Warum wir Willensfreiheit und Bewusstsein haben und KI-Systeme nicht

Inhaltsverzeichnis

1. Der Beweis der Willensfreiheit

Abstract

1. Zunächst wird der Unterschied zwischen Wirklichkeit und Beschreibung bestimmt. Davon ausgehend kann gezeigt werden, dass die physikalische Kausalität – im Folgenden als "Kausalität von unten" bezeichnet – *unvollständig* ist.

2. Dies ist eine notwendige Bedingung für die Annahme von Kausalität in komplexeren Ebenen der Wirklichkeit, die durch nicht-physikalische Gesetze geregelt werden. Diese Art von Kausalität – im Folgenden "Kausalität von oben" genannt – wird durch ein Beispiel erläutert und dann allgemein begründet.

3. Die Begründung gilt auch für das menschliche neuronale Netz. Daraus folgt, dass die geistige Ebene die *kausale Ebene* des Netzes ist.

4. Im Unterschied zu den Gesetzen der Physik sind die Gesetze der geistigen Ebene veränderbar. Da die geistigen Prozesse ursächlich sind, müssen auch diese Veränderungen der geistigen Tätigkeit zugeschrieben werden.

5. Für eine Willensentscheidung gilt daher Folgendes:

a) Sie ist kein physikalischer, sondern ein geistiger Prozess.

b) Der Entscheidungsprozess kann die Gesetze ändern, die vor seinem Beginn galten. Wenn aber erst durch diesen Prozess selbst bestimmt wird, was geschehen wird, kann die Entscheidung vorher nicht festgelegt sein.

Sie ist also frei.

1.1. Der Unterschied zwischen Wirklichkeit und Beschreibung

In unserem Universum scheint ganz allgemein Folgendes zu gelten:

***Alles, was existiert*, besteht aus elementaren Objekten, die miteinander wechselwirken. Wie sich diese Objekte verhalten, *wird vollständig durch physikalische Gesetze geregelt*. Somit folgt die gesamte zukünftige Entwicklung aus sogenannten "Anfangsbedingungen" – der Gesamtheit der Attribute aller Objekte zu irgendeinem Zeitpunkt – und physikalischen Gesetzen.**

In diesem Bild, das von der Naturwissenschaft so überzeugend präsentiert wird, ist anscheinend für nichts anderes Platz als für Physik. Gleichgültig, wie komplex die Aggregate auch sein mögen, zu denen sich die elementaren physikalischen Objekte zusammenfügen, gleichgültig, welch phantastische Kreationen die Evolution auch hervorbringt – *letztlich* bleibt alles Physik. Es ist einfach kein Platz für irgendetwas Anderes.

Dieser Sachverhalt lässt sich so konkretisieren:

In der soeben vorgestellten, *reduktionistisch* genannten Sichtweise der Wirklichkeit bleibt die Kausalität immer "unten", d.h. in der elementaren Schicht der Wirklichkeit. Alle anderen, komplexeren Schichten haben ihre Selbständigkeit verloren. Beschreibungen, die sich auf diese Schichten beziehen – etwa neuronale oder psychologische Beschreibungen menschlicher Handlungen, sind bloß vereinfachte, näherungsweise gültige Zusammenfassungen von Prozessen, die *eigentlich* physikalischer Natur sind.

Die Konsequenzen dieser Hypothesen sind ziemlich seltsam, um nicht zu sagen bizarr. Wenn wir etwa annehmen, wir hätten einen Gedanken B deshalb geäußert, ***weil er aus einem anderen Gedanken A folgt***, dann wäre das eine Selbsttäuschung; Es würde bedeuten, eine Kausalität auf der Ebene geistiger Prozesse zu postulieren, also eine Kausalität von "oben". Die Vorstellung, dass *das Denken selbst* zu korrekten Ergebnissen führt, setzt offensichtlich 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: B wäre dann "kausal überbestimmt".

Aus reduktionistischer Sicht gäbe es nur eine einzige Möglichkeit, dass B tatsächlich der Logik entsprechen könnte: Sie bestünde darin, dass die Evolution die physikalischen Prozesse in unserem Gehirn den Erfordernissen der Wirklichkeit soweit angepasst hätte, dass wir uns in einem für unser Überleben ausreichenden Maß logisch verhalten und denken. Aber ich wiederhole: die Überzeugung, dass wir uns *deshalb* so verhalten oder denken, *weil* es logisch ist, wäre eine Täuschung, eine List der Evolution, unser angepasstes Verhalten durch ein angenehmes Gefühl zu verstärken. Und, nebenbei gesagt, wir würden auch niemals feststellen können, ob so etwas wie "Logik" überhaupt existiert, da ja etwas *einzusehen* ebenfalls ein geistiger Prozess wäre, den es *als solchen* gar nicht gibt. Einsichten wären keine Einsichten, Gedanken keine Gedanken, der Geist wäre verschwunden, *wir selbst* hätten uns im Nebel der Selbsttäuschungen verflüchtigt...

Es ist also ein völlig absurdes Bild, das der Reduktionismus entwirft, und ich glaube, dass er nur deshalb so verbreitet ist, weil kein Reduktionist je die Konsequenzen seines Standpunkts vollständig berücksichtigt hat. (Wenn es doch einen gäbe, wäre er allerdings längst verstummt und daher unauffindbar.)

Ich will noch kurz auf die beiden populärsten Versuche eingehen, das Problem zu "entschärfen".

Der erste Einwand ist, dass wegen der quantenmechanischen Unschärfe eine "objektive Unbestimmtheit" in der Natur selbst existiert, sodass nicht behauptet werden kann, dass "die

Zukunft aus Anfangsbedingungen und Gesetzen folgt". Es lässt sich aber behaupten, dass "die Zukunft ausschließlich von Anfangsbedingungen und Gesetzen abhängt" – nur dass diese Gesetze eben nicht mehr deterministisch sind. Die nachstehenden Schlussfolgerungen bleiben dann gültig.

Am häufigsten wird gegen den Reduktionismus eingewendet, dass eine vollständige Reduktion in den meisten Fällen nicht gelungen ist und wohl auch niemals gelingen wird. Ich halte diesen Einwand für unzureichend: Ob es eine Reduktion *gibt,* kann nicht dadurch entschieden werden, ob *wir* dazu imstande sind, sie durchzuführen – das oben skizzierte Bild der Wirklichkeit, das die Grundlage des unglaublichen Erfolgs der Naturwissenschaft ist, wird durch die Einschränkungen, denen *unsere* Mittel und Fähigkeiten unterworfen sind, nicht in Frage gestellt, und das gilt auch für die Folgerungen aus diesem Bild.

Um diesen seltsamen Folgerungen zu entgehen, ist es daher notwendig, das Bild selbst in Frage zu stellen. Also fragen wir uns: *Ist die Behauptung A wahr?*

A: ***Alles, was geschieht, folgt aus physikalischen Gesetzen und Anfangsbedingungen.***

Beginnen wir mit einem Gedankenexperiment:

Wir betrachten folgendes Szenario: eine große Anzahl beliebiger materieller Objekte im leeren Raum, die sich relativ zueinander auf zufällige Weise bewegen, aber so, dass sie gravitativ aneinander gebunden bleiben.

Nehmen wir an, wir wären imstande, die Anfangsbedingungen – also die Gesamtheit der Attribute aller Objekte des Systems – *vollständig genau* zu erfassen und auf eine Beschreibung zu übertragen. Wir kümmern uns also nicht darum, dass wir nicht unendlich genau messen können oder dass wir nicht einmal dazu imstande sind, auch nur den Wert eines einzigen Attributs unendlich genau aufzuschreiben bzw. zu speichern. Außerdem nehmen wir an, dass unser Gravitationsgesetz *richtig* ist und dass wir alle erforderlichen Berechnungen mit unendlicher Genauigkeit durchführen können.

Jetzt vergleichen wir die Lage im *wirklich existierenden System* mit der Lage im *Beschreibungs-system*.

Unter den oben genannten Voraussetzungen wird sich *im existierenden System* ohne Zweifel genau das ereignen, was wir erwarten: jeder Körper wird sich exakt *so* verhalten, wie die Gravitation es ihm vorschreibt. Die Behauptung A scheint sich hier also zu bestätigen.

Und *im Beschreibungssystem*? Nun, hier ereignet sich zunächst *überhaupt nichts*. Obwohl wir in unsere korrekten Gleichungen die unendlich genauen Werte aller Attribute eingesetzt haben, so dass sie die Objekte und ihre zeitliche Entwicklung eigentlich perfekt repräsentieren, verhalten sich die Gleichungen doch nicht so wie die Objekte selbst: Während sich die *wirklich existierenden Objekte* von dem Zeitpunkt an, den wir zur Messung ihrer Attribute gewählt haben, *von selbst* weiter bewegen und auf diese Weise die gravitativ determinierte Dynamik des Systems vollziehen, tun das die Gleichungen offensichtlich nicht – sie bleiben einfach unverändert so stehen, wie wir sie notiert haben.

Das ist eigentlich vollkommen selbstverständlich. Ich war trotzdem ein wenig ausführlicher als nötig, weil wir damit auf einen außerordentlich wichtigen Sachverhalt gestoßen sind, der aber – vermutlich gerade *wegen* seiner trivial erscheinenden Selbstverständlichkeit – weder von der Philosophie noch von der Naturwissenschaft zur Kenntnis genommen worden ist.

Er lautet:

Satz:

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

Kehren wir zu unserem Gedankenexperiment zurück. Wir haben festgestellt: Im *existierenden System* wird sich jeder Körper exakt *so* verhalten, wie die Gravitation es ihm vorschreibt. Wird dadurch tatsächlich die Behauptung A bestätigt?

Die Antwort ist: *Nein, das wird sie nicht!* Wir haben ja dem wirklich existierenden System etwas hinzugefügt, was in A nicht enthalten ist: *Aktivität*.

Dass die Wirklichkeit *aktiv* ist, bedeutet, dass sich an jedem Punkt zu jeder Zeit genau das vollzieht, 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. [1]

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

Haben wir ein solches Verfahren? Zunächst ist klar, dass sich für eine "große Anzahl" von Körpern, die sich zufällig bewegen, unsere Gleichungen nicht lösen lassen. Tatsächlich haben wir nur eine einzige Möglichkeit, etwas über die weitere Entwicklung des Systems zu erfahren: Da wir das Gravitationsfeld kennen, können wir für jeden Körper berechnen, wohin er sich nach einem bestimmten Zeitintervall Δt *in diesem Feld* bewegt haben würde – und hier ist der Konjunktiv erforderlich, weil er sich selbstverständlich *nicht* in *diesem* Feld bewegt: es bewegt sich ja nicht nur der eben betrachtete Körper, sondern auch alle anderen, und das bedeutet, dass auch das Feld sich permanent verändert. Um aber überhaupt irgendetwas berechnen zu können, müssen wir für kleine Zeitintervalle das Feld als *statisch* annehmen. Wir führen dann dieselbe Art der Berechnung für alle Körper durch. Anschließend machen wir dasselbe 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.

Damit haben wir gezeigt, dass die Behauptung A falsch ist. Da es kein Verfahren gibt, mit dem man von der Gegenwart in die Zukunft gelangt, kann sie nicht aufrechterhalten werden.

Satz:

Es gibt Systeme, deren künftige Entwicklung nicht aus physikalischen Gesetzen und Anfangsbedingungen folgt.

Aber wird uns nicht *durch die Wirklichkeit selbst* andauernd vor Augen geführt, dass die Zukunft aus der Gegenwart folgt? Keineswegs. Was wir sehen, 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 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

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

wir uns gezwungen, nun unser "folgt aus" durch eine Reihe logischer Schritte zu ersetzen. Somit landen wir zwangsläufig wieder bei einem mathematischen Verfahren, und zuletzt wieder bei der Tatsache, dass kein solches Verfahren existiert – selbst dann nicht, wenn wir uns vorstellen, wir wären von allen Beschränkungen des Messens und Rechnens befreit.

Die Zukunft folgt also nicht immer aus der Gegenwart. Was ergibt sich daraus?

Die wichtigste Folge ist, dass dadurch ein *logischer Freiraum* entstanden ist: Wenn Anfangsbedingungen und physikalische Gesetze hinreichen würden, um daraus die Zukunft abzuleiten, dann wäre in der Menge der Bedingungen für die Ableitung der Zukunft kein Platz mehr; Da sie aber *nicht* hinreichen, ist in dieser Menge nun Raum für weitere Bedingungen.

Satz:

Die Kausalität von unten ist unvollständig. Es ist Raum für Kausalität von oben.

1.2. Nicht-physikalische Kausalität

Unser nächster Schritt wird sein, zu klären, um welche "weiteren Bedingungen" es sich handeln könnte, von denen die künftige Entwicklung von Systemen abhängt – zusätzlich zu Anfangsbedingungen und physikalischen Gesetzen. Sind es andere Arten von Daten? Oder andere Arten von Gesetzen? Um das zu ermitteln wechseln wir den Schauplatz.

Wir betrachten 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. Hier müssen wir also weder auf die physikalischen Objekte – die Glasmoleküle – noch auf die physikalische Wechselwirkung – den Elektromagnetismus – eingehen, um den Ton vorauszusagen. 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*** nennen werde.

Vergleichen wir unsere beiden Szenarien, das der gravitierenden Körper und das des schwingenden Gefäßes:

Im Gravitationsszenario sind die Anfangsbedingungen als ***lokale Parameter*** gegeben, als Attribute der einzelnen Körper. Ihre Werte werden in das ***physikalische Gesetz*** – das Gravitationsgesetz – eingesetzt. Obwohl alles, was sich ereignet, vollständig diesem Gesetz entspricht, ist es dennoch unmöglich, die weitere Entwicklung vorauszusagen. Die Zukunft des Systems ***folgt nicht*** aus seiner Gegenwart.

Im Glasszenario sind es nicht die Attribute der Glasmoleküle, die in das Gesetz eingehen, sondern die Abmessungen des Glases, also ***globale Parameter***. Das Gesetz ist kein physikalisches Gesetz, sondern ein ***Strukturgesetz***. Aus den globalen Parametern und dem Gesetz lässt sich die weitere Entwicklung ableiten. Die Zukunft des Systems ***folgt*** aus seiner Gegenwart.

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.

Zuvor hatten wir uns gefragt, welche Arten von Daten und Gesetzen es neben physikalischen Anfangsbedingungen und Gesetzen noch geben könnte. Das einfache Beispiel des schwingenden Gefäßes hat uns eine Antwort geliefert:

1. neue Daten in der Form *globaler Parameter*
2. neue Gesetze in der Gestalt von *Strukturgesetzen,* die auf den globalen Parametern beruhen.

Da sich mittels dieser neuen Daten und Gesetze die Zukunft des Systems voraussagen lässt, sind sie tatsächlich Elemente der "Menge der Bedingungen für die Ableitung der Zukunft", mit der wir uns oben beschäftigt haben.

Am wichtigsten für unsere Überlegungen ist aber zweifellos Folgendes:

Die lokalen Parameter – 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.

(Voraussetzung dafür, dass es sich dabei tatsächlich um "Kausalität von oben" handelt, ist allerdings, dass im betreffenden System die physikalische Kausalität – die "Kausalität von unten" – *unvollständig* ist, genauso, wie wir das im Gravitationsszenario nachgewiesen haben. Da das Glasgefäß aber nur zur Demonstration dienen sollte, worum es geht, brauchen wir uns nicht darum zu kümmern, ob diese Bedingung hier erfüllt ist.)

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

1.3. Das menschliche neuronale Netz

Gegenstand unserer Untersuchung ist die folgende Frage:

Welcher Art von Kausalität gehorcht das neuronale Netz?

Im Netz finden wir drei Ebenen ansteigender Komplexität vor:

1. die physikalische Ebene
2. die neuronale Ebene
3. die geistige Ebene

Bezogen auf diese Einteilung lautet unsere Frage also:

Von welcher Art von Prozessen hängt es ab, was im Netz geschieht? Von physikalischen, von neuronalen oder von geistigen Prozessen? Welche Ebene ist die kausale Ebene? – oder, anders gefragt: Welche Ebene ist dominant?

Zunächst zur *physikalischen Ebene*. Nehmen wir an, wir hätten vollständiges Wissen über die Werte der Attribute aller physikalischen Objekte des Netzes und könnten somit das Gleichungssystem aufstellen, das den Zustand des Netzes und seine weitere Entwicklung repräsentiert. (Natürlich ist diese Vorstellung völlig absurd, aber in der Form eines Gedankenexperiments ist sie zulässig – *im Prinzip* muss dieses Gleichungssystem ja existieren.)

Jetzt sind wir aber wieder mit dem Problem konfrontiert, das schon beim Gravitationsszenario die Berechnung der Entwicklung des Systems verhindert hat: Eine ungeheure Zahl von Prozessen läuft zeitgleich ab, und jeder von ihnen ist mit etlichen anderen direkt vernetzt. Um aber irgendeinen Prozess berechnen zu können, müssen wir 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.

Wir sind also, wie beim Gravitationsszenario, auf Näherungen angewiesen, die schon nach kurzer Zeit erheblich von der Wirklichkeit abweichen können. Es ist nicht möglich, die Entwicklung des Netzes vorauszusagen. Die Behauptung "Was im Netz geschieht, folgt aus Anfangsbedingungen und physikalischen Gesetzen" ist falsch.

Und auch hier gilt wieder: Die Wirklichkeit tut, wozu wir nicht in der Lage sind: aufgrund ihrer *Aktivität* arbeitet sie zeitgleich die ungeheure Zahl von Prozessen ab, sodass wir den Eindruck gewinnen, alles "folge aus" Anfangsbedingungen und physikalischen Gesetzen.

Satz:

Im neuronalen Netz ist die physikalische Kausalität unvollständig. Es ist Raum für Kausalität von oben.

Betrachten wir nun die *neuronale Ebene*. Sie 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. Die neuronale Aktivität wird durch ein Gesetz geregelt, das aus dem neuronalen Input-Output-Mechanismus folgt.[2]

Dieses Gesetz kann als *Wechselwirkungsgesetz der Neuronen* aufgefasst werden. (Es dient auch als Grundlage für Computersimulationen.)

Auch auf dieser Ebene erscheint es uns im ersten Moment wieder völlig selbstverständlich, dass aus den Anfangsbedingungen der Neuronen und ihrem Wechselwirkungsgesetz folgt, was sich im Netz ereignen wird. Und abermals müssen wir erkennen, dass wir wieder derselben Täuschung erlegen sind, indem wir Wirklichkeit und Beschreibung nicht voneinander unterschieden oder miteinander verwechselt haben:

Da ja das neuronale Wechselwirkungsgesetz eine Zusammenfassung physikalischer Sachverhalte ist, bleibt auch das Argument gültig, mit dem wir gerade eben die Behauptung widerlegt haben, dass alles aus Anfangsbedingungen und physikalischen Gesetzen folgt. Für die neuronale Ebene gilt somit: Der hohe Vernetzungsgrad der Neuronen – die permanente Rückkopplung, die sich daraus ergibt – schließt die Existenz eines mathematischen Verfahrens zur Berechnung der weiteren Entwicklung aus.

2 Mit der Bezeichnung "Input-Output-Mechanismus" ist Folgendes gemeint: Die Dendriten jedes Neurons werden über Synapsen durch andere Neuronen stimuliert oder inhibiert. Die auf diese Weise verursachte elektrische Erregung wird zum Zellkörper weitergeleitet und dort aufsummiert. Wenn eine bestimmte Grenze überschritten ist, wird sie an das Axon abgegeben und auf dessen Verzweigungen verteilt, sodass sie schließlich über synaptische Verbindungen weitere Neuronen beeinflusst.

Satz:

Auch die Beschreibung durch neuronale Anfangsbedingungen und das neuronale Wechselwirkungsgesetz lässt Raum für Kausalität von oben.

Damit kommen wir zuletzt zur komplexesten Ebene, der *Ebene des Geistes*. 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.[3]

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

Stellen wir uns ein neuronales Netz vor, in dem es noch keine Repräsentationen gibt. 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 dazu auch die Bemerkung auf Seite 22 oben.)

Zuvor haben wir festgestellt:

Unter der Voraussetzung, dass die Kausalität von unten unvollständig ist, folgt aus der Existenz von Attraktoren, dass das betreffende System, falls es im Attraktor-Zustand selbst oder diesem Zustand "nahe genug" ist,[4] durch Kausalität von oben bestimmt wird.

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.

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*

3 "Sachverhalt" muss hier im weitest-möglichen Sinn aufgefasst werden.

4 Ohne den Begriff des Phasenraumes lässt sich dieses "nahe genug" nicht wirklich definieren. Das neuronale Netz ist jedenfalls *immer* "nahe genug" an einem Attraktor-Zustand.

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! Oder – um an die bei der Kritik des Reduktionismus verwendeten Formulierungen anzuschließen: Einsichten sind Einsichten, Gedanken sind Gedanken, der Geist ist in seine Rechte gesetzt, *wir selbst* sind wir selbst...

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 bzw. die neuronale Kausalität durch die geistige Kausalität ersetzt. Damit haben wir erreicht, dass unser Geist nicht durch physikalische oder neuronale 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.

1.4. Der Unterschied zwischen physikalischen und geistigen Gesetzen

Menschliche neuronale Netze unterscheiden sich stark voneinander, und zwar auch dann, wenn noch keine Strukturierung durch äußere Reize stattgefunden hat. Daraus folgt unmittelbar, dass auch die Muster, die etwas repräsentieren, bei allen Menschen verschieden sind, selbst dann, wenn der repräsentierte Sachverhalt identisch ist.

Die Reihenfolge der 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,[5] und abgebaut, wenn sie inaktiv sind. Das bedeutet zugleich, dass jede geistige Aktivität die Struktur des Netzes beeinflusst.

5 Diese Erkenntnis geht auf Donald Hebb zurück, der 1949 in *The Organization of Behavior* feststellte: When an axon of cell A is near enough to excite B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

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.

1.5. Die Begründung der Freiheit

Die offensichtlichste Folgerung der Verstärkung aktiver neuronaler Verbindungen ist allerdings, dass das, was wir *immer* denken, fühlen und tun, sich selbst verstärkt. Es ist aber im Grunde selbstverständlich, dass auch das Gegenteil eintreten kann:

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.

Postskriptum

Bei der Durchsicht des Textes schien es mir, als wäre ich meinem Ziel, das Thema so kurz und einfach wie möglich darzustellen, ein wenig zu radikal gefolgt. Deshalb will ich abschließend versuchen, die wichtigsten Punkte meiner Argumentation nochmals zu erläutern:

Nehmen wir an, wir hätten ein System zu beschreiben, das aus einer großen Zahl physikalischer Prozesse besteht, die miteinander verkoppelt sind. Die Gleichungen der Prozesse sind also ebenfalls miteinander vernetzt. Für eine exakte Beschreibung benötigen wir dann *in jedem Augenblick* die Werte aller Parameter aller Prozesse, um sie in die Gleichungen der jeweils anderen Prozesse einzusetzen – mit anderen Worten: es ist (außer in sehr einfachen Fällen) unmöglich, *mit physikalischen Mitteln* über das System, das aus allen diesen Prozessen besteht, genaue Voraussagen zu machen, und zwar aus *prinzipiellen* Gründen, und nicht nur wegen der Einschränkungen des Messens und Rechnens.

Und damit wären wir am Ende unserer Möglichkeiten angelangt – *es sei denn,* die betrachteten Prozesse könnten als Elemente einer "Struktur höherer Ordnung" aufgefasst werden, in der weitere Gesetze gelten. Diese "Gesetze höherer Ordnung" sind dann aber *keine physikalischen Gesetze* mehr*,* und damit haben wir den Bereich der Physik verlassen.

Falls diese neuen Gesetze eine Voraussage über die Entwicklung des Gesamtsystems ermöglichen, dann gilt somit Folgendes:

1. Die Entwicklung des Gesamtsystems ***folgt nicht aus physikalischen Gesetzen***.
2. Die Entwicklung des Gesamtsystems ***folgt aus Gesetzen höherer Ordnung***.

Natürlich geschieht auch weiterhin alles ***in Übereinstimmung*** mit den physikalischen Gesetzen – aber diese Gesetze vollziehen sich nun innerhalb einer ***übergeordneten Struktur***. (Wie beim schwingenden Glasgefäß.)

Die Kausalität ist also nicht mehr *unten* – im elementaren, physikalischen Bereich: sie ist *nach oben* gewandert, in einen Bereich höherer Ordnung, in dem ***neue, nicht-physikalische Gesetzmäßigkeiten*** gelten.

Genau diese Verhältnisse finden wir im neuronalen Netz vor, und zwar mehrfach:

In einem Neuron laufen zahlreiche physikalische Prozesse zeitgleich ab. Die physikalische Betrachtungsweise ermöglicht uns zwar ein Verständnis dessen, was im Neuron vor sich geht, aber die Verkopplung der Prozesse verhindert eine exakte Berechnung der weiteren Entwicklung. Diese Prozesse sind jedoch durch die *Form und Struktur des Neurons* in ein System höherer Ordnung eingebettet, sodass sie einem "Strukturgesetz" gehorchen, das wir zuvor "neuronales Input-Output-Gesetz" genannt haben.

Nun gilt aber wiederum, dass uns auch *dieses* Gesetz keine genaue Voraussage über die künftige Entwicklung von vielen aneinander gekoppelten Neuronen ermöglicht. Die Neuronen sind jedoch selbst wiederum Elemente eines Systems höherer Ordnung – eben des neuronalen Netzes mit seinen

aufgeprägten Mustern (Attraktoren). Damit sind also auch die Neuronen einem neuen Gesetz unterworfen: einem Strukturgesetz abermals höherer Ordnung: dem Gesetz der Abfolge neuronaler Muster, und das heißt: ***dem Gesetz des Geistes***. Somit ist der Geist die *kausale* Ebene. Er bestimmt die im Netz ablaufenden Prozesse – auch diejenigen, die dieses Gesetz selbst verändern.

Zuletzt nochmals der Hinweis auf den Unterschied zwischen *Beschreibung* und *Wirklichkeit*:

Um in der ***Beschreibung*** eines Systems von der Gegenwart in die Zukunft zu gelangen, benötigen wir irgendwelche Verfahren. Das können mathematische Verfahren sein, Algorithmen oder Gleichungen, aber auch Methoden, Sachverhalte so zusammenzufassen, dass sich daraus Schlüsse ziehen lassen. In manchen Fällen gelingt uns das so gut, dass wir behaupten können, B *folge* aus A.

In der ***Wirklichkeit*** ist das alles nicht notwendig. Wenn an jedem Ort zu jeder Zeit geschieht, was zu geschehen hat, dann entsteht die Zukunft *von selbst,* dann entwickeln sich alle komplexen Objekte und Strukturen samt ihren Gesetzmäßigkeiten *von selbst.*

Aber daraus, dass in der Wirklichkeit der Vollzug elementarer Prozesse für die Entstehung der Zukunft hinreicht, kann nicht geschlossen werden, dass die Zukunft aus elementaren Prozessen ***folgt***, denn das würde voraussetzen, das, was in der Wirklichkeit ***von selbst*** geschieht, in eine ***Reihe logischer Schritte*** zu übersetzen, und das ist unmöglich.

Bemerkung:

In dieser Begründung der Willensfreiheit ist es *nicht* notwendig, dass im Weltgeschehen eine "Verzweigung" existiert. Der entscheidende Punkt ist hier, dass die Zukunft nicht in der Gegenwart enthalten ist – dass sie also nicht aus der Gegenwart *folgt,* sondern bloß aus ihr *entsteht,* und dass die Gründe für das, was sich dann tatsächlich ereignen wird, geistiger Art sind.

Für den nun folgenden Beweis, dass *wir selbst* Empfindungen und Bewusstsein haben, *KI-Systeme* hingegen empfindungslos und bewusstlos bleiben, werden die im Beweis der Willensfreiheit abgeleiteten Resultate vorausgesetzt.

2. Warum wir empfindungsfähig sind und KI-Systeme nicht

2.1. Ontologische Voraussetzungen

Wir beginnen mit dem Unterschied zwischen Wirklichkeit und Beschreibung, den wir im Abschnitt über Willensfreiheit vorgestellt haben:

Wirklich existierende Objekte sind aktiv, Objekte in einer Beschreibung sind dagegen nicht aktiv. Somit muss zur Existenz wirklicher Objekte etwas gehören, was Objekten in einer Beschreibung fehlt.

Dieses Element der Existenz wirklicher Objekte bezeichnen wir als ***Substanz***. ***Substanz ist also dasjenige, wovon die Aktivität existierender Objekte ausgeht***.

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

Dieses Element ihrer Existenz bezeichnen wir 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**".

Es gilt somit folgender

Satz:

Wirklich existierende Objekte bestehen aus Substanz und Akzidenzien, Objekte in einer Beschreibung 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.)

Für uns besteht also ***jedes existierende Objekt*** 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 beschrieben werden kann, und aus ***Akzidenzien*** – das ist der Teil von Existenz, der beschrieben und definiert werden kann.

Im physikalischen Bereich der Wirklichkeit – oder sagen wir besser: im Bereich der Materie – sind uns diese Verhältnisse vertraut. Wir wissen, dass *Masse* Gravitation bewirkt, und dass *elektrische Ladung* die elektromagnetische Wechselwirkung verursacht. Wir wissen also, *dass da* ***etwas*** *sein muss*, was Ursache der Dynamik ist und benennen es, aber wir wissen nicht, was es "**ist**".

Nun müssen wir bestimmen, was im Bereich des Geistes als Substanz und Akzidenzien aufzufassen ist. Im Abschnitt über Willensfreiheit haben wir bewiesen, dass die geistige Ebene die ***kausale Ebene*** ist. Somit befinden wir uns nicht mehr im physikalischen Bereich, und die dort gültige Systematik kann nicht angewendet werden. Zunächst müssen die Objekte der geistigen Wirklichkeit definiert werden, und danach muss bestimmt werden was ihre Substanz und ihr Akzidens ist.

Im Abschnitt über Willensfreiheit haben wir 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 wie folgt:

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 *definiert,* 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.

(Ich werde geistige Zustände als ***Qualia*** bezeichnen. Der Ausdruck *Quale* steht also für den ganzen geistigen Zustand und nicht bloß für den Empfindungsteil.)

Mit den obigen Bestimmungen ist zugleich klar, was die Substanz und das Akzidens des geistigen Zustands sind:

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

Also ist *Informationsverarbeitung* das Akzidens des Quale.

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

Also ist *Empfindung* die Substanz des Quale.

Und das bedeutet:

***Empfindung* ist der Antrieb der Dynamik des Geistes.**

2.2. Durchführung des Beweises

Nun sind wir vorbereitet, zu begründen, warum wir Empfindungen haben und KI-Systeme nicht.

Zunächst benötigen wir folgende

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, werden wir diese Art der Aktivität als ***zugeführte Aktivität*** bezeichnen, 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 *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.)

Der letzte Baustein unseres Beweises ist folgender

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.

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. 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])

6 Das soll aber nicht etwa heißen, dass es nun zwei Substanzen gibt – vielmehr ist die zweite Substanz als aus der ersten Substanz hervorgehend gedacht, und die Frage, die wir uns stellen, lautet demnach: Warum verwandelt sich *für uns* die erste Substanz im Fall der Qualia in die zweite Substanz Empfindung?

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.

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.

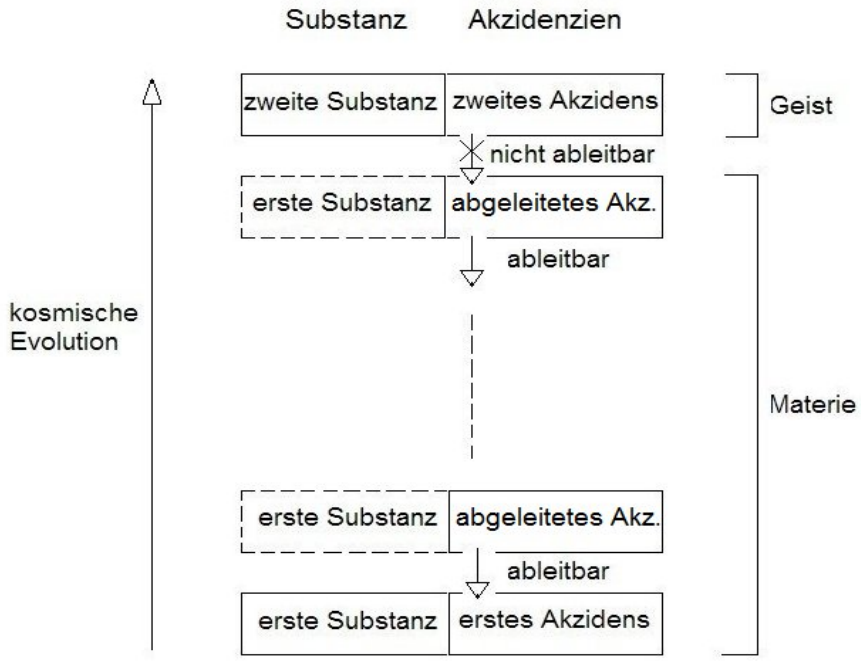
Aufgrund der ***Untrennbarkeit*** von *erster Substanz* und *erstem Akzidens* gilt jedoch:

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 <u>verschieden</u> ist, dann muss auch die zugehörige Substanz von der ersten Substanz <u>verschieden</u> sein.

Hier eine Skizze zur Veranschaulichung:



Der im Rahmen unserer Argumentation entscheidende Punkt ist, dass die Verwandlung des Wesens des Seienden sich nur dann ereignen kann, wenn die Dynamik des betrachteten Systems sich aus der ***untrennbaren Einheit*** von Substanz und Akzidenzien ergibt. ***Nur dann*** folgt aus der Tatsache, dass die Akzidenzien nun nicht mehr 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 aber 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 das neuronale (analog-digitale) Input-Output-Gesetz 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 überhaupt 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.

2.3. Über Empfindung und Künstliche Intelligenz

Üblicherweise wird gefragt:

"Wieso gibt es im Geist etwas Undefinierbares, wie 'Farbe' oder 'Schmerz', und sonst nirgends?"

Wir haben uns stattdessen die Frage gestellt:

"Wieso ändert das Undefinierbare, das es überall in der Wirklichkeit gibt, seinen Charakter, wenn es im Geist auftritt?"

Es wird also nicht nach dem Grund der *Existenz* dieses Undefinierbaren gefragt – was überflüssig wäre, weil es – wie wir gezeigt haben – *in allem Seienden* zu finden und somit selbstverständlich ist[7] – sondern nach dem Grund seiner *Veränderung*.

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

In der zweiten Version lässt sich die Frage aber beantworten, und diese Antwort enthält überdies den Beweis, dass ***Empfindung –*** die *geistige Erscheinungsform* dieses "Undefinierbaren" – in Systemen, die ***nicht durch Evolution entstanden***, sondern ***von uns konstruiert*** sind, ***nicht existiert***.

Ich habe den Begriff "Empfindung" abweichend von seinem üblichen Gebrauch bestimmt. Ich will nun abschließend 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

7 Siehe Seite 13 oben.

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 dienten 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 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. [8]

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.

---

8 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.

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.[9]

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

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 gilt somit: ***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 <u>wahrnehmen</u>***.
   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 <u>erleben</u>***.
   Sie haben keine Gefühle.
3. ***KI-Systeme können nichts <u>wollen</u>***.
   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.

<u>Bemerkung:</u>

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

<u>Bemerkung:</u>

*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*.

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

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 Selbst-Repräsentation) ist reproduzierbar – gleichgültig, welche technischen Schwierigkeiten seiner Simulation auch im Weg stehen, während der zur Empfindung gehörende Teil (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.

Bemerkung:

Für unseren Beweis ist es nicht hinreichend, die Kausalität "nach oben" zu verschieben. Der Grund dafür ist wie folgt: Nehmen wir an, es könnte ein neuronales Netz konstruiert werden, das geeignet ist, Attraktoren auszubilden und zu vernetzen[10] – so, wie wir das bei menschlichen neuronalen Netzen voraussetzen – und nehmen wir außerdem an, dieses Attraktor-Netzwerk sei die *kausale Ebene* des Systems. Dennoch bliebe das System *empfindungslos*: die Bedingung (Seite 17 oben), dass seine Dynamik auf ***wesensgemäßer Aktivität*** beruht – dass sie also aus der *untrennbaren Einheit seiner Substanz und Akzidenzien* hervorgeht – wäre nicht erfüllt.

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:

$$\textit{Wahrnehmen} = \textit{Wiedererkennen},$$

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 Kompo-

10 Die gegenwärtig populären künstlichen neuronalen Netze (z.B. GPTs) sind dafür nicht geeignet.

nenten 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.

Bemerkung:

Zuletzt noch ein Kommentar zum Szenario der gravitierenden Körper aus dem Abschnitt über Willensfreiheit:

Sogar ein Laplacescher Dämon mit unendlichen Ressourcen an Raum, Zeit und Information würde an der Berechnung scheitern: Um die Zukunft des Systems *exakt* zu ermitteln, muss der Dämon die Berechnung für unendlich kleine aufeinander folgende Zeitintervalle durchführen. Falls die Intervallgrenzen genauso dicht liegen wie die *reellen* Zahlen, wird er sogar in unendlich langer Zeit nicht fertig, falls sie aber weniger dicht liegen (wie z.B. die *rationalen* Zahlen), wird es geschehen, dass ihm eine Instabilität entgeht, die *zwischen* zwei Zeitpunkten seiner Berechnungen liegt.

Tatsächlich haben wir aber auch mit dieser Argumentation noch immer nicht das ganze Ausmaß des Problems erfasst: Wir haben ja angenommen, dass wir aufgrund der vollständigen Kenntnis der Anfangsbedingungen auch das Gravitationsfeld kennen. Diese Annahme ist jedoch falsch, und zwar aus folgendem Grund:

Bezeichnen wir den Zeitpunkt, zu dem wir genaue Kenntnis der Anfangsbedingungen besitzen und an dem unsere Berechnung beginnen soll, mit t_0. Wenn wir für irgendeinen der Körper, sagen wir: den Körper A, berechnen wollen, wohin er sich im ersten Zeitintervall bewegt, dann müssen wir sämtliche Wirkungen kennen, denen A zur Zeit t_0 vonseiten der anderen Körper ausgesetzt ist.

Betrachten wir z.B. den Körper B: wir kennen den Ort, an dem er sich zur Zeit t_0 befindet. Die von B stammende Wirkung, der A zur Zeit t_0 ausgesetzt ist, geht jedoch ***nicht*** von ***diesem*** Ort aus, sondern von einem Ort, an dem sich B ***vorher*** befand – und zwar genau *so lange* vorher, wie die Gravitation benötigte, um ***von dort aus*** den Körper A zur Zeit t_0 zu erreichen. Um die Wirkung von B auf A zur Zeit t_0 zu ermitteln, müssen wir daher B auf seiner Bahn *in die Vergangenheit versetzen*, und genau dasselbe gilt auch für alle anderen Körper: sie alle müssen in die Vergangenheit versetzt werden – umso weiter, je weiter sie von A entfernt sind.

Bevor wir überhaupt damit ***beginnen*** können, die Bahn von A zu ermitteln, müssen wir also zunächst die Bahnen aller anderen Körper bestimmen. Dafür ist es aber erforderlich, auch die Wirkung zu kennen, die A zur Zeit t_0 auf die anderen Körper ausübt, und deshalb müssen wir auch A selbst auf seiner Bahn in die Vergangenheit versetzen, d.h. auf der Bahn, die uns nicht bekannt ist, weil wir sie ja soeben erst berechnen wollten!

Dasselbe gilt für *jeden* Körper: um ihn in die Vergangenheit zu versetzen, müssen wir die Bahnen *aller anderen* Körper kennen. Da uns aber *keine einzige* dieser Bahnen bekannt ist, ist es unmöglich, die genauen Positionen zu bestimmen, wo sich die Körper vorher befanden, und damit ist es auch unmöglich, die Wirkungen zu ermitteln, denen sie zum Zeitpunkt t_0 ausgesetzt sind.

Mit anderen Worten: Wir – und mit "wir" meine ich uns alle ***und*** den Laplaceschen Dämon – sind nicht nur außerstande, eine ***exakte Berechnung*** der Zukunft ***auszuführen***, wir sind nicht einmal in der Lage, damit auch nur ***zu beginnen***.

Das Szenario ist nicht berechenbar. Die *Wirklichkeit* ist nicht berechenbar. *Wir selbst* sind nicht berechenbar.

Die formale Version unserer ontologischen Argumentation zur Willensfreiheit lautet also wie folgt:

Das Verhalten aller elementaren Objekte wird von physikalischen Gesetzen bestimmt. Versucht man aber, die Zukunft (oder, falls objektiver Zufall einbezogen werden soll: *irgendeine* Version der Zukunft) auf physikalische Weise abzuleiten, dann scheitert das an der Tatsache, dass dafür überabzählbar viele logische Operationen erforderlich wären.

In manchen Fällen kann jedoch die überabzählbare Menge logischer Operationen durch eine endliche Menge von Aussagen über eine höhere, *nicht-physikalische* Schicht der Wirklichkeit ersetzt werden. Die Fakten, auf die sich diese Aussagen beziehen, können dann als Ursachen (oder *Gründe*) für den künftigen Zustand aufgefasst werden.

Heinz Heinzmann, Wien 2023

(Von dieser Arbeit gibt es auch eine längere Version, in der die Folgen für die KI analysiert werden.)