

Neuronale Vernunft

1. Fragestellung

Platon sagt, genauer er lässt Sokrates sagen: *Ich weiß, dass ich nichts weiß.*

Immanuel Kant sagt: *Es gibt synthetische Urteile a priori.*

Kant widerspricht Platon, denn entschlüsselt lautet seine Behauptung etwa so: Es gibt inhaltsvolles (synthetisches) absolut sicheres (a priori). Wissen (Urteile). Inhaltsleer ist z.B. das Wissen, dass ein Kreis keine Ecken hat. Wer hat Recht, Platon oder Kant? Kant hat seine Behauptung in der „Kritik der reinen Vernunft“ /4/ bewiesen, die vor 222 erschienen ist. Es ist sicher lohnend, über Kants Satz aus der Sicht der modernen Wissenschaft, insbesondere der Gehirnforschung und der KI-(Künstlichen-Intelligenz-)Forschung neu nachzudenken. Kant fragt: *Was kann ich wissen?* Er sucht die Antwort, indem er eine andere Frage untersucht: *Wie erlange ich Wissen?* Aus heutiger Sicht kann man konkreter fragen: *Wie gelangt Wissen ins Bewusstsein, wie in den Computer und wie in die graue Materie?* Die sogenannte graue Materie des Gehirns besteht aus vielen Milliarden untereinander vernetzter Neuronen. Sie sind der stoffliche Träger des Bewusstseins, des Denkens und des Gedächtnisses. Außerdem stellt sich die Frage, ob Kants Schlussfolgerungen den Ergebnissen der modernen Wissenschaft standhalten. Ich werde einige Antworten auf diese Fragen zu geben versuchen.

Zur Einstimmung auf das hohe Abstraktionsniveau, auf dem wir Kant folgen müssen, führe ich den Begriff der Aussage ein: Die Verknüpfung eines Objekts mit einem Merkmal (Attribut) heißt Aussage, z.B.: Ich wiege 55 kg. Je nach Zusammenhang kann eine Aussage als Aussagesatz, Gedanke, Bewusstseinsinhalt, Wissen, Erkenntnis, Urteil oder Urteilsform bezeichnet werden. Beispielsweise ist ein Aussagesatz eine artikulierte, ein Gedanke eine gedachte Aussage. Ein Urteil ist eine Aussage, die mindestens einen allgemeinen Begriff enthält, z.B.: Kornblumen sind blau. Wo Kant von Erkenntnis oder von Urteilen spricht, sprechen Informatiker i. Allg. von Wissen und meinen damit allgemeingültiges Wissen. In jedem Falle handelt es sich um Aussagen

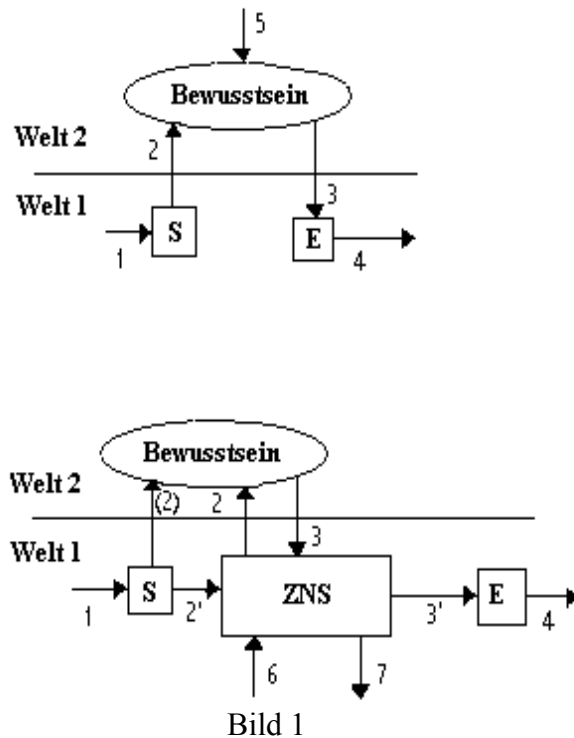
2. Psychologische und neurophysiologische Sichten

Kant beantwortet die Frage *Was kann ich wissen?* in den Grenzen der Möglichkeiten seiner Zeit, d.h. der Möglichkeiten der psychologischen Methode. Denn nur diese stand ihm zur Verfügung, und ihm als Philosophen sogar nur die introspektive Methode. Aus psychologischer und speziell aus introspektiver Sicht ist das Gehirn ein "schwarzer Kasten", in den man nicht hineinsehen kann. Man "sieht" weder, was im Schädel ist, noch was „draußen“ ist, sondern das, was im *Bewusstsein* ist.

Im oberen Teil von Bild 1 ist das Areal sinnbildlich dargestellt, in das Kant bei der Suche nach Antworten auf seine Frage „hineinsehen“ konnte, nämlich sein eigenes Bewusstsein. Den Eingang des Bewusstseins bilden die Sensoren (S) der Sinnesorgane, den Ausgang die Effektoren (E), die den Bewegungsapparat steuern, den Sprechapparat eingeschlossen. Pfeil 1 bzw. 4 stellt die Einwirkungen *aus* bzw. *auf* die Umwelt dar. Die Pfeile 2 und 3 symbolisieren fiktive „Wechselwirkungen“ zwischen dem Bewusstsein und seiner Außenwelt, zu der auch der eigene Körper gehört. Über Pfeil 2 treten Sinnesempfindungen ins Bewusstsein; über Pfeil 3 „steuert“ das Bewusstsein mittels der Effektoren (E) den Bewegungsapparat und damit Handeln und Sprechen.

Die Pfeile 2 und 3 symbolisieren Schnittstellen zwischen dem Bewusstsein und seiner Außenwelt; sie durchstoßen die waagerechte Linie, die den sogenannten *kartesischen Schnitt*

symbolisiert. Er trennt die sinnlich wahrnehmbare, physische Welt von der Welt der Gedanken, oder Welt 1 von Welt 2, wie Karl Popper die beiden Welten nennt. Die Pfeile 2 und 3 stellen also keine Wechselwirkung im physikalischen Sinne dar, im Gegensatz zu den Pfeilen 1 und 4 (siehe die Bemerkung weiter unten). Pfeil 5 stellt die Wirkung einer höheren Instanz dar, an deren Existenz man glauben kann aber nicht muss. Sie wird im Weiteren nicht in die Betrachtung einbezogen, was nicht heißt, dass sie grundsätzlich negiert wird.



Mehr konnte Kant nicht sehen. Er sah im Grunde nur in sein eigenes Bewusstsein hinein. Wie lässt sich aus dieser Sicht eine Erkenntnistheorie aufbauen? Das scheint unmöglich zu sein, denn das Denken, das die Theorie entwickelt, ist gleichzeitig Mittel und Gegenstand der Nachforschung. Die Theorie ist also zwangsläufig in einem Zirkel gefangen. Dessen war sich Kant bewusst. Er musste diese Zirkularität an einem bestimmten Punkte abbrechen. Dieser Punkt ist durch die *reinen Anschauungs-* und *Urteilsformen* markiert, auf die später eingegangen wird.

Einer modernen Erkenntnistheorie steht ein größeres Areal zur Verfügung, innerhalb dessen nach Antworten auf Kants Frage gesucht werden kann. Sinnbildlich ist es im unteren Teil von Bild 1 dargestellt. Es enthält außer dem Bewusstsein eines Menschen dessen Zentralnervensystem (ZNS). Den Ein- bzw. Ausgang des ZNS bilden die Sensoren (S) bzw. die Effektoren (E). Die Pfeile 1 und 4 entsprechen denen im oberen Bild. Aus dem dortigen Pfeil 2 ist die Pfeilfolge 2'-2 und aus Pfeil 3 ist die Pfeilfolge 3-3' geworden. Pfeil (2) deutet an, dass evtl. eine direkte „Einwirkung“ der Sensoren auf das Bewusstsein möglich ist. Die Pfeile 2, (2) und 3 symbolisieren Schnittstellen zwischen dem Bewusstsein und seiner Außenwelt. Zu Kants Zeiten lagen sie völlig im Dunkel. Heute werden bestimmte Abschnitte des ZNS, genauer des assoziativen Kortex, als der stoffliche Träger des Bewusstseins angesehen. Die Pfeile 6 und 7 symbolisieren Wechselwirkungen mit der Umwelt, die nicht über die Sinnesorgane stattfinden, wie möglicherweise die Wirkung des Wetters oder des Elektroschocks. Ferner könnten die Pfeile 6 und 7 zur Erklärung parapsychologischer Phänomene ad hoc postuliert werden. Sie werden im Weiteren nicht betrachtet.

Bemerkung. Unsere Denkgewohnheiten könnten dazu verführen, die Pfeile 2 und 3 als physische Ursache-Wirkungsbeziehungen zu missdeuten. Diese hätten jedoch zwischen

Welt 1 und Welt 2 keinen Sinn. Tatsächlich stellen die Pfeile lediglich Zuordnungen dar, sodass man sie zu einer einzigen Verbindungslinie zusammenfassen kann, die bedeutet, dass zwischen mentalen und neuronalen Phänomenen eine Zuordnung besteht (vgl. die Brückenhypothese in Abschnitt 4.2.2).

Die Erweiterung des Suchareals um das ZNS öffnet der Erkenntnistheorie einen zweiten Weg, den neurophysiologischen, d.h. den Weg über die Erforschung neuronaler Prozesse, die im Gehirn ablaufen. Das Erforschen des Denkens und Erkennens ist nicht mehr auf Psychologie und Introspektion eingeschränkt, der Zirkel ist aufgebrochen.

Den beiden genannten Wegen der erkenntnistheoretischen Forschung entsprechen zwei Wege der KI-Forschung. KI ist die Abkürzung für *Künstliche Intelligenz*, womit eine mittels Computer simulierte natürliche Intelligenz gemeint ist. Die klassische KI geht den psychologischen Weg, die alternative KI den neurophysiologischen. Beide Forschungsbereiche werden der Informatik zugerechnet.

3. Wie kommt Wissen in den Computer?

Um menschliche Intelligenz zu simulieren, geht der klassische KI-Ingenieur folgendermaßen vor. Er analysiert das Vorgehen des Menschen (psychologische Methode), überführt es in einen Algorithmus und diesen in ein Computerprogramm, das dann in das Gedächtnis eines Computers einspeichert wird. Nach dem Bild des Nürnberger Trichters „trichtert“ er dem Computer das eigene Wissen ein. Nach der Eingabe liegt das Wissen im Computer binär codiert vor, d.h. in Form von Zeichenketten, in denen nur zwei verschiedene Zeichen vorkommen, z.B. Nullen und Einsen. Beispielsweise ist 01001110 eine Kette von Binärzeichen. Sie kann der Binärcode irgendeines Objektes, einer Zahl, einer Rechenregel oder irgendeiner Aussage (eines Aussagesatzes) sein. Die binäre Codierung ist notwendig, weil das Gedächtnis eines Computers binär arbeitet, d.h. jedes Glied der zu speichernden Zeichenkette wird durch ein elektronisches Bauelement dargestellt, das zwei Zustände einnehmen kann. Der eine Zustand stellt die Null, der andere die Eins dar. Wenn zur Speicherung Transistoren verwendet werden, können die beiden Zustände *leitend* und *nichtleitend* oder *hohes Potenzial* und *niedriges Potenzial* sein; im Falle von Lochkarten sind es die Zustände *geloht* und *nichtgeloht*. Zur Abspeicherung obiger Kette wären 8 solcher Bauelemente erforderlich. Den Gesamtzustand aller 8 Bauelemente nenne ich *codierenden Zustand*. Dieser Begriff spielt im Weiteren eine wichtige Rolle. Seine allgemeine Begriffsbestimmung lautet: Ein codierender Zustand ist ein ausreichend stabiler physikalischer Zustand eines Speichermediums, der eine Bedeutung codiert (darstellt). Der codierende Zustand einer Lochkarte ist das Muster aller Löcher.

In der Definition tritt das Wort *Bedeutung* auf, das seinerseits definiert werden muss. Es ist ein schwieriges Wort, denn es schlägt die Brücke zwischen dem codierenden Zustand und dem Bewusstsein desjenigen, der das codierte Wissen eingegeben bzw. ausgelesen hat. Tatsächlich gehört die Bedeutung eines Wortes oder Satzes im umgangssprachlichen Sinne zur Welt 2, denn wenn man von Bedeutung spricht, meint man damit, genau betrachtet, dasjenige, was das Wort bzw. der Satz ins Bewusstsein treten lässt. Damit ergibt sich folgende Begriffsbestimmung:

Die Bedeutung einer Zeichenkette, z.B. eines Wortes oder Satzes, ist der Bewusstseinsinhalt, den der Sender (Schreiber oder Sprecher) in die Zeichenkette hineingelegt hat, m. a. W. den der Sender mittels der Zeichenkette artikuliert hat, bzw. sie ist der Bewusstseinsinhalt, den die Zeichenkette im Bewusstsein eines Empfängers (Lesers oder Hörers) auslöst, m. a. W. sie ist die Interpretation der Zeichenkette durch den Empfänger. Die Bedeutung einer Zeichenkette ist an den Artikulierer bzw. Interpretierer gebunden.

Ein Code, der einer bestimmten Bedeutung zugeordnet wird, ist arbiträr, d.h. er kann beliebig gewählt werden, doch ist eine einmal getroffene Wahl für immer verbindlich.

Die Antwort auf die Frage *Wie kommt Wissen in den Computer?* lautet also: Das Einbringen von Wissen in den Computer besteht im Realisieren entsprechender codierender Binärzustände. Wenn beispielsweise einem Computer das Wissen eines Botanikers eingegeben werden soll oder konkreter der Inhalt eines Blumenerkennungsbuches, geht der klassische KI-Ingenieur üblicherweise folgendermaßen vor. Zunächst legt er eine Tabelle an und trägt in die erste Spalte die Namen der Blumen ein, die der Computer kennen soll. In die übrigen Spalten trägt er die jeweiligen Merkmale ein, in jede Zeile die Merkmale der betreffenden Blume, z.B. Typ und Farbe der Blüte, Standort, Zeit der Blüte u.ä.m. Das Einspeichern der Tabelle erledigt der Computer weitgehend selbständig, sowohl das Umcodieren in Binärketten als auch das Zuweisen von Speicherplätzen. Nach erfolgter Einspeicherung wird sich irgendwo im Speicher der Binärcode einer Blume, z.B. einer Kornblume, und an einer anderen Stelle der Code der entsprechenden Farbe, also blau, befinden. Auf Anfrage kann der Computer beide Binärketten zu der Aussage „Kornblumen sind blau“ verbinden, oder er kann nach vorgegebenen Merkmalen eine Blume als Kornblume „erkennen“. Wie die klassische KI im Einzelnen vorgeht, ist in /3/ dargestellt.

Der alternative KI-Ingenieur wendet eine Methode an, die auf den Erkenntnissen der Neurophysiologie basiert. Er übernimmt die Rolle eines Lehrers und belehrt ein künstliches neuronales Netz solange, bis es in der Lage ist, Kornblumen zu erkennen. Dazu muss der Computer, den er belehren will, über ein Programm zur Simulation eines neuronalen Netzes verfügen, beispielsweise des in Bild 2 dargestellten Netzes. Im Weiteren kann nur sehr kurz auf neuronale Netze eingegangen werden. Interessenten steht eine umfangreiche Literatur zur Verfügung, u.a. die nichtmathematische Einführung /9/ und die mathematische Einführung /8/.

Die Kreise in Bild 2 stellen Neurone dar, genauer Binärneurone. Ein Binärneuron kann sich in zwei Zuständen befinden, die ich mit 0 und 1 bezeichne. Über die Pfeile wird der Zustand an die Nachfolgerneuronen weitergegeben. An jeder Pfeilspitze befindet sich eine sogenannte Synapse, die den Eingabewert mit einem Faktor zwischen 0 und 1, Gewicht genannt, multipliziert. Das Gewicht kann eingestellt werden. Ein Neuron geht in den Zustand 1 über, wenn die Summe aller Eingangswerte einen festgelegten Wert überschreitet.

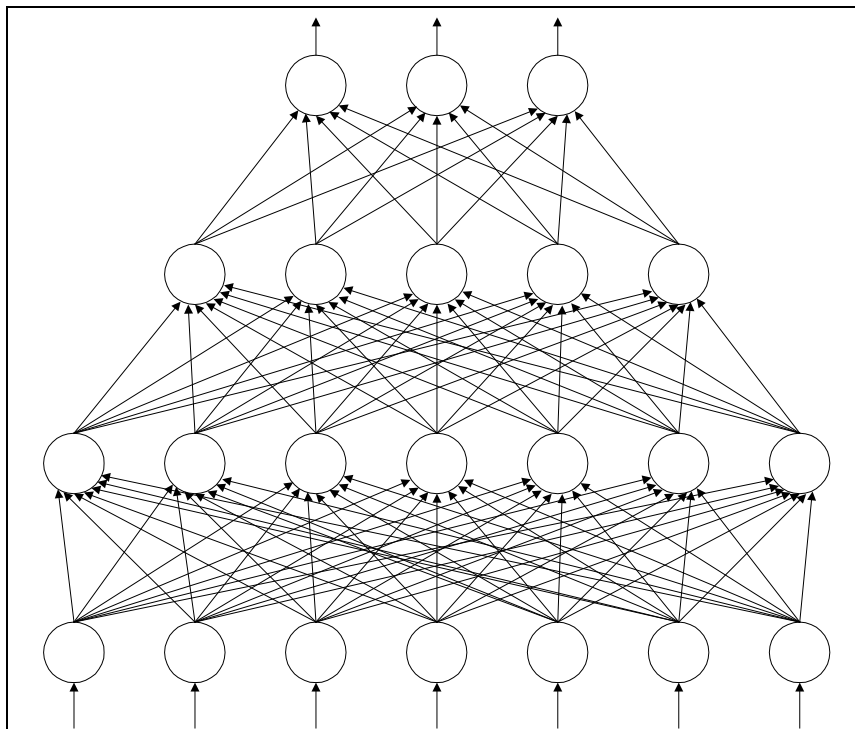


Bild 2. Beispiel eines künstlichen neuronalen Netzes, eines sog. Mehrschicht-Perzeptrons (nach /8/).

Derartige Netze - ich nenne sie Binärneuronennetze - besitzen eine erstaunliche Eigenschaft; sie können jede Zuordnung zwischen Eingabe- und Ausgabebinärketten erlernen, sofern sie ausreichend viele Neuronen und Verbindungen besitzen. Man stelle sich vor, dass das Netz von Bild 2 nicht nur 22, sondern viele Tausend Neuronen enthält und dass die Eingabeneuronen (Neuronen mit einem freien Eingabepfeil) und die Ausgabeneuronen (Neuronen mit einem freien Ausgabepfeil) keine Kette, sondern je ein Rechteck bilden, das Ein- bzw. Ausgabebild, sodass jedes Neuron einen Bildpunkt, sogenanntes Pixel, darstellt, wie bei einem Fernseher oder einer Digitalkamera. Wenn mittels des künstlichen Auges ein Eingabebild erzeugt wird, erscheint am Ausgang ein Ausgabebild, das auf dem Bildschirm angezeigt werden kann. Wie das Bild aussieht, hängt von den Synapsengewichten ab. Die Bilder können auch Texte sein.

Ein solches System kann lernen, Kornblumen zu erkennen. Das angelernte System gibt immer dann das Wort Kornblume oder irgendein anderes vereinbartes Symbol aus, wenn ihm irgendeine Kornblume gezeigt wird. Das Lernen erfolgt durch wiederholtes Zeigen von Kornblumen bei gleichzeitiger schrittweiser Änderungen der Synapsengewichte. Die Änderungen werden durch ein Programm vorgenommen, dem der *Lehrer* (Experimentator) mitteilt, ob die Antwort des Systems richtig oder falsch war (sogenanntes Lernen mit Lehrer). Die Synapsengewichte werden solange verändert, bis das System sich nicht mehr bzw. kaum noch irrt. Auf die gleiche Weise kann das System lernen, Kornblumen und Margeriten zu unterscheiden. Ein entsprechend angelerntes System kann Objekte erkennen und klassifizieren, es kann also zwei fundamentale Fähigkeiten des menschlichen ZNS simulieren. Im folgenden Abschnitt wird sich herausstellen, dass nach Kant diese beiden Fähigkeiten die primären Voraussetzungen für die Gewinnung von Erkenntnissen durch den Menschen sind.

4. Wie erlangt der Mensch Wissen?

4.1 Antwort Immanuel Kants auf die Frage *Wie erlangt der Mensch Wissen?*

4.1.1 Mechanismen der Erkenntnisgewinnung

Der KI-Ingenieur, der sein Wissen und seine Problemlösungen dem Computer mitteilen will, geht von gegebenem Wissen aus. Kant muss tiefer ansetzen, wenn er das zu erlangende Wissen nicht als bereits existent voraussetzen will. Um die Frage *Was kann ich wissen?* zu beantworten, muss er nach dem Mechanismus der Erkenntnisgewinnung und nach den Wurzeln fragen, aus denen der Mensch die Fähigkeit, Erkenntnisse zu gewinnen, letztendlich bezieht.

Kant hat sein Programm in der „*Kritik der reinen Vernunft*“ durchgeführt. Er war sich mit den Empiristen darin einig, dass der Mensch sein Wissen über die Welt aus den Sinnesempfindungen bezieht. Im Gegensatz zu ihnen war er jedoch überzeugt, dass Sinnesempfindungen allein nicht ausreichen. Denn um sie in Erkenntnisse zu überführen, muss der Mensch über entsprechende Fähigkeiten verfügen, die ihm *vor* jeder Sinnesempfindung, *vor* jeder Erfahrung, „*a priori*“ gegeben sein müssen. Um diese Fähigkeiten zu finden, analysierte er zunächst den *Mechanismus* der Erkenntnisgewinnung.

Kant ging davon aus, dass der Mensch sich aus der Mannigfaltigkeit der Sinnesempfindungen ein *einheitliches* Bild von der Welt erschafft. Es *muss* einheitlich sein, weil es auf ein einziges Zentrum bezogen ist, auf das denkende Ich, von dem das Bild *geschaffen* wird. Die mannigfaltigen *Sinnesempfindungen* müssen zu einem *sinnvollen* Bild zusammengefügt werden. Das führte Kant zu der Überzeugung, dass Erkenntnisgewinnung auf Synthese beruht, genauer, dass aus den Sinnesempfindungen schrittweise ein einheitliches Bild der Welt *synthetisiert* wird. Ich will auf die ersten beiden Syntheseschritte näher

eingehen.

Im ersten Syntheseschritt werden verschiedene Empfindungen zu Empfindungskomplexen verknüpft, wodurch *Objekte der Anschauung* entstehen. Dabei werden die einzelnen Objekte nicht erdacht, sondern *erschaut* und zwar aufgrund ihrer Lage oder Bewegung im Raum sowie anderer Merkmale wie z.B. Form und Farbe. Nebenbei sei bemerkt, dass jeder neugeborene Erdenbürger dadurch die Welt zu ergreifen und zu begreifen beginnt, dass er lernt, aus dem Strom der Sinneseindrücke bestimmte „Objekte der Anschauung“ herauszuheben. Diese Art des „Begrreifens“ setzt sich durch das ganze Leben hindurch fort.

Im zweiten Syntheseschritt verknüpft der *Verstand* Anschauungen zu Begriffen. Dabei geht der Verstand folgendermaßen vor. Wenn gleichartige Sinnesempfindungen an verschiedenen Objekten wahrgenommen werden (in verschiedenen Anschauungen enthalten sind), werden sie zu einem *Merkmalsbegriff* verknüpft, z.B. zum Merkmalsbegriff *weich* oder *blau*. Außerdem werden die betreffenden Objekte zu einem *Objektbegriff* verknüpft, z.B. zum Objektbegriff *Blume* oder *Kornblume*. Es ist also zwischen *Merkmalsbegriffen* und *Objektbegriffen* zu unterscheiden. In beiden Fällen sind die Begriffe von Anschauungen und letztlich von Sinnesempfindungen "abgezogen" (abstrahiert), sie sind abstrakter, *begrifflicher* Natur. Dieser Prozess kann sich fortsetzen, indem der Verstand aus Begriffen allgemeinere Begriffe abstrahiert. Objektbegriffe und Merkmalsbegriffe werden zu allgemeingültigen Aussagen oder Urteilen, wie Kant sie nennt, verknüpft, beispielsweise zum Urteil *Kornblumen sind blau*.

Die Begriffe und Urteile, die ein Mensch mit Hilfe seines Verstandes „synthetisiert“ hat, stellen nach Kant die *Erfahrung* des betreffenden Menschen dar. Die Urteile nennt er *Erfahrungsurteile* oder *Urteile a posteriori*. Begriffe und Urteile werden nicht wie die Objekte der Anschauung *erschaut*, sondern vom Verstand *erfunden*. Neu erfundene Urteile stellen Erkenntnisgewinn dar.

Außer den Begriffen und Urteilen muss der Verstand Regeln erfinden, nach denen abstrakte Begriffe mit konkreten Objekten der Anschauung verbunden werden, z.B. der Begriff *Hund* mit der Vorstellung eines Hundes. Solche Regeln nennt Kant Schemata. Interessanterweise sind die Begriffe Objekt und Schema ohne wesentlichen Bedeutungswandel in die Programmierungstechnik übernommen worden. Ein Schema wird dabei zu einem Computerprogramm oder Programmpaket. Beispielsweise kann das Schema eines Objektes *Kreis* ein Programm zum Zeichnen eines Kreises sein oder ein Programmpaket zur Berechnung von Elementen eines Kreises.

Das Ergebnis der Analyse des Mechanismus der Erkenntnisgewinnung sei noch einmal zusammengefasst: Erkenntnisgewinnung beginnt mit zwei Syntheseschritten. Im ersten Schritt werden Sinnesempfindungen zu Objekten der Anschauung verknüpft. Im zweiten Schritt verknüpft der Verstand Objekte der Anschauung zu Objektbegriffen und Merkmalsbegriffen und diese zu Urteilen (Erkenntnissen), und er stellt Regeln (Schemata) auf, nach denen Begriffe mit Objekten der Anschauung verbunden werden.

Eine Gegenüberstellung von Kants Denkresultaten mit den oben beschriebenen Ergebnissen der alternativen KI zeigt, dass die beiden Syntheseschritte, die Kant herausgearbeitet hat, identisch sind mit dem Erkennen und Klassifizieren von Objekten durch neuronale Netze. Als ich begann, mich ernsthaft mit Kants Erkenntnistheorie zu beschäftigen, war ich mit den Eigenschaften neuronaler Netze bereits vertraut. Das führte dazu, dass die Lektüre Kants bei mir einen frappierenden Aha-Effekt auslöste, und ich glaube, dass es jedem anderen an meiner Stelle ähnlich ergeht.

4.1.2 Wurzeln der Erkenntnis

Damit hat Kant den Synthesemechanismus der Erkenntnisgewinnung erklärt. Doch gräbt er noch tiefer, bis zu den Wurzeln des Syntheseprozesses. Er fragt, wie überhaupt das Erschauen

von Objekten und das Erfinden von Begriffen möglich ist. Seine Antwort lautet: durch die *reinen Anschauungs- und Urteilsformen*; sie sind die Wurzeln des menschlichen Erkenntnisvermögens. *Rein* bedeutet: „*ohne jede empirische Beimischung*“, d.h. jeder Erfahrung vorangehend.

Die *reinen Anschauungsformen* sind die Vorstellungen des Raumes, in dem sich Objekte befinden, und der Zeit, in der sie sich bewegen bzw. verändern. Sie sind die Voraussetzung für das Verknüpfung von Sinnesempfindungen zu Objekten der Anschauung. Raum und Zeit sind für Kant wie für Newton absolut und a priori gegeben.

Die *reinen Urteilsformen* sind allgemeingültige, absolut sichere Aussagen, an deren Wahrheit kein Zweifel möglich ist, beispielsweise die Aussage *Zwei Objekte der Anschauung sind entweder von gleicher Größe oder das eine ist größer als das andere*, oder die Aussage *Alles, was geschieht hat eine Ursache*. Die Begriffe, die in reinen Urteilsformen auftreten, nennt Kant reine Verstandesbegriffe oder – in Anlehnung an Aristoteles – Kategorien. Nach Kant sind z.B. *Größe (Quantität)* und *Ursache* Kategorien. Allem Denken liegen Kategorien zugrunde. Sie sind die Voraussetzung für das Verknüpfen von Objekten der Anschauung zu Begriffen. Sie lassen sich nicht weiter analysieren und folglich auch nicht definieren, weil sie selber für die Analyse notwendig wären, was in einen Zirkel führen würde. Die Kategorien nennt Kant auch die wahren Stammesbegriffe des reinen Verstandes.

Zu den Kategorien gelangte Kant über eine Analyse der allgemeinsten Aspekte, unter denen Objekten Attribute zugeordnet werden, beispielsweise Aspekte der Quantität oder der Kausalität (Ursache-Wirkungs-Relation). Konsequenterweise führt Kant auch den Begriff des reinen Schemas ein. Reine Schemata sind Vorschriften, „*wie reine Verstandesbegriffe auf Erscheinungen überhaupt angewendet werden können*“ (I/4, S. 237) Kant selbst gibt folgendes Beispiel. „*Das reine Schema der Größe aber (quantitatis) als eines Begriffs des Verstandes, ist die Zahl, welche eine Vorstellung ist, die die sukzessive Addition von Einem zu Einem (gleichartigen) zusammenfasst*“ (I/4, S. 241). Aus diesem Schema entwickelt sich durch Erweiterung die gesamte Mathematik. Sowohl die Definition als auch die Schlussfolgerung mag beim ersten Lesen ziemlich dunkel erscheinen. Dem Mathematiker ist aber sofort klar, dass beides sein Pendant in der Algorithmentheorie hat, speziell in der Erkenntnis, dass die Berechnung jeder rekursiven (effektiv berechenbaren) Funktion auf das Inkrementieren (Erhöhen um Eins) zurückführbar ist, eine Einsicht, zu der Kant etwa 150 Jahre früher gelangt ist, wenn er auch nicht die Sprache der modernen Algorithmentheorie verwendete. Auf dieser Erkenntnis beruhte Kants Überzeugung, dass die Formeln der Mathematik Urteile a priori sind; auf ihr beruht aber auch die Universalität des Computers.

Nach Kant sind nicht nur die Formeln der Mathematik, sondern auch die der Physik synthetische Urteile a priori. Zu diesem Schluss gelangte Kant durch folgenden, sehr verkürzt dargestellten Gedankengang. Physikalische Gesetze und ganz allgemein Naturgesetze in der Form mathematischer Formeln, haben drei Quellen: Beobachtung, Kausalität und Quantifizierung. Beobachtung beruht auf den Formen der reinen Anschauung Raum und Zeit; Kausalität ist eine der Kategorien, und durch vollständige Quantifizierung, d.h. durch zahlenmäßige Angabe aller Merkmale, werden physikalische Urteile zu mathematischen. Aus der Tatsache, dass alle drei Quellen dem Verstand a priori gegeben sind, zieht Kant den Schluss, dass derartige Naturgesetze a priori gelten, dass sie synthetische Urteile a priori sind. Aufgrund dieser Einsichten formulierte Kant den berühmten Satz: „*Der Verstand schöpft seine Gesetze nicht aus der Natur, sondern schreibt sie dieser vor*“ (I/5 § 36, S. 99).

Kant sah in dieser Erkenntnis eine „kopernikanische Revolution“, allerdings in umgekehrter Richtung. Kopernikus hatte den Menschen als Beobachter der Sternenbewegung *aus* dem Zentrum der Welt *herausgenommen*, um die beobachtete Sternenbewegung auf einfache Weise erklären zu können. Kant stellte das denkende Ich als Konstrukteur seines Weltbildes *in* den Mittelpunkt dieses Bildes *hinein*, um das Erkenntnisvermögen des Menschen erklären zu können. Das denkende Ich ist der Bildner des naturwissenschaftlichen

Weltbildes und damit ist es selber der Gesetzgeber der Natur. Dieser Schluss war für Kant folgerichtig, denn er war, wie dargelegt, zu der Einsicht gelangt, dass das „naturwissenschaftliche Weltbild“, nicht aus der Welt, sondern aus dem denkenden Ich stammt, und dass es mit absoluter Sicherheit richtig sein *muss*, denn es besteht aus Urteilen a priori. Doch macht Kant einen Unterschied zwischen Mathematik und Physik insofern, als die Physik sich „*dennoch auf Erfahrung und deren durchgängige Bestätigung*“ stützen muss (I/5, § 40, S. 108).

4.1.3 Korrekturen an Kants Schlussfolgerungen durch die moderne Physik

Es bleibt die letzte eingangs gestellte Frage zu beantworten, ob Kants Schlussfolgerungen den Ergebnissen der modernen Wissenschaft standhalten. Die Frage soll im Hinblick auf die Ergebnisse der Physik des 20. Jahrhunderts beantwortet werden. Kant war gerade angesichts der Erfolge der Physik seiner Zeit, vor allem der Erfolge der newtonschen Mechanik, insbesondere der Theorie der Planetenbewegung, fest davon überzeugt, dass zweifelsfreie und allgemeingültige naturwissenschaftliche Erkenntnisse möglich ist. Die logische Konsequenz und Detailliertheit, mit der Kant die Möglichkeit absolut sicherer Erkenntnis über die Welt nachweist, ist bewundernswert, doch halten die Voraussetzungen, von denen er ausgeht, den Ergebnissen der modernen Physik nicht stand, zumindest in den beiden folgenden Punkten.

1. Raum und Zeit haben ihren absoluten Charakter verloren. Richtig ist sicherlich, dass jeder Mensch alles, was er wahrnimmt, in Raum und Zeit wahrnimmt, d.h. dass alle Gegenstände und Ereignisse, die in sein Bewusstsein treten, in seiner Vorstellung in einen dreidimensionalen Raum und eine eindimensionale, vom Raum unabhängige Zeit eingebettet sind. Doch hat Einstein mit seiner Relativitätstheorie gezeigt, dass diese Vorstellung nur dann zutreffend ist, wenn die durch die Masseverteilung verursachte Krümmung des beobachteten Ausschnittes des vierdimensionalen „Weltraumes“ vernachlässigbar klein ist und wenn die beobachteten Geschwindigkeiten massebehafteter Objekte klein sind im Vergleich zur Lichtgeschwindigkeit. Folglich kann es keine *reinen* Anschauungsformen von Raum und Zeit geben.

2. Die Kausalität hat ihren deterministischen Charakter verloren. Zur Zeit Kants wurde das Kausalitätsprinzip im deterministischen Sinne verstanden, d.h. nicht nur die Möglichkeit eines Ereignisses, sondern der Wert jedes beteiligten Merkmals einschließlich des Zeitpunktes hat eine Ursache und ist durch diese eindeutig bestimmt (determiniert). Die Quantenmechanik hat gezeigt, dass das deterministische Kausalitätsprinzip nur gültig ist, solange Prozesse beobachtet werden, in die relativ große Energien involviert sind, genauer gesagt, solange Wirkungen und Drehimpulse beobachtet werden, die bedeutend größer sind als das plancksche Wirkungsquantum h . In der Quantenmechanik wird die *deterministische* Kausalität zu einer *probabilistischen*. Wenn beispielsweise ein Atom der Sorte Radium A in ein Atom der Sorte Radium B und ein Betaeilchen (Elektron) zerfällt, so ist die Ursache für die *Möglichkeit* des Zerfalls bekannt. Sie liegt in der Energiebilanz des Zerfalls. Unbekannt und nicht vorhersagbar ist dagegen der *Zeitpunkt*, in dem der Zerfall erfolgt. Für den Zeitpunkt gibt es keine Ursache, oder richtiger gesagt, hinsichtlich des Zeitpunktes hat der Begriff der Ursache keinen Sinn. Das Entsprechende gilt für die Richtung, in die das Betaeilchen ausgestrahlt wird. Das Kausalitätsprinzip ist also nicht dem allgemeinen Sinne gültig, wie Kant annahm. Folglich kann der Begriff der Ursache kein reiner Verstandesbegriff, keine Kategorie sein.

Beide Widersprüche zu Kants Schlussfolgerungen beruhen auf Beobachtungen, die außerhalb des Wirklichkeitsausschnittes liegen, der den *natürlichen* (unbewaffneten) Sinnesorganen des Menschen zugänglich ist und durch sie wahrgenommen werden kann. Zu Kants Zeiten gab es zwar schon das Fernrohr, doch machte dieses gerade diejenigen Beobachtungen erst möglich, die zur newtonsche Mechanik führten, an die Kant fest glaubte.

Mit Hilfe moderner physikalischer und astronomischer Beobachtungs- und Messmethoden konnte nachgewiesen werden, dass Kants reine Anschauungsformen und Verstandesbegriffe weder im Makrokosmos noch im Mikrokosmos konsequent anwendbar sind, sondern nur im sogenannten Mesokosmos, in dem wir leben, d.h. in dem sich unser Lebensprozess vollzieht und in dem wir unsere *Alltagserfahrungen* sammeln, der damit die Quelle unseres „natürlichen“ Wissens über die Natur ist. Dieses Wissen kann sich also nur auf den Mesokosmos beziehen, und nur auf ihn kann sich die „reine Naturwissenschaft“ im Sinne Kants beziehen.

Anders verhält es sich mit mathematischem Wissen, allgemein mit logischen Schlüssen. Sie scheinen wirklich unabhängig von jeder Erfahrung zu sein. Doch rührt sich der Verdacht, dass auch hier vielleicht der Schein trügt, dass auch unsere Logik auf Erfahrung beruht und dass es weder reine Anschauungsformen noch reine Urteilsformen gibt. Sollte vielleicht auch die Logik das Resultat einer Anpassung sein? Dieser Verdacht führt zu einem neuen, zum evolutionären Aspekt der Frage, wie Wissen ins Gehirn kommt.

4.2. Antworten der evolutionären und der neurophysischen Erkenntnistheorien auf die Frage *Wie erlangt der Mensch Wissen?*

4.2.1 Wurzeln der Erkenntnis

Für Kant gab es keine Möglichkeit, die Wurzeln der Erkenntnis, die reinen Anschauungs- und Urteilsformen, aus noch tieferliegenden Wurzeln abzuleiten. Für uns, die Nachfahren Darwins, gibt es diese Möglichkeit. Nach heutiger Auffassung ist das ZNS der Träger des Bewusstseins, des Denkens und Erkennens. Das ZNS aber ist ein Produkt der Evolution. Folglich liegen die Wurzeln der Erkenntnis in der Evolution. Das würde bedeuten, dass Erkenntnis und Wissen letztlich das Produkt eines Ausleseprozesses ist. Wie die Auslese vor sich geht, ergibt sich unmittelbar aus der Funktion des ZNS. Es dient nämlich der optimalen Steuerung des Verhaltens des Individuums. Optimal bedeutet: am besten an die Lebensumstände angepasst, sodass die Lebens- und Erfolgchancen maximiert werden. Das bessere ZNS hat also die größeren Chancen, in den Nachfahren „weiterzuleben“. Die Anpassung ist umso besser möglich, je sicherer das Individuum in jeder Lebenssituation die optimale Verhaltensweise aus vielen möglichen herausfinden und sich für diese entscheiden kann. Das wiederum ist umso besser möglich, je mehr das Individuum über seine Umwelt weiß, je umfassender, präziser und konsistenter sein inneres Modell der Welt ist und je genauer das Modell aus der Vergangenheit auf die Zukunft schließen lässt. Das Schließen auf die Zukunft ist am zuverlässigsten und genauesten, wenn es quantitativ erfolgt und Entscheidungen auf Berechnungen beruhen, m. a. W. wenn das Modell der Welt ein mathematisches ist.

Das Spitzenprodukt der Phylogenese ist das ZNS des *homo sapiens*, zumindest nach Meinung der Menschen. Es besitzt die Fähigkeit, mathematische Modelle der Welt, d.h. eine exakte Naturwissenschaft zu entwickeln. Deren Endziel ist ein geschlossenes mathematisches Modell der Welt. Es wird oft das *physikalische Weltbild* genannt, korrekter wäre die Bezeichnung *mathematisches Weltbild*. Es entwickelt sich durch schrittweise Synthese, durch Zusammenfügen von Wissensbruchstücken, sodass ein immer umfassenderes und in sich geschlosseneres Weltbild entsteht, etwa so wie beim Puzzlespiel aus vielen Bausteinen ein Bild entsteht.

Stellt man nun noch einmal die von Kant als unbeantwortbar zurückgewiesene Frage, woher die reinen Anschauungs- und Urteilsformen und die Kategorien stammen, so lautet die Antwort der Evolutionstheorie: Ihre Wurzeln liegen in fest verankerten Strukturen des ZNS, die sich im Laufe der phylogenetischen und ontogenetischen Evolution durch Anpassung an eine relativ stabile Umwelt herausgebildet haben. Die Strukturen können erblich angelegt sein

oder sich sehr früh in das ZNS des Individuums eingepägt haben, sodass sie durch keinerlei äußere Einflüsse mehr verändert werden können.

Diesen Strukturen entsprechen auf der Bewusstseinssebene fest eingepägte Denkstrukturen, z.B. die Vorstellung von Raum und Zeit, die Überzeugung, das alles was geschieht, eine Ursache hat, sowie alle rein logischen Schlüsse. Das ist die Erklärung des Satzes „Es gibt synthetischen Urteile a priori“ durch die *Evolutionäre Erkenntnistheorie*, die von dem Verhaltensforscher und Nobelpreisträger Konrad Lorenz begründet worden ist /7/, /10/

Man beachte, dass unser Gedankengang mehrmals zwischen den Welten 1 und 2 hin und her gesprungen ist, denn das Zentralnervensystem gehört der physischen Welt an, Wissen, Schließen und das bewusste Entwickeln von Modellen dagegen der Welt der Gedanken. Der Sprung ist insofern zulässig, als das ZNS Träger der Gedanken und damit des inneren Modells ist. Dem Sprung liegt die Vorstellung zugrunde, dass es eine Entsprechung zwischen ZNS-Struktur und Denken gibt. Auf sie wird im folgenden Abschnitt näher eingegangen.

4.2.2 Mechanismus der Erkenntnisgewinnung

Es folgen einige Überlegungen, deren Grundideen ich bereits in meinem Buch „Kausale Informatik“ /3/ skizziert habe. Ich wage es, diese Ideen im Weiteren etwas näher auszuführen, obwohl ich mit vielen Bedenken und Einwänden rechnen muss, zumal ich weder Philosoph noch Neurophysiologe, sondern Physiker und Informatiker bin.

Ich frage ganz unbefangen, wie der Mechanismus der Erkenntnisgewinnung aus biologischer, genauer aus neurophysiologischer Sicht aussehen könnte? Eine Erkenntnistheorie, die hierauf Antwort gibt, nenne ich neurophysisch, weil die Antwort auf der Erforschung der physischen Realität, insbesondere der Eigenschaften der neuronalen Struktur und Funktion des Gehirns basiert und nicht auf psychologischen Untersuchungen oder auf Introspektion wie bei Kant. Eine solche Theorie existiert noch nicht, doch ihre Konturen lassen sich bereits erkennen. Die Theorie hat gegenwärtig noch weitgehend hypothetischen Charakter. Einige Hypothesen möchte ich darlegen. Ob sie zutreffen, wird die Gehirnforschung früher oder später zeigen.

Vielleicht wird mir manch einer Reduktionismus vorwerfen. Mit diesem Vorwurf muss jeder rechnen, der ernsthaft und bedingungslos versucht, die Welt naturwissenschaftlich zu erklären, selbst wenn er weiß, dass ihm das nur unvollständig gelingen kann. Kant selber fordert dazu auf, die Welt, bis hinab zu ihren „Endursachen“, auf Physik - er sagt auf Mechanik – zurückzuführen, wenn er schreibt, dass „*die größtmögliche Bestrebung, ja Kühnheit in Versuchen sie [die Endursachen] mechanisch zu erklären, nicht allein erlaubt ist, sondern wir auch durch Vernunft dazu aufgerufen sind, unerachtet wir wissen, dass wir damit aus subjektiven Gründen der besonderen Art und Beschränkung unseres Verstandes niemals auslangen können,...*“ (/6/, § 82, S. 377).

Ich hatte dargelegt, dass Wissen im Computer binär codiert abgespeichert wird. Die Frage ist erlaubt, ob die Natur ein ähnliches Prinzip anwendet. Das würde bedeuten, dass es im Gehirn codierende Zustände geben muss. Ob es sie tatsächlich gibt, ist wahrscheinlich, aber nicht absolut sicher. Angenommen, es gibt sie, welcher Natur müssen sie dann sein? In Analogie zum Computer könnte man annehmen, dass auch das Gehirn binär codiert und dass jedes Binärzeichen durch ein Neuron dargestellt wird. Auf dieser Idee beruhen sehr erfolgreiche Simulationsexperimente der alternativen KI-Forschung, in denen Netze aus künstlichen Neuronen mit zwei stabilen Zuständen – ich hatte sie oben Binärneuronen genannt - untersucht werden. Doch die Idee entspricht nicht den biologischen Tatsachen, denn das natürliche Neuron besitzt nur einen einzigen stabilen Zustand, den Ruhezustand. Es kann zwar aus dem Ruhezustand in einen angeregten Zustand überführt werden, doch kehrt es innerhalb von etwa einer hundertstel Sekunde in den Ruhezustand zurück. Auf diese Weise kann das Neuron elektrische Impulse generieren; man sagt, es *feuert*. Der Impuls wird über

eine Ausgabelitung, Axon genannt, an nachgeschaltete Neuronen weitergeleitet. Ein Neuron feuert, wenn es innerhalb kurzer Zeit ausreichend viele Impulse von vorgeschalteten Neuronen empfangen hat. Wären die elektrischen Impulse Lichtblitze, so wäre das Gehirn von einem ständigen Blitzen oder Funkeln erfüllt, ähnlich wie ein Glühwürmchenschwarm. Künstliche Neurone dieser Art nenne ich Impulsneurone. Auch Impulsneuronennetze sind seit einigen Jahren Gegenstand der Untersuchung. Es kann als experimentell und theoretisch erwiesen angesehen werden, dass sich in Impulsneuronennetzen, ebenso wie in Binärneuronennetzen, stabile kollektive Anregungszustände ausbilden können. Doch handelt es nicht um statische, sondern um dynamische Stabilität, d.h. um stationäre Zustände..

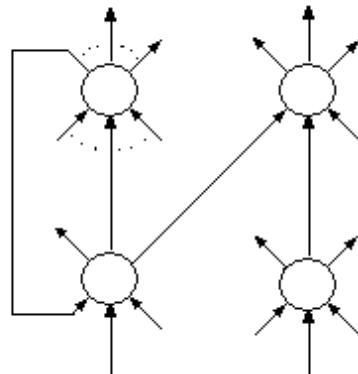


Bild 3

Ein sehr einfacher derartiger Zustand ist das ununterbrochene Kreisen eines Impulses in einem Neuronenring. Bild 3 zeigt vier der vielen Milliarden Neuronen des ZNS. Zwei von ihnen bilden einen Ring. Die Impulse werden in Pfeilrichtung weitergegeben. Damit es in dem Ring zum Kreisen eines Impulses kommt, muss ein einziger Impuls ausreichen, um das Nachfolgerneuron zum Feuern anzuregen. In der Regel muss ein Neuron viele Impulse empfangen haben, bis es feuern kann. Da sich stabile dynamische Zustände tatsächlich ausbilden können, liegt folgende Hypothese nahe:

Codierungshypothese: Als codierende Zustände dienen im ZNS elektrische, ausreichend stabile dynamische (stationäre) Anregungszustände.

Im Gegensatz zum Computer muss das Gehirn selbst einem codierenden Zustand eine Bedeutung zuordnen. Wie ist das möglich? Die Antwort kann gegenwärtig wiederum nur eine Hypothese sein. Ich nenne sie Brückenhypothese, weil sie eine Brücke zwischen den beiden Welten in Bild 1 schlägt und damit eine Brücke zwischen Psychologie und Physiologie.

Brückenhypothese: Einem Bewusstseinsinhalt entspricht ein ausreichend stabiler dynamischer (stationärer) Anregungszustand im ZNS. Denkprozessen entsprechen Folgen stationärer Anregungszustände, oder etwas verkürzt: Mentalen Zuständen und Prozessen entsprechen neuronale Zustände und Prozesse. In diesem Sinne ist Pfeil 2 in Bild 1 unten zu interpretieren. Ein Anregungszustand, der einem Bewusstseinszustand zugeordnet ist, wird als dessen neuronaler Code bezeichnet. Gestützt wird die Hypothese durch experimentelle Ergebnisse der Gehirnforschung, wonach Denkprozesse mit neuronalen Anregungsprozessen einhergehen und bestimmten Bewusstseinsinhalten bestimmte Anregungszustände entsprechen können. Voraussetzung dafür scheint zu sein, dass ausreichend viele Neuronen bestimmter Gebiete im Kortex ausreichend lange und ausreichend intensiv synchron feuern.

Damit ist die eigentliche Natur der Beziehung zwischen neuronalen Anregungen und Bewusstsein freilich noch nicht erklärt. Zudem schließt die Brückenhypothese nicht aus, dass einem Bewusstseinsinhalt viel komplexere und nicht nur neuronale Strukturen entsprechen.

Nach der Brückenhypothese entspricht einem Denkprozess eine Folge codierender Zustände einschließlich der Übergabeprozesse der Anregung von Zustand zu Zustand. Wenn die Brückenhypothese zutrifft, ergibt sich die philosophisch schwerwiegende

Schlussfolgerung, dass der *Logik* eines Denkprozesses die *Kausalität* des zugeordneten neuronalen Prozesses entspricht. Die Brückenhypothese wird heute von vielen Menschen akzeptiert, nicht nur von Gehirnforschern, sondern auch von Natur- und Geisteswissenschaftlern, selbst von Vertretern der Kirche. Es ist zu beachten, dass die Brückenhypothese nicht die Reduktion des Denkens und des Bewusstseins auf Neurophysiologie und letztlich auf Physik beinhaltet, denn die Pfeile 2 und 3 in Bild 1 stellen keine physikalische Wechselwirkung, sondern lediglich eine Entsprechung dar. Was sich dahinter verbirgt, bleibt völlig offen.

Die Codierungs- und die Brückenhypothese reichen nicht aus, um zu erklären, wie Wissen in das Gehirn hineinkommt und wie es gespeichert ist, sodass es auch noch nach Jahren ins Bewusstsein treten kann. Das Einspeichern muss offenbar ein *Einprägen* in die graue Materie sein, eine Strukturbildung im neuronalen Netz. Die Strukturen selber stellen noch keine codierenden Zustände dar, doch müssen sie so beschaffen sein, dass sich in ihnen codierende Zustände ausbilden können. Voraussetzung dafür ist eine passende Strukturierung des ZNS. Damit wäre die Frage, wie Wissen ins Gehirn gelangt, folgendermaßen zu beantworten: Strukturierungshypothese: Das Erwerben von Wissen über die Welt beruht auf Selbststrukturierung der grauen Materie des Gehirns.

Für diese Hypothese sprechen viele Experimente und Erkenntnisse der Gehirnforschung. Es ist eine gesicherte Tatsache, dass sich die neuronale Struktur des ZNS unter dem Einfluss der Anregungsprozesse, die in ihm ablaufen, verändert. Die Strukturänderung kann durch Auf- und Abbau von Leitungen erreicht werden oder durch Änderung der Durchlassfähigkeit der Synapsen; die in Bild 3 – ebenso wie in Bild 2 – durch Pfeilspitzen dargestellt sind. Die Durchlässigkeit war oben als Synapsengewicht bezeichnet worden. Für den Mechanismus der Gewichtsänderung hat Donald Hebb vor vielen Jahren eine Hypothese aufgestellt, die inzwischen als zutreffend nachgewiesen werden konnte. Danach erhöht sich das Gewichte einer Synapse immer dann, wenn gleichzeitig das postsynaptische Neuron (das Neuron, auf das der betreffende Pfeil in Bild 3 zeigt) feuert und die Synapse vom präsynaptischen Neuron (Neuron, von dem der Pfeil ausgeht) einen Impuls empfängt. Das Strukturieren nach diesem Mechanismus wird als hebbisches Lernen bezeichnet.

Das Strukturieren erfolgt sowohl in der Wachstumsperiode des Gehirns als auch später und zwar durch empfangene Sinnesempfindungen und durch Aktivitäten der Neuronen unabhängig von Sinnesempfindungen, z.B. während des Nachdenkens oder im Schlaf. William H. Calvin nimmt an, dass die Strukturierung nach den Prinzipien der darwinschen Evolutionstheorie erfolgt /1/.

Fasst man die Verbindungsstruktur als ein Wegenetz auf, so entspricht der neuronalen Netzstrukturierung eine Veränderung des Wegenetzes durch Entstehen und Verschwinden von Wegen sowie durch das Austreten vorhandener Wege.

Die zunehmende Strukturierung des ZNS als Träger des inneren Modells der Welt bedeutet eine Erweiterung des Modells, was man sich mit Hilfe der Puzzlemetapher veranschaulichen kann. Sie war ursprünglich auf der Ebene des Bewusstseins angesiedelt und bedeutete dort, dass sich das innere Modell der Welt aus Wissensbruchstücken „zusammenpuzzelt“. Auf neuronaler Ebene bedeutet sie, dass die Anregungszustände, die den Wissensbruchstücken entsprechen, sich aneinander anpassen, dass sie *kompatibel* werden; das heißt, dass die Zustände sich nach erfolgter Anpassung entweder gegenseitig anregen können oder dass sie in einen einzigen Zustand zusammenfließen können. Dies bedeutet wiederum auf der Bewusstseins Ebene, dass entweder der eine Bewusstseinsinhalt mit dem anderen assoziiert bzw. aus ihm abgeleitet werden kann, oder dass zwei Bewusstseinsinhalte zusammenfließen, z.B. dass zwei Begriffe zu einem einzigen Begriff konvergieren. Das Zusammenfließen bedeutet die Herausbildung allgemeinerer Begriffe und eines umfassenderen, in sich geschlosseneren Modells der Welt. In der Glühwürmchenmetapher entspricht dem Zusammenfließen zweier Zustände die Synchronisierung zweier Bereiche, von denen jeder

für sich synchron, die untereinander aber zunächst asynchron blinken. Diese Vorstellungen führen zu folgender Hypothese:

Konvergenzhypothese: Wenn zwei codierende Zustände ausreichend häufig gleichzeitig oder sequentiell angeregt werden, ändert sich die Struktur des betreffenden Netzbereiches, bis die Zustände kompatibel werden, d.h. bis ein Zustand den anderen aktivieren kann oder bis beide Zustände einen gemeinsamen Zustand bilden, auch dann, wenn sie sich zunächst gegenseitig stören, sodass sie sich nicht gleichzeitig oder sequentiell vollständig aktivieren lassen. Auch für diese Hypothese, insbesondere für das synchrone Feuern einer großen Anzahl von Neuronen, sprechen viele Experimente der Gehirnforschung. Das Zusammenfließen und allgemein das Kompatibelwerden von Anregungszuständen nenne ich neuronale Konvergenz. Sie stellt das neuronale Pendant nicht nur der Synthese bei Kant, sondern auch des dialektischen Prinzips bei Hegel dar, wonach These und Antithese in der Synthese „aufgehoben sind“.

Nach diesen Vorstellungen beruht die Entwicklung des physikalischen (besser des mathematischen) Modells der Welt auf neuronaler Konvergenz, und naturwissenschaftliche Erkenntnisse in mathematischer Form entstehen letztlich durch Selbststrukturierung des ZNS mittels Synapsenänderungen während der Arbeit des Gehirns am inneren Modell der Welt. Die Strukturierung führt dazu, dass Erfahrungsurteile mit mathematischen Urteilen durch neuronale Konvergenz kompatibel werden. Kompatibilität bedeutet hier, dass die quantifizierten Beobachtungen zur Interpretation eines Kalküls geworden sind. Der Motor dieser Selbststrukturierung des ZNS könnte ein Belohnungssystem im ZNS sein, das auf erfolgreiche Konvergenz mit der Festigung der erreichten Kompatibilität reagiert und im Bewusstsein das Gefühl der Befriedigung oder Freude auslöst, das jedem Denker und Erfinder bekannt ist.

Die neuronale Konvergenz ist der Kern der neurophysischen Erkenntnistheorie, genauso wie die schrittweise Synthese der Kern der Erkenntnistheorie Kants ist. Unter Einbeziehung der Brückenhypothese kann damit die eingangs gestellte Frage, wie Wissen in die graue Materie gelangt, folgendermaßen beantwortet werden:

Wissen gelangt dadurch in die graue Materie, dass Sinnesempfindungen das neuronale Netz in einer solchen Weise strukturieren, dass sich stabile Anregungszustände ausbilden können, denen im Bewusstsein Aussagen über die Welt zugeordnet sind. Wird ein codierender Zustand angeregt, tritt das zugeordnete Wissen ins Bewusstsein und wird zu *gedachtem* Wissen, zum *Gedanken*. Unter dem Einfluss von Sinnesempfindungen bilden sich Codierende Zustände und die zugehörigen Gedanken *gemeinsam* aus. Ein Gedanke und sein neuronaler Code bilden eine Einheit. Dies ist meine Antwort auf das Leib-Seele-Problem. Eine Lösung des Problems ist es freilich nicht, denn es bleibt offen, worin die Einheit besteht. Eine physikalische Einheit zu postulieren, würde einen radikalen Reduktionismus und eine Negierung des Denkens und des Geistes als eigenständiger Bestandteile der Welt bedeuten. Man beachte, dass dieser Standpunkt an keiner Stelle der vorangehenden Überlegungen bezogen wurde.

5. Was spricht für die Theorie?

Auf der Basis obiger Hypothesen lassen sich die Grundprinzipien einer Erkenntnistheorie entwickeln, welche die Frage, wie Wissen in der grauen Materie entsteht, beantworten könnte. Doch wirft die skizzierte Theorie viele Fragen auf, und viele Einwände können gegen sie erhoben werden, vor allem der Vorwurf, dass sie weitgehend spekulativ ist und nur heuristischen Wert besitzt, solange die vier Hypothesen nicht verifiziert sind. Das aber kann die experimentelle Technik gegenwärtig nicht leisten.

Doch sind die Hypothesen nicht aus den Luft gegriffen, sondern basieren auf Ergebnissen der Gehirnforschung bzw. der alternativen KI. Wie bereits erwähnt, zeigen Untersuchungen

am aktiven Gehirn, dass Denkprozesse mit Anregungsprozessen im ZNS einhergehen, und unterstützen damit die Brückenhypothese. Simulationsexperimente mit künstlichen neuronalen Netzen zeigen, dass diese Netze gerade diejenigen Eigenschaften besitzen, die von den restlichen drei Hypothesen vorausgesetzt werden:

- In neuronalen Netzen können sich stabile Anregungszustände ausbilden (Voraussetzung der Codierungshypothese).
- Stabile Zustände können sowohl auf dem Wege der Fremdstrukturierung (mit „Lehrer“) als auch auf dem der Selbststrukturierung „erlernt“ werden.
- In Impulsneuronennetzen neigen zwei Zustände nach mehrmaliger gleichzeitiger oder sequentieller Anregung dazu, sich gegenseitig anzuregen oder einen gemeinsamen Zustand zu bilden (Voraussetzung der Konvergenzhypothese).

Da die Voraussetzungen für Codierung, Strukturierung und Konvergenz durch künstliche neuronale Netze erfüllt werden, ist zu erwarten, dass sich der Erkenntnisprozess früher oder später wird simulieren lassen. Das wäre ein Argument dafür, dass die Theorie die entsprechenden Gehirnprozesse im Prinzip richtig wiedergibt. Ein zweites Argument ist die oben erwähnte Tatsache, dass das Erkennen und das Klassifizieren von Objekten die beiden fundamentalen Prozesse der Erkenntnisgewinnung sind und zwar sowohl der natürlichen Erkenntnisgewinnung durch den Menschen (Kant) als auch der künstlichen Erkenntnisgewinnung durch neuronale Netze (alternative KI)

6. Resümee

Kants Schlussfolgerung, dass es naturwissenschaftliche Erkenntnis a priori gibt, hat sich als unhaltbar herausgestellt. Kant hat zwar Recht wenn er sagt, dass die Gesetze der Naturwissenschaften vom Menschen gemacht, d.h. mathematisch formuliert werden. Aber die Natur muss sich nicht unbedingt und nicht streng nach ihnen richten und künftige Erfahrungen können zeigen, dass sie korrigiert werden müssen, dass neue Formeln erfunden werden müssen. Für die newtonsche Mechanik ist das tatsächlich eingetreten. Aber wird der Mensch die Formeln, die er erfunden hat, auch verstehen? Angesichts der Relativitätstheorie und der Quantenmechanik scheint so, als würde die Frage *Was können wir wissen?* durch die Frage *Was können wir verstehen?* verdrängt.

Falls auch Logik auf Erfahrung beruht, möglicherweise auf phylogenetischer Erfahrung, dann gibt es auch keine mathematische Erkenntnis a priori, und Platon hat Recht, wenn man das eingangs angeführte Zitat ergänzt zu *Ich weiß, dass ich nichts absolut sicher (a priori) weiß.* Man könnte die Frage stellen, ob vielleicht eine andere als die angeborene Logik notwendig ist, um die moderne Physik „verstehen“ zu können. Vielleicht bildet sich eine neue Logik aus, wenn die Menschen ständig mit Lichtgeschwindigkeit im Raum – evtl. im virtuellen Raum – umherkutschieren oder wenn sie sich tagaus tagein mit Quantencomputern unterhalten.

Ich möchte mit einem Zitat enden, das dem Buch „Der Teil und das Ganze“ von Werner Heisenberg /2/ entnommen ist. Das Buch enthält ein Gespräch mit der Überschrift „Quantenmechanik und Kantsche Philosophie“. Heisenberg lässt seinen Freund Carl Friedrich von Weizsäcker das Gespräch mit dem Resümee beenden, *„dass die Kantsche Analyse der Erkenntnis echtes Wissen, nicht nur unbestimmtes Meinen enthält, und dass sie überall dort richtig bleibt, wo lebendige Wesen, die reflektieren können, zu ihrer Umwelt in die Beziehung treten, die wir vom menschlichen Standpunkt aus ‚Erfahrung‘ genannt haben. Aber auch das Kantsche ‚a priori‘ kann später aus seiner zentralen Stellung verdrängt und Teil einer sehr viel umfassenderen Analyse des Erkenntnisprozesses werden. Es wäre an dieser Stelle sicher falsch, naturwissenschaftliches oder philosophisches Wissen mit dem Satz ‚Jede Zeit hat ihre eigene Wahrheit‘ aufweichen zu wollen. Aber man muss sich doch gleichzeitig vor Augen halten, dass sich mit der historischen Entwicklung auch die Struktur des menschlichen*

Denkens ändert. Der Fortschritt der Wissenschaft vollzieht sich nicht nur dadurch, dass uns neue Tatsachen bekannt und verständlich werden, sondern auch dadurch, dass wir immer wieder neu lernen, was das Wort ‚Verstehen‘ bedeuten kann.“

Literatur

- /1/ Calvin, William H.: *Die Sprache des Gehirns. Wie in unserem Bewusstsein Gedanken entstehen*. München; Wien: Carl Hanser Verlag 2000
 - /2/ Heisenberg, Werner: *Der Teil und das Ganze. Gespräche im Umkreis der Atomphysik*. München: R. Piper & Co Verlag, 1971
 - /3/ Jungclaussen, Hardwin: *Kausale Informatik. Einführung in die Lehre vom aktiven sprachlichen Modellieren durch Mensch und Computer*. Wiesbaden: Deutscher Universitäts-Verlag, 2001
 - /4/ Kant, Immanuel: *Kritik der reinen Vernunft*. Reclams Universalbibliothek Nr. 6461-70.
 - /5/ Kant, Immanuel: *Prolegomena*. Reclams Universalbibliothek Nr. 2468-70.
 - /6/ Kant, Immanuel: *Kritik der Urteilskraft*.. Reclams Universalbibliothek Nr. 1026-30.
 - /7/ Lorenz, Konrad: *Die Rückseite des Spiegels*.. München; Zürich: Piper, 1973
 - /8/ Nauck, Detlef; Frank Klawonn; Rudolf Kruse: *Neuronale Netze und Fuzzy-Systeme*. Braunschweig, Wiesbaden 1996
 - /9/ Spitzer; Manfred: *Geist im Netz. Modelle für Lernen, Denken und Handeln*, Heidelberg; Berlin: Spektrum Akademischer Verlag, 2000
 - /10/ Vollmer, Gerhard: *Was können wir wissen?* Stuttgart: Hirzel Verlag 1988
-