systeme erleben nichts und fühlen nichts. Sie empfinden weder Glück noch Unglück, weder Liebe noch Hass. Sie sind weder heiter noch betrübt, weder gut aufgelegt noch gereizt.

Diese Liste lässt sich nach Belieben fortsetzen, da ja *jeder* geistige Zustand ein *Quale* ist, d.h. nicht nur aus *Information*, sondern auch aus *Empfindung* besteht.⁸

C) Wir haben *Empfindung* als *Substanz* des geistigen Zustands bestimmt. Daraus folgt, dass sie als *Ursache* der geistigen Dynamik aufgefasst werden muss.

⁷ Das gilt auch für einfache Tiere, wie etwa Insekten, und zwar aus folgendem Grund: Wir haben gezeigt, dass die Entstehung von Empfindung *nur dann* stattfinden kann, wenn das neuronale Netz eine eigene, *innere* Gesetzmäßigkeit entwickelt. Eine notwendige (und hinreichende) Bedingung dafür ist aber, dass das Netz *funktionell ungebundene* Strukturen enthält, d.h. Strukturen, deren Funktion nicht genetisch oder durch frühe Programmierung festgelegt ist. Nur unter dieser Voraussetzung kann (und wird) sich das *Netzwerk aus neuronalen Zuständen* (Attraktoren) ausbilden, das wir als *Geist* auffassen.

Der *Besitz von Augen* ist für uns gleichbedeutend mit der Fähigkeit *zu sehen*. Das ist jedoch falsch. Für ein Tier, das eine lichtempfindliche Zelle besitzt, ist die Welt keineswegs *hell* – das Tier verfügt lediglich über die *Information*, aus welcher Richtung das Licht kommt.

⁸ Natürlich gibt es auch Aktivitäten *ohne* Empfindung, wie Reflexhandlungen oder automatisch ausgeführte Abfolgen von Bewegungen. Das sind dann aber keine *geistigen*, sondern *neuronale* Aktivitäten.

Demnach muss alles, was **unser** Denken und Handeln antreibt, einen *Empfindungsanteil* besitzen. Es gibt kein Handeln oder Denken ohne ein Motiv. Selbst rein logisches Schlussfolgern kann nur stattfinden, wenn wir die korrekte Lösung finden **wollen**.

Umgekehrt folgt aus dem *Fehlen* von Empfindung:

KI-Systeme können nichts wollen oder nicht-wollen. Sie kennen weder Motiv noch Interesse, weder Neugier noch Ablehnung.

In diesem Bereich ist der Mangel an Differenziertheit des Sprachgebrauchs besonders problematisch. Programmierer sprechen von "Zielen" oder "Absichten" eines KI-Systems, von dem, was es "anstrebt". Es handelt sich dabei aber in allen Fällen nur um die Steigerung eines Parameterwertes und nicht um *Ziele* oder *Absichten*, wie wir sie als Elemente menschlichen Handelns verstehen, die immer mit Emotionen verknüpft sind.

Ich fasse zusammen:

1. ***KI-Systeme können nichts wahrnehmen.***
Ihnen fehlt das "*innere Theater*", das "*Bild*" der Umgebung: Sie *sehen* nicht. Ebenso gilt: sie hören nicht, fühlen nicht, riechen nicht, schmecken nicht. Für sie gibt es *nur Information*.
2. ***KI-Systeme können nichts erleben.***
Sie haben keine Gefühle.
3. ***KI-Systeme können nichts wollen.***
Ihnen fehlt Intentionalität und Motivation.

Gleichgültig, wie die Zukunft der KI auch immer aussehen mag, KI-Systeme werden aufgrund der oben genannten Einschränkungen *niemals* eine neue, überlegene Spezies sein. Die Dystopien, in denen wir ihnen ausgeliefert sind, gehören in den Bereich der Fantasy.

(Im dritten Teil der Arbeit [Willensfreiheit](#) untersuche ich, inwiefern die Leistungen gegenwärtiger und zukünftiger KI durch ihre Empfindungslosigkeit begrenzt werden.)

Bemerkung:

Auch die sogenannte "Simulationshypothese" ist durch unseren Beweis widerlegt: wäre unsere Wirklichkeit eine Simulation, dann hätten *wir selbst* keine Empfindungen und kein Bewusstsein.

Bemerkung über Bewusstsein:

Wie schon zuvor festgestellt, gilt Folgendes:

*Alles, was **definiert** werden kann*, ist durch Informationsverarbeitung **erreichbar**, *alles, was **nicht definiert** werden kann*, ist für Informationsverarbeitung **prinzipiell unerreichbar**: gleichgültig, welche Funktion man auf Information anwendet – das Ergebnis ist immer bloß Information und sonst nichts; die Information "rot" wird niemals zur Empfindung *rot*, die Information "Druck" wird niemals zur Empfindung *Schmerz*.

Deshalb bilden "**Information**" und "**Empfindung**" (in der [oben](#) festgelegten Bedeutung) **das einzige Begriffspaar**, das es ermöglicht, zwischen künstlicher Intelligenz und menschlichem Geist eine klare und eindeutige Grenze zu ziehen und dafür eine Begründung zu liefern.

Daraus folgt, dass der häufig im Mittelpunkt der Diskussion stehende Begriff "**Bewusstsein**" nur dann für diese Grenzziehung geeignet ist, wenn die geistigen Phänomene, die ihm (in seiner jeweiligen Definition) zugeschrieben werden, gemäß ihrer Zugehörigkeit zu *Information* oder *Empfindung* analysiert und eingeteilt werden: der zur *Informationsverarbeitung* gehörende Teil des

Bewusstseins (z.B. jede Art von Objekt-Repräsentation, auch Selbst-Repräsentation) ist *reproduzierbar* – gleichgültig, welche technischen Schwierigkeiten seiner Simulation auch im Weg stehen, während der *über Information hinaus* gehende Teil (Wahrnehmen, Erleben, Wollen, Wünschen, Leiden, Mitfühlen usw.) für KI *unzugänglich* bleibt.

Es wäre also eine unnötige und überdies auf Abwege führende Komplikation, den Unterschied zwischen KI und Geist auf den Begriff "Bewusstsein" zu gründen.

Heinz Heinzmann

Wien, Juni 2026

Bemerkung:

Die vorliegende Arbeit ist das zweite Kapitel meiner Abhandlung "[Woraus besteht die Welt?](#)"

Da darin nur bisher ungelöste Fragen beantwortet werden, sodass die etablierte physikalische und philosophische Orthodoxie nicht in Frage gestellt wird wie in allen anderen Kapiteln, habe ich mich dafür entschieden, dieses Kapitel auch als eigenständige Schrift zu präsentieren.