

# Über die Möglichkeit künstlichen Bewusstseins

-

## Grundlegende philosophische Überlegungen und Probleme

Eine Hausarbeit von Florian Boge  
zum Seminar  
*Ausgewählte Texte zum Problem des Bewusstseins*  
Von apl. Prof. Dr. Ulrich Baumgarten

# Inhalt

Einleitung .....	S. 03
1. Das Problem des Bewusstseins im Allgemeinen .....	S. 04
1.1. Merkmale des Bewusstseins .....	S. 04
1.2. Physikalismus .....	S. 05
1.3. Dualistische Positionen .....	S. 07
1.4. Physikalistische Kriterien für Bewusstsein .....	S. 08
1.5. Versuch einer eigenen Charakterisierung des Bewusstseins ...	S. 10
2. Der Spezialfall künstlichen Bewusstseins .....	S. 13
2.1. Über die generelle Möglichkeit künstlichen Bewusstseins .....	S.13
2.2. Eine kurze Erläuterung des Konnektionismus .....	S. 15
2.3. Das Problem der Verifikation künstlichen Bewusstseins .....	S. 16
2.4. Zusammenfassung und Kontemplation der angestellten Überlegungen .....	S. 18
Fazit .....	S. 22
Verwendete Literatur .....	S. 24
Eidesstattliche Versicherung.....	S. 25

## Einleitung

Wie kann es überhaupt sein, dass es Bewusstsein gibt? Und was genau ist eigentlich Bewusstsein? Und wem kann man berechtigter Weise ein Bewusstsein zuschreiben? Diese Fragen quälen Philosophen seit tausenden von Jahren und in der jüngeren Zeit auch zunehmend Wissenschaftler aus anderen Disziplinen. Zu diesen Disziplinen zählen neben der Psychologie und den Kognitions- und Neurowissenschaften auch die Computerwissenschaften – zumindest der Teil davon, der sich mit künstlicher Intelligenz befasst. Denn die Frage nach der Möglichkeit eines künstlichen Bewusstseins ist vielleicht noch viel interessanter, als die nach der Möglichkeit künstlicher Intelligenz.

Um uns dieser hochinteressanten Frage, die vor dem Hintergrund des Erstarkens reduktionistisch-materialistischer Theorien, auf Basis neuerlicher empirischer Erkenntnisse und der fortschreitenden Entwicklung der Computer-Technologie, immer mehr an Aktualität gewinnt, überhaupt nähern zu können, müssen wir uns zunächst mit Theorien des Bewusstseins im allgemeinen auseinandersetzen. Ich werde also im ersten Teil dieser Arbeit einige Grundlegende Konzeptionen davon, was überhaupt Bewusstsein ist, bzw. was es ausmacht, vorstellen, um dort direkt einige Grundprobleme zu thematisieren.

Danach werde ich zu den speziellen Problemen, die sich im Hinblick auf künstliches Bewusstsein ergeben, übergehen. Es wird dabei allerdings nicht möglich sein, über theoretische Vorüberlegung und philosophische Spekulation hinauszugehen, da eine echte bewusste Maschine gegenwärtig noch eine Utopie zu sein scheint. Anhand der Beispiele, die als Exemplare künstlicher Intelligenz vorgeschlagen wurden – entsprechend dem *Turing Test* (Vgl. Turing 1950) – können allerdings ein paar Fragen zum Zusammenhang von künstlicher Intelligenz und künstlichem Bewusstsein erläutert und diskutiert werden.

Ich werde also mit dieser Arbeit versuchen, die Grundprobleme künstlichen Bewusstseins zu skizzieren und dabei Stellung für die mir am plausibelsten erscheinenden Positionen demgegenüber, sowie zu Bewusstseinstheorien im Allgemeinen zu beziehen und jeweils Argumente dafür zu liefern. Beginnen wir nun also im ersten Abschnitt mit der kritischen Betrachtung verschiedener Theorien des Bewusstseins, sowie dem Versuch einer eigenen Skizzierung dessen, was ‚Bewusstsein‘ bedeuten könnte.

# **1. Das Problem des Bewusstseins im Allgemeinen**

Bewusstsein ist etwas, mit dem wir im Alltagsleben umgehen, als ob es eine Selbstverständlichkeit sei. In alltagssprachlichen Kontexten kommt der Begriff des Bewusstseins in Bezug auf Sachverhalte vor (*Umweltbewusstsein*), als Zustandsbeschreibung für Menschen und Tiere (*bei Bewusstsein sein*) oder als Angabe über die aktuelle Kenntnis bestimmter Tatsachen (*sich einer Tatsache bewusst sein*). Was alle diese Verwendungen vereint, ist die Tatsache, dass sie sich auf ein lebendes Individuum beziehen, in den meisten Fällen einen Menschen. Aber auch bei vielen Tierarten tut sich der menschliche Hausverstand nicht schwer, ihnen Bewusstsein zuzuschreiben.

Doch was berechtigt uns eigentlich dazu irgendwem außer uns selbst Bewusstsein zuzuschreiben? Die Ähnlichkeit zu uns selbst? Das Verhalten? Hinter dieser Zuschreibung versteckt sich meiner Auffassung nach immer ein abduktiver Analogieschluss der Form: *Objekt x zeigt Verhalten y. Ich zeige Verhalten y. Verhalten y wird bei mir durch Bewusstseinsakte hervorgerufen. Also wird Verhalten y auch bei x durch Bewusstseinsakte hervorgerufen.*

Es scheint die beste, nächstliegende Erklärung für uns zu sein, wenn sich ein unserer Außenwelt angehöriges Objekt auf diese und jene Weise verhält und wir dieses Verhalten von uns selbst kennen und es als bewusstes Verhalten identifizieren können, dass dieses Objekt ebenfalls ein Bewusstsein besitzt.

Doch die Basis für dieses Schlussmuster stellt selbstverständlich das Vorhandensein eines eigenen Bewusstseins dar. In dem Moment, in dem ich diese Zeilen hier verfasse, ist es für mich persönlich undenkbar kein Bewusstsein zu haben, denn die Eindrücke, die ich erfahre, den Monitor meines Computers, die Tastatur unter meinen Fingern und den Straßenlärm von draußen, erlebe ich *bewusst*. Mein Bewusstsein *besteht* sozusagen darin, dass ich all das erlebe. Mit welchen Begriffen erfasse ich jedoch ein derartiges Phänomen, das zunächst in *sicherer* Weise lediglich einer so begrenzten Sphäre – meinem ganz privaten Inneren – angehört? Sehen wir uns dazu an, was einige Philosophen als Grundlegende Merkmale des Bewusstseins vorgeschlagen haben.

## **1.1. Merkmale des Bewusstseins**

Dass Privatheit ein Merkmal für das Bewusstsein sein mag, liegt gewissermaßen auf der Hand und wird auch mit Sicherheit bei genauerer Betrachtung von kaum jemandem in Frage

gestellt werden. Manchmal glauben wir zwar in Alltagskontexten genau zu wissen, was im Bewusstsein eines anderen Menschen oder Tieres vor sich geht, was dieser oder dieses fühlt und denkt. Diese Alltagsintuition wird zuweilen sogar durch eintreffende Voraussagen gestützt, wie sich das entsprechende Wesen in gewissen Situationen verhalten wird, wann die betreffende Person anruft, oder auch durch Zustimmung dieser bei dem eigenen Versuch, ihre Gefühle zu beschreiben. Doch müssen wir hier aufpassen, nicht unberechtigter Weise von der Kenntnis typischer Verhaltensweisen eines Lebewesens auf die Kenntnis des Inhaltes seines Bewusstseins zu schließen! Nur weil wir möglicherweise ein reliables Wissen über das Standardvokabular haben, dass eine Person nutzt um ihre Gefühle zu beschreiben, oder eine äquivalente Beschreibung im eigenen Vokabular hervorbringen können, wissen wir nicht, wie sich die Gefühle dieser Person *wirklich für sie anfühlen!* Und nur weil wir auf eine für uns selbst nicht unbedingt transparente Weise abschätzen können, wann eine Person sich melden wird, wissen wir nicht, wie es ist, *die Gedanken dieser Person zu denken*<sup>1</sup>.

Thomas Nagel hat diesen Punkt in einem Aufsatz von 1974 anhand des Beispiels einer Fledermaus deutlich gemacht, indem er darauf verwies, dass wir selbst bei genauer Kenntnis des Wahrnehmungsapparates einer Fledermaus doch keine Ahnung haben, wie es ist, diese Fledermaus *zu sein*; das heißt, wir haben insbesondere keine Ahnung, wie es ist, den Wahrnehmungs- und Bewegungsapparat einer Fledermaus zu haben, also zum Beispiel „die Umwelt mit einem System reflektierter akustischer Signale aus Hochfrequenzbereichen [wahrzunehmen].“ (Nagel 1974, 264).

Ein ganz entscheidender Punkt, den wir aus Nagels Aufsatz aber gewinnen können, der uns etwas vom recht abstrakten und nicht ganz klar definierten Begriff der Privatheit wegbringt, ist die Erkenntnis, was wir mit Bewusstsein in gewisser Hinsicht meinen können: „Die Tatsache, dass ein Organismus überhaupt bewusste Erfahrung hat, heißt im wesentlichen, *dass es irgendwie ist, dieser Organismus zu sein.*“ (Nagel 1974, 262)

## **1.2. Physikalismus**

In Nagels Auffassung stellt dieses Verständnis von Bewusstsein, welches unserem Alltagsgebrauch des Wortes recht gut entsprechen mag, ein Problem für bestimmte philosophische Theorien des Bewusstseins dar, vor allem für den *Physikalismus*. Unter Physikalismus versteht man eine ontologische Theorie, die besagt, dass letztlich alles

---

<sup>1</sup> Es ist allerdings unklar, ob dies nicht bloß kontingenter Weise, also in Abhängigkeit von gegenwärtigen technischen Möglichkeiten, bzw. bekannten Methoden so ist. Es ist durchaus denkbar, dass es irgendwann oder bereits irgendwo die Möglichkeit geben *könnte*, die Bewusstseinsinhalte eines anderen zu erfahren. Die Realität oder Realisierbarkeit dieser Möglichkeit unterliegt allerdings soweit der reinen Spekulation.

physischer Natur ist (siehe z.B. Beckermann 2000, 1). Damit steht der Physikalismus im starken Gegensatz zu klassischen, dualistischen Theorien cartesischer Tradition, welche ich später noch genauer erläutern will.

Das Problem in der Erklärung des Bewusstseins ist für den Physikalismus nach Nagel nun das folgende:

„Wenn der Physikalismus verteidigt werden soll, müssen phänomenologische Eigenschaften [also Bewusstseinsinhalte – d.Verf.] selbst physikalisch erklärt werden. Wenn wir aber ihren subjektiven Charakter untersuchen, scheint so etwas unmöglich zu sein. Der Grund dafür ist, dass jedes subjektive Phänomen mit einer einzelnen Perspektive verbunden ist; und es scheint unvermeidlich, dass eine objektive physikalische Theorie von dieser Perspektive abstrahieren wird.“ (Nagel 1974, 263)

Um es mit anderen Worten zu sagen kann in Nagels Auffassung der Physikalismus niemals das *Charakteristische* des Bewusstseins erfassen, weil er gerade von diesem abstrahiert. Was das Bewusstsein ausmacht, ist die Tatsache, dass es nicht bloß einer subjektiven Perspektive unterliegt, sondern gewissermaßen die subjektive Perspektive eines Organismus *ist*. Eine Theorie wie der Physikalismus, die danach strebt, von jeder subjektiven Perspektive zu abstrahieren um eine möglichst objektive Beschreibung der Welt zu erlangen, schließt diese Klasse von Phänomenen also von vorne herein aus und scheint somit etwas auszulassen.

Ich denke, dass Nagel hiermit sicher einen wichtigen Punkt anspricht und sicherlich ein gewisses Erklärungsproblem für den Physikalismus aufwirft. Dennoch ist dies meiner Meinung nach kein K.O.-Kriterium für eine physikalistische Ontologie. Denn nur weil eine *Perspektive* von einer Theorie ausgelassen wird, heißt das nicht, dass die Theorie falsch ist. Möglicherweise wird hier jedoch eine Grenze für die Leistungsfähigkeit des Physikalismus aufgezeigt. Da der Physikalismus den Anspruch erhebt, *alle* Phänomene der Welt physikalisch zu erklären, muss er auch eine physikalische Erklärung für das Bewusstsein bieten<sup>2</sup>. Dies ist zwar möglich aber nur aus einer abstrahierenden, objektivierenden Perspektive, die etwas Charakteristisches auslässt. Man könnte also sagen, eine physikalistische Ontologie könne vermutlich alle *Phänomene* erfassen, jedoch nicht *alle Perspektiven auf alle Phänomene* (und damit unter Umständen nicht alle Aspekte eines

---

<sup>2</sup> Auf eine solche darf man bekannter Weise am ehesten aus den Neurowissenschaften hoffen. Dass man neurowissenschaftliche Erklärungsmodelle als physikalisch ansehen kann, hängt mit der plausiblen Möglichkeit zusammen, dass bestimmte wissenschaftliche Beschreibungsebenen (Biologie, Neurowissenschaften) auf anderen (Physik) supervenieren. Diese Beziehung wird auch *humesches Mosaik* genannt (Vgl. z.B. Nolan 2005, 27 ff.).

jeweiligen Phänomens). Dieses Problem haben jedoch auch andere Ontologien und damit ist es kein Spezialproblem des Physikalismus.

### **1.3. Dualistische Positionen**

Selbstverständlich ist eine physikalistische Ontologie nicht der einzige – wenngleich der für diese Arbeit interessanteste – Standpunkt, den man bezüglich der Erklärung des Bewusstseins einnehmen kann. Eine traditionellere philosophische Position gegenüber der Seele des Menschen und damit auch gegenüber seinem Bewusstsein, ist der Substanzdualismus, wie er in erster Linie auf René Descartes zurückgeht (Descartes 1641, 73 ff.). Dieser Position zur Folge besteht der Mensch aus zwei von einander verschiedenen Substanzen: der Seele als denkendem Ding und dem Körper als ausgedehntem Ding. Dieser Ansatz ist gänzlich konträr zu dem des Physikalismus, da letzterer versucht die Welt substanzial einheitlich, also *monistisch* zu erklären. Der Substanzdualismus hingegen sieht Gründe zu einer grundlegenden Trennung zwischen mentalen und physikalischen Phänomenen.

Zwar muss man zugeben, dass der Substanzdualismus eine ganze Reihe von philosophischen Problemen mit sich bringt, wie zum Beispiel das Interaktionsproblem, also die Frage, wie Seele und Körper auf einander einwirken können, oder das Problem kausaler Überdetermination, also die Tatsache, dass für viele Phänomene plötzlich eine körperliche und eine seelische oder geistige Ursache gleichermaßen in Frage kommen (was vielen Theorien der Kausalität widerspricht). Auch kann man festhalten, dass aufgrund dieser Probleme und aufgrund der massiven Fortschritte bestimmter Wissenschaftszweige im vergangenen Jahrhundert, der Substanzdualismus sehr in den Hintergrund getreten ist, bzw. weniger Anklang in der gegenwärtigen Philosophie findet. Aber dennoch muss man auch hier sagen, dass diese Tatsachen keine K.O.-Kriterien darstellen, sondern lediglich Abschwächungen des, oder Probleme für den Dualismus. Zudem kann man alltäglich beobachten, dass die Annahme einer substanzialen Verschiedenheit der Seele vom Körper der Intuition vieler Menschen entspricht und zum Beispiel auch in religiöse Vorstellungen stark integriert ist.

Zwischen dem monistischen Physikalismus und dem klassischen Substanzdualismus gibt es noch diverse andere Positionen. Der Eigenschaftsdualismus geht z.B. davon aus, dass „sich die *phänomenalen Eigenschaften* unseres Bewusstseins nicht auf physikalische Vorgänge im Gehirn reduzieren lassen [...]“. (Löbel 2008, 6). Hier wird also nicht mehr davon ausgegangen, dass es zwei grundverschiedene Substanzen gibt, sondern dass bewusste

Organismen lediglich Eigenschaften haben, die in einer rein physikalischen Beschreibung ihrer nicht erfasst werden können.

Doch insofern, als die vorgestellten Standpunkte sich lediglich mit dem ontologischen Status des Bewusstseins befassen, liefern sie noch keinen Aufschluss darüber, wann wir einem Organismus (außer uns selbst) überhaupt Bewusstsein *zuschreiben* können. Der Substanzdualismus ist in dieser Hinsicht besonders wenig hilfreich. Denn wenn Bewusstsein etwas ist, das nicht einmal von physikalisch mess- und erfassbaren Eigenschaften eines Organismus abhängt, sondern lediglich vom Vorhandensein einer nicht räumlichen Substanz, kann man sich letztlich beliebige Kriterien einfallen lassen, die Indikatoren für Bewusstsein darstellen. Sehen wir uns daher einmal an, wie man aus physikalistischer Perspektive mit diesem Problem umgehen könnte.

#### **1.4. Physikalistische Kriterien für Bewusstsein**

Ein Vorschlag, welche Kriterien wir für die Annahme des Vorhandenseins eines Bewusstseins einsetzen könnten, kommt vom erklärten Physikalisten Robert Kirk. In seiner Auffassung bedeutet ein Bewusstsein zu haben für ein System, dass es *Information für* dieses System gibt (Vgl. Kirk 1995, 645). Zwar liefert Kirk keine explizite Definition für seinen Informationsbegriff, gibt aber implizit Aufschluss über dessen Bedeutung indem er seine Rolle im Zusammenhang mit einem weiteren zentralen Begriff klärt: dem Begriff der *Grundausrüstung* wobei mit Grundausrüstung zunächst nur eine Menge mit einander verknüpfter Kompetenzen zur Informationsverarbeitung gemeint ist.

„Systeme, die die ‚Grundausrüstung‘ besitzen, sammeln nicht nur Information, die Information für sie ist: Sie speichern auch Information, sie haben Bedürfnisse oder Ziele, sie sind mit einer Art von Bewertung ihrer Situation und einer Art von Entscheidungsfindung beschäftigt und sie initiieren und kontrollieren ihr Verhalten auf der Grundlage von Information.“ (Kirk 1995, ebd.)

Nun mag diese Auflistung von Eigenschaften bezüglich des Umgangs mit Information nicht unglaublich aufschlussreich sein, im Hinblick auf die Frage was Bewusstsein ist, aber dennoch haben wir hier eine Liste mehr oder weniger gut überprüfbarer Kriterien, die wir verwenden können, um einem Organismus oder einem System Bewusstsein zu unterstellen. Und genau das war es ja, was wir uns erhofft hatten.

Zwar kann man sagen, dass dieses Problem nicht auf den Bereich der Theorien des Bewusstseins beschränkt ist und auch andere Theorien, wie zum Beispiel Wahrheitstheorien

oft entweder gute, intuitionsnahe Definitionen für das zu erfassende Phänomen, oder aber ein Kriterium zur Feststellung desselben liefern, meist aber nicht beides in einem. Wir können aber in diesem Falle sogar eine vage Charakterisierung des Bewusstseins aus Kirks Ausführung abstrahieren: Bewusstsein zu haben heißt im physikalistischen Sinne (nach Kirk), ein System zu sein, das eine gewisse Grundausstattung besitzt, die das Erfassen von und den Umgang mit Information ermöglicht.

Die Ausführungen von Kirk gehen in gewisser Weise mit einigen anderen Überlegungen zum Bewusstsein überein, wie zum Beispiel in einem Artikel von Giorgio Buttazzo von 2001 dargelegt. Buttazzo spricht eine mögliche Perspektive auf das Bewusstsein an, in der das Haben eines Bewusstseins in einer Stufenfunktion der Gehirnkomplexität beschrieben werden kann. Erst ab einer gewissen Komplexität des Gehirns scheinen Bewusstseinsvorgänge vorhanden zu sein; das heißt es gibt einen kritischen Schwellenwert der Komplexität von Gehirnen, ab denen diese ein bewusstes Erleben hervorzubringen scheinen. Dieser Schwellenwert der Komplexität könnte der von Kirk genannten Grundausstattung entsprechen, die es dem System möglich macht, Information in der von ihm genannten Weise zu verarbeiten und zu benutzen, ja, gewissermaßen das haben von Information zu *erleben*.

Was den genauen Aufbau solcher Systeme bzw. entsprechend komplexer Systeme angeht, so gibt es in den Kognitionswissenschaften zwei rivalisierende Paradigmen, von denen das eine in den letzten Jahren in den Vordergrund getreten ist. Die Rede ist hier von *konnektionistischen* Modellen gegenüber *klassischen*. Während im klassischen Verständnis von Kognition der Umgang mit symbolischen Strukturen und die Verarbeitung von Symbolen im Vordergrund stehen, orientiert sich der Konnektionismus eher am Aufbau des Gehirns. Hierbei interagieren – wie die verschiedenen Module eines Gehirns – verschiedene Einheiten mit einander und leiten aufgenommene Inputs in Form von elektrischer Ladung weiter (siehe z.B. Stillings et al. 1995).

Ohne an dieser Stelle im genauen auf die Details eingehen zu wollen, können wir festhalten, dass für einige Kognitionswissenschaftler der *materielle Aufbau* eines kognitiven Systems eine entscheidende Rolle spielt, für andere hingegen lediglich die systematische Struktur der *Verarbeitungsprozesse*. Wir werden im späteren Verlauf noch einmal auf diese beiden Modelle zu sprechen kommen. Für unsere Frage nach der Möglichkeit eines künstlichen Bewusstseins bleibt jedenfalls festzuhalten, dass man zwei Aspekte als relevant für Kognition und damit vermutlich auch für Bewusstseinsfähigkeit halten kann. Zudem kann man bei diesen beiden Konzeptionen sehen, dass die eine – der Konnektionismus – eigentlich nur eine physikalistische Sichtweise zulässt, während die andere sich ontologisch neutral

verhält. Eine klassische Konzeption, die sich mit dem Umgang mit Symbolen beschäftigt, könnte z.B. auch die Struktur des *Geistes* als einer nicht-physischen Entität erfassen.

An dieser Stelle haben wir bereits einen großen Schritt zu unserem Hauptthema gemacht, da man sich Anhand dieser beiden Konzeptionen die Frage der künstlichen Reproduzierbarkeit der gesamten mentalen Sphäre<sup>3</sup> stellen kann. Bevor wir uns nun aber mit detaillierteren Fragen zu künstlichem Bewusstsein auseinandersetzen, möchte ich noch einmal ein paar eigene Gedanken zum Bewusstsein im Allgemeinen festhalten.

### **1.5. Versuch einer eigenen Charakterisierung des Bewusstseins**

Die jüngere Debatte über die Möglichkeit einer vollständigen physikalischen Beschreibung des Universums in besonderem Hinblick auf Bewusstseinsphänomene regt einen immer wieder dazu an, darüber nachzudenken, ob das Bewusstsein ein ontologisches Extra ist, oder nicht. Das heißt, der Physikalismus steht und fällt ja gewissermaßen damit, ob wir es schaffen, das Bewusstsein mit physikalischer Terminologie zu erfassen oder nicht. Meines Erachtens müssen wir uns zunächst Gedanken darüber machen was für einen *ontologischen Charakter* das Bewusstsein überhaupt hat, wie dies ja auch in der Geschichte der Philosophie immer wieder geschehen ist.

Eine Feststellung, die mir bei der Kontemplation der Bewusstseinsfragen in den Sinn gekommen ist, ist die Tatsache, dass Bewusstsein immer etwas relationales zu sein scheint. Wenn ich im alltäglichen Gebrauch von Bewusstsein rede, so meine ich damit immer ein Bewusstsein *von* etwas. So zum Beispiel in Begriffen oder Phrasen wie *Umweltbewusstsein* oder *das ist mir bewusst*, aber auch *Selbstbewusstsein* exemplifiziert. Doch auch in Verwendungen des Bewusstseinsbegriffes, die diesen relationalen Charakter nicht explizit machen, wie die Formulierung *bei Bewusstsein sein*, glaube ich eine gewisse Relationalität erkennen zu können; bei Bewusstsein zu sein bedeutet, sich selbst und die Umwelt *aktiv wahrnehmen* und *erleben* zu können.

Ein weiterer Punkt der Auffällt ist, dass Bewusstsein sowohl eine *reflexive*<sup>4</sup>, als auch *irreflexive* Relation sein kann. Ich kann zwar ein Bewusstsein meiner selbst haben und dies wird auch die meiste Zeit über, in der ich bei Bewusstsein bin der Fall sein; ich meine jedoch auch im Nachhinein schon mehrfach festgestellt zu haben, dass ich bei intensiver Kontemplation gewisser Sachverhalte keinerlei Bewusstsein meiner selbst (das heißt z.B. meines körperlichen Befindens, meiner äußeren Erscheinung, meines Wirkens auf andere, der

---

<sup>3</sup> Hier ontologisch neutral zu verstehen; also lediglich als ein thematischer Aspekt von lebenden Organismen, nicht als spezieller ontologischer Bereich (wie z.B. im Substanzdualismus angenommen).

<sup>4</sup> Im rein logischen Sinne gemeint und nicht speziell auf die Tätigkeit der aktiven Selbstreflexion bezogen.

Tatsache meiner Denkaktivität usw.) hatte. Ich denke jeder, der schon einmal eine intensive theoretische Auseinandersetzung mit einem bestimmten Thema vollzogen hat, kann sich vorstellen, was ich meine. Mit anderen Worten ist die Reflexivität des Bewusstseins nach meinem Verständnis im Grunde also bloß *dispositional*, wenn auch die meiste Zeit über aktual.

Nun, ob variierend zwischen reflexiv und irreflexiv oder nicht, Bewusstsein scheint jedenfalls etwas relationales zu sein. Doch das ist noch kein sonderlich aussagekräftiges Charakteristikum für eine ontologische Komponente der Welt. Was Hingegen – ganz in Nagels Sinne – das Bewusstsein auszeichnet, ist die Einzigartigkeit jeden Bewusstseins gegenüber jedem anderem Bewusstsein.

Darüber hinaus, so denke ich, ist es kein Zufall, dass die Begriffe *Bewusstsein* und *Wissen* etymologisch verwandt sind! Habe ich ein Bewusstsein von etwas so bedeutet dies, dass ich einen *epistemischen Zugang* zu etwas habe. Und ein Zugang bedeutet, wie gesagt, eine Relation zwischen mir und diesem Etwas. Wenn wir nun alles zusammenfügen, so haben wir meiner Auffassung nach eine relativ griffige Definition dessen, was Bewusstsein bedeuten könnte:

Ein System oder Organismus hat ein Bewusstsein gdw. dieses System oder dieser Organismus einen *einzigartigen epistemischen Zugang zu der es oder ihn umgebenden Welt und zu sich selbst* hat.

Wie genau ein solches System, das dieses Kriterium erfüllt, aufgebaut ist, bzw. welche Mindestanforderungen es erfüllen muss, ist eine andere Frage. Und ebenso ist es eine andere Frage, ob z.B. der Physikalismus klären kann, welche Kriterien dies sind, also wie ein solches System (bzw. ein solcher Organismus) im Detail aufgebaut sein muss. Die Frage danach, was der Bewusstseinsbegriff bedeutet, kann meines Erachtens nur auf die philosophische Weise geklärt werden, wie ich es soeben versucht habe und auch andere Philosophen wie Nagel vor mir dies getan haben.

Meiner Auffassung nach kennzeichnet die Einzigartigkeit der Relation, die ein Bewusstsein bedeutet, nicht notwendiger Weise ein ‚ontologisches Extra‘ gegenüber dem Physikalismus; es besteht die Möglichkeit, dass, gegeben der individuellen materiellen Komposition und Genese eines Systems oder Organismus, sein einzigartiger epistemischer Zugang zur Welt ebenfalls immer schon gegeben ist

Allerdings führt uns die obige Frage nach den *physikalischen Kriterien* wieder an den Rand unseres Hauptthemas zurück. Denn wenn man sich die Frage stellt, welche Kriterien erfüllt sein müssen, damit man vom Vorhandensein eines Bewusstseins ausgehen kann, lässt sich direkt im Anschluss die Frage stellen, wer alles diese Kriterien erfüllt und ob es möglich ist, dass künstliche Systeme Bewusstseinsfähigkeiten entwickeln können. Wenden wir uns daher als endlich dieser Problemstellung zu!

## **2. Der Spezialfall künstlichen Bewusstseins**

Wenn im folgendem von künstlichem Bewusstsein die Rede ist, so soll damit das Bewusstsein einer von Menschen gebauten Maschine – also eines Roboters oder Computers – gemeint sein. Die Fragen, die sich um die Existenz eines solchen Bewusstseins ranken, stellen insofern einen Spezialfall dar, als vermutlich die allerwenigsten Menschen bezweifeln, dass ihre Mitmenschen ein Bewusstsein haben, und auch bezüglich der meisten höher entwickelten Tiere der Konsens mittlerweile recht groß ist, dass diese eine Form von Bewusstsein besitzen. Bei Computern oder Robotern sollte es aber eher umgekehrt sein; die wenigsten Menschen gehen hier offenbar davon aus, dass es bereits bewusste Maschinen gibt. Zwar mag es vorkommen, dass man im Alltag mit einem Computer oder einem anderen Gerät redet, oder umgeht, als sei es ein bewusstes Wesen, doch dieses Verhalten resultiert vermutlich aus der bloßen Gewohnheit, dass viele Objekte unserer Alltagswelt mit denen wir interagieren (Menschen und andere Tierarten) ein Bewusstsein zu haben scheinen, als aus dem tatsächlich vorhandenen Glauben an ein Bewusstsein in dem betreffenden Apparat.

Es beschäftigen uns also bei diesem Thema gegenwärtig noch ganz andere Fragen, als bloß die Frage nach dem Verständnis davon, was Bewusstsein ist, oder die, aufgrund welcher Beobachtungen wir Bewusstsein unterstellen. Die letztere dieser beiden Fragen ist zwar vielleicht von großer Relevanz, wenn wir uns nach künstlichem Bewusstsein fragen, ihr muss aber *systematisch* gesehen zumindest die Frage vorausgehen, ob es *überhaupt* möglich ist, dass eine Maschine Bewusstsein besitzt oder entwickelt.

### **2.1. Über die generelle Möglichkeit künstlichen Bewusstseins**

Wenn wir versuchen uns die genannte Frage vor dem Hintergrund dessen, was wir bis hierhin über das Bewusstsein festgehalten haben, zu beantworten, so müssen wir uns zunächst darüber im Klaren sein, welche Art von Möglichkeit wir meinen. Von der rein logischen Möglichkeit kann nicht die Rede sein. Denn ganz offenbar führt uns die Vorstellung eines künstlichen Bewusstseins nicht in einen logischen Widerspruch (und meines Erachtens auch nicht in einen analytischen). Die Art von Möglichkeit, die uns hier interessiert muss zunächst diejenige sein, die oft *metaphysische Möglichkeit* genannt wird. Das bedeutet, dass wir uns die Frage stellen müssen, ob die Welt im Grunde so strukturiert ist, dass diese Struktur es zulässt, dass neben biologischen Organismen noch andere Systeme ein Bewusstsein – also meinem Verständnis nach eine epistemische Relation zu sich und ihrer Umwelt – haben.

Hier kommen unsere ganzen Vorüberlegungen aus dem ersten Teil erneut ins Spiel! Wenn der klassische Dualismus die richtige Metaphysik sein sollte, es also eine immaterielle Substanz gibt, die den Bereich des Mentalen erschöpft, so wäre die Frage nach der Möglichkeit künstlichen Bewusstseins eine ganz andere, als in einer physikalistischen Ontologie:

Im ersten Falle müssten wir uns nach der Verankerung des immateriellen Bewusstseins in künstlichen Apparaten fragen. Wir müssten uns die Frage stellen, wie etwas nicht-materielles Einfluss auf Materie nehmen kann und ob dabei die spezifische materielle Komposition des jeweiligen Objektes eine Rolle spielt, bzw. wenn ja, dann welche.

Im zweiten Falle hingegen fällt ein Teil dieser Fragestellung weg. Es bliebe lediglich die Frage übrig, ob das spezifische Material, aus dem ein System besteht, entscheidend dafür ist, ob dieses System ein Bewusstsein entwickeln kann, oder ob es lediglich auf seine Komplexität, seine spezifische Struktur, das Zusammenspiel seiner Komponenten usw. ankommt.

Ich werde mich im Folgenden auf die Fragen, die in einer physikalistischen Perspektive aufgeworfen werden, konzentrieren. Dies liegt vor allem daran, dass es in der Gegenwärtigen Debatte kaum noch Anhänger des Substanzdualismus zu geben scheint und die Forschung ebenfalls fast komplett unter Präsupposition eines physikalistischen Paradigmas geschieht. Diese Tatsache hängt aber wiederum in großem Maße damit zusammen, dass konkrete Forschung über immaterielle Entitäten mit empirischen Mitteln kaum oder sogar gar nicht möglich ist. Zudem verkompliziert die Annahme einer zweiten, substanziell verschiedenen Klasse von Entitäten, die zu der der materiellen Entitäten und ihrer Zusammenhänge hinzukommt, das gesamte Unterfangen ohne einen wirklichen Nutzen zu bringen. Es handelt sich also um einen Fall in dem Ockhams Rasiermesser Anwendung finden könnte oder sogar sollte<sup>5</sup>, da wir nicht nur ohne die Annahme einer immateriellen Substanz auskommen, sondern ohne diese vielleicht sogar besser fahren. Wir müssen uns im Folgenden also auch noch über die Frage der *physikalischen Möglichkeit*, das heißt, der konkreten Realisierbarkeit eines künstlichen Bewusstseins in unserer als rein physikalisch angenommenen Welt, Gedanken machen.

Um nun etwas konkreter zu werden, müssen wir uns wieder auf das besinnen, was wir innerhalb des physikalistischen Paradigmas zum Bewusstsein festgehalten haben. Wenn zum Beispiel das Bewusstsein tatsächlich einer Stufenfunktion der Gehirnkplexität entspricht,

---

<sup>5</sup> Ich möchte hier erneut betonen, dass dies keine *Falsifikation* des Substanzdualismus darstellt! Die Welt könnte dennoch dual aufgebaut sein, auch wenn dies für uns pragmatische Schwierigkeiten birgt. Praktikabilität oder Einfachheit kann kein Argument für die Korrektheit einer These sein – es handelt sich hierbei meines Erachtens lediglich um einen *pragmatischen Wert* einer Theorie, wenn diese einfacher oder besser verwendbar ist, als eine andere.

dann liegt die Frage nahe, ob nicht auch andere, dem Gehirn in ihrer Funktionsweise ähnliche Systeme ab einem gewissen Grad der Komplexität ein Bewusstsein besitzen oder entwickeln könnten. Die vielversprechendsten Versuche ein solches System zu konstruieren, sind offenbar konnektionistischer Art. Wir sollten uns daher den Konnektionismus noch einmal etwas näher ansehen.

## **2.2. Eine kurze Erläuterung des Konnektionismus**

Konnektionistische Modelle des menschlichen Verstandes, oder des Verstandes im Allgemeinen, sind, wie bereits erwähnt, in gewisser Weise am Aufbau des Gehirns orientiert. In der gegenwärtigen Kognitionsforschung erfreuen sich Systeme, die auf den Grundprinzipien des Konnektionismus basieren immer größerer Beliebtheit, da ihre Verarbeitungskapazitäten den kognitiven Fähigkeiten eines Menschen deutlich näher kommen, als Systeme, die auf dem klassischen Paradigma basieren (Vgl. z.B. Schlagel 2004, 9 ff.). Konnektionistische Systeme bestehen aus einer Reihung von verschiedenen Einheiten mit verschiedenen Funktionen.

Diese Einheiten sind auf verschiedene Schichten aufgeteilt: eine zur Eingabe von Werten (zum Beispiel 1 und 0), eine oder mehrere versteckte Schichten zur weiteren Verarbeitung der Eingaben und eine zur Ausgabe von Resultaten der verarbeiteten Impulse (also beispielsweise akustische Signale, Zeichen auf einem Monitor usw.). Desweiteren gibt es Verbindungen zwischen den Einheiten verschiedener Schichten, die *gewichtet* sind, was bedeutet, dass sie mit verschiedenen Ausgangswerten, ausgedrückt durch positive oder negative Zahlen oder die Zahl Null, versehen sind. Diese Zahlen werden mit dem numerischen Wert des jeweiligen Inputs multipliziert und die Ergebnisse werden addiert um die Stärke des Outputs zu bestimmen (Vgl. Stillings et al 1995, 65). Entscheidend für die Art des Outputs ist natürlich, welche Symbole oder Signale den Input-Einheiten zugeordnet sind.

Eine Besonderheit ist, dass die einzelnen Einheiten innerhalb einer Schicht nicht miteinander verbunden sind, sondern parallel arbeiten. Die Verbindungen die zwischen den Einheiten bestehen können zudem sehr unterschiedlich sein und Bestimmte Einheiten können mehrere, oder auch gar keine Verbindung zu bestimmten anderen Einheiten haben. Erst dieses komplexe Zusammenspiel aus verschiedenen Schichten und gewichteten Verbindungen bringt eine Art Verarbeitung von Signalen hervor, die manche Wissenschaftler bereits als *kognitiv* verstehen:

„The cognitive task that a network performs is determined in part by the arrangement of the units and connections, in part by the representations that are employed at the input and output levels, and in part by the weight of the connections.” (Stillings et al 1995, 67)

Ohne nun noch weiter auf die Details einzugehen, können wir festhalten, dass bei dieser Art von System das Netzwerk als Ganzes die verarbeitende Leistung hervorbringt, was man in gewisser Weise mit der komplexen Funktionsweise eines Gehirns vergleichen kann. Konnektionistische Systeme würden also als potentielle Kandidaten von bewussten künstlichen Systemen in Frage kommen, wenn man von der bereits mehrfach angesprochenen Hypothese ausgeht, dass die Komplexität des Systems das entscheidende Merkmal für Bewusstseinsfähigkeit ist.

Doch diese Behauptung steht und fällt mit der Annahme über den Zusammenhang zwischen der Komplexität eines Systems und seinen Fähigkeiten! An diesem Punkt kommen wir zu einem ganz entscheidenden Problem, das wir bereits im Hinblick auf Bewusstsein im Allgemeinen angesprochen hatten, das sich und aber bei Maschinen noch um so deutlicher stellt: Das Problem, wie wir überhaupt erkennen können, ob etwas – in diesem Falle: eine Maschine – ein Bewusstsein hat oder nicht.

### **2.3. Das Problem der Verifikation künstlichen Bewusstseins**

Das Problem der Verifikation stellt in diesem Kontext ein recht zentrales, praktisches Problem der Philosophie künstlichen Bewusstseins dar, das in einer gewissen Tradition steht. Der bereits in der Einleitung dieser Arbeit angesprochene Turing Test<sup>6</sup> ist ein weltbekannter Lösungsvorschlag für das Verifikationsproblem bei künstlicher *Intelligenz*. Alan Turing hatte 1950 vorgeschlagen, dass man einen Menschen in einen abgetrennten Raum setze und ihn schriftlich mit einem Computer und einem Menschen kommunizieren lasse, wobei er das Ziel habe, herauszufinden welcher der Gesprächspartner Mensch und welcher Maschine sei. Gelingt dies aber nach einer längeren Befragungsperiode nicht, so sei der Computer als denkfähig, also intelligent einzustufen.

Mit Sicherheit kann man diverse Kritikpunkte an diesem Verfahren hervorbringen. Zum Beispiel ist es fragwürdig, ob hier wirklich von Gedanken und Intelligenz die Rede sein kann; also ob unsere Konzepte davon, was Intelligenz und Gedanken sind, sich wirklich auf die Berechnungsprozesse eines Computers anwenden lassen. Zudem kann man sich die Frage stellen, ob es wirklich der Computer ist, der denkt, oder ob nicht nur seine Konstrukteure ein

---

<sup>6</sup> Im Folgenden abgekürzt durch *TT*.

besonders hohes Denkvermögen gezeigt haben, indem sie eine so raffinierte Maschine konstruiert haben. Eine Frage, die sich aber meines Erachtens direkt anschließt und für uns hier besonders interessant ist, ist die Frage, ob, wenn man einen Computer nun tatsächlich als intelligent einstuft, man ihm auch gleichzeitig ein Bewusstsein zusprechen muss.

Zunächst ist dies eine analytische Frage. Ist Denkvermögen immer an ein Bewusstsein gekoppelt, hängen also unsere Konzepte von ‚Denken‘ und ‚Bewusstsein‘ so eng zusammen, dass Bewusstsein eine notwendige Bedingung für Denken ist? Angenommen, dies wäre nicht der Fall und es stellte sich heraus, dass die Mehrheit der Menschen nicht unbedingt voraussetzt, dass etwas Denkendes auch ein Bewusstsein hat. Dann bleibt die *ontologische* Frage zurück, ob nur bewusste Systeme *tatsächlich* Gedanken hervorbringen können, oder ob es auch so etwas wie ‚bewusstseinslose Denker‘ gibt.

Mit der jeweiligen Antwort auf diese beiden Fragen steht und fällt die Relevanz des *TT* für unsere Bewusstseinsproblematik. Würde sich die Losgelöstheit des Bewusstseins vom Denken irgendwie herausstellen lassen und würden wir den *TT* als valide Methode zur Überprüfung der Denkfähigkeit einer Maschine akzeptieren, so bräuchten wir ein Äquivalent des *TT*, das speziell auf Bewusstsein ausgerichtet wäre. Doch wie könnte so ein Test bloß aussehen? Vielleicht wäre hier deutlich mehr Interaktion mit einer entsprechenden Maschine vonnöten; doch dazu müsste diese auch das Erscheinungsbild eines Menschen haben, da man ja sonst sofort erkennen würde, dass es sich um eine Maschine handelte. Ganz offensichtlich ist es jedenfalls noch schwieriger einen solchen Test für das Vorhandensein eines Bewusstseins zu modellieren, als für irgendeine Form von Intelligenz.

Eine bemerkenswerte Feststellung bezüglich des ursprünglichen *TT* ist jedenfalls, dass es heutzutage zwar keine Maschine gibt, die den Test in Bezug auf Konversation bestanden hat, einige Maschinen auf anderen Gebieten jedoch erstaunliche Leistungen hervorgebracht haben, durch die man sie in Turings Sinne als intelligent einstufen könnte. Zum Beispiel im Bereich des Schachspiels, bei dem der weltbekannte Schachspieler Garry Kasparov 1997 gegen den IBM-Computer *Deep Blue* verlor und während des Spiels bezweifelte, dass er wirklich gegen einen Computer spielte (Vgl. Buttazzo 2001, 25-26). Ebenso wurde 1997 von Musikprofessor Steve Larson von der Universität von Oregon eine Version des *TT* für Musik entworfen, bei dem die entsprechenden Computer erstaunlich gut abschnitten:

“The audience classified many computer pieces as authentic compositions and vice versa. A loose interpretation of these results could indicate that, with regard to musical composition, the computer passed the Turing test.” (Buttazzo 2001, 26)

Doch der allzu offensichtliche Einwand bleibt, dass es schwer fällt, Computern ein tatsächliches Verständnis von der von ihnen beherrschten Materie zuzusprechen. Denn letztlich geht jede dieser Leistungen auf klare mathematische Programmierung zurück. Akzeptiert man diesen Einwand und setzt man Verständnis von etwas als Bedingung für Gedanken darüber voraus, so scheinen die genannten Beispiele den *TT* eher als inadäquat zu erweisen, als zu zeigen, dass es bereits Computer mit bereichsspezifischen Denkfähigkeiten gibt.

Doch ein möglicher Gegeneinwand dazu ist, dass es bei fortschreitendem Erfolg neurowissenschaftlicher Untersuchungen des Gehirns und Interpretationen des menschlichen Verstandes in diesem Paradigma es sein kann, dass wir auch menschliches Denken irgendwann mit Hilfe klarer mathematischer Algorithmen erfassen und beschreiben können. Man könnte der Auffassung sein, dass die Mathematik die echten metaphysischen Strukturen und Gesetze der Welt erfasst und wiedergibt. Nur weil wir im einen Falle die Erzeuger der Systeme sind, die diesen Gesetzen gehorchen und im anderen Falle die Systeme sind, die selbigen gehorchen, wäre es bei weitem nicht gerechtfertigt, Maschinen Denkfähigkeiten abzusprechen und sie uns selbst zuzusprechen.

Doch diese Überlegungen hängen noch stark von theoretischer Spekulation ab und sind somit noch nicht im Entferntesten entscheidbar. Angesichts der genannten Schwierigkeiten, die sich bereits für die Verifikation eines Denkvermögens und somit vermutlich erst recht für die eines Bewusstseins ergeben, sollten wir vermutlich eher von *Plausibilisierungsstrategien* für das Vorhandensein eines Bewusstseins sprechen, als von einer Verifikation. Es lassen sich aber vielleicht ethische Gründe nennen, warum man selbst bei solch losen, angreifbaren Testverfahren wie dem *TT* eher davon ausgehen sollte, dass etwas Bewusstseinsartiges vorhanden ist, als vom Gegenteil: Würden wir zum Beispiel – trotz einer geringen Indikation für Bewusstsein – eine tatsächlich bewusste Maschine wie ein bewusstseinsloses Objekt behandeln und ihr vermeidbaren Schaden zufügen, *den sie als solchen erlebt*, würden wir uns in einem gewissen Sinne eines Unrechtes gegen diese Maschine schuldig machen. Behandelten wir hingegen eine Maschine ohne Bewusstsein, die in Frage kommt ein solches zu haben, als ob sie tatsächlich ein Bewusstsein habe, also mit einer gewissen Vorsicht, so vermieden wir, ihr unrecht zu tun.

#### **2.4. Zusammenfassung und Kontemplation der angestellten Überlegungen**

Wir hatten festgestellt, dass wir, um eine philosophisch stabile Konzeption dessen, was künstliches Bewusstsein ist, oder um zu der Annahme gelangen zu können, eine Maschine

habe ein Bewusstsein, erst einmal eine stabile Konzeption davon brauchen, was Bewusstsein *überhaupt* ist, und welche Kriterien erfüllt sein müssen, damit einem System oder einem Organismus ein Bewusstsein zugesprochen werden kann. Dabei hatten wir mit dem Konnektionismus eine Position kennengelernt, die das Bewusstsein potentiell als eine Eigenschaft, die auf systematischer Komplexität basiert, erklären könnte. Eine mögliche Gegenposition dazu habe ich noch nicht diskutiert: Die Auffassung, dass ein Bewusstsein an das spezifische *Material* gebunden ist, das ihm zu Grunde liegt. Man könnte ohne weiteres der Auffassung sein, dass für das Vorhandensein eines Bewusstseins das Vorhandensein eines zentralen Nervensystems die Voraussetzung ist. Diese Position ist nicht gerade abwegig, sondern sogar eher naheliegend, da die Objekte, denen wir intuitiv im Alltag Bewusstsein zuschreiben alle ein solches besitzen.

Die beiden Positionen sind allerdings nicht inkompatibel. Ebenso kann man der Auffassung sein, dass ein Nervensystem zwar eine notwendige, aber noch keine hinreichende Bedingung für Bewusstsein ist. Es ist denkbar, dass nur Nervensysteme *ab einem gewissen Grad der Komplexität* ein Bewusstsein besitzen, was eine Kombination der beiden Positionen darstellte. Dies würde bedeuten, nur ein komplexes Netzwerk, bestehend aus Neuronen und Synapsen brächte tatsächlich ein Bewusstsein hervor. Jede Position, die das Vorhandensein eines bestimmten biologischen Materials voraussetzt, bedeutet aber auch gleichzeitig den Ausschluss der Möglichkeit eines künstlichen Bewusstseins<sup>7</sup>! Setzen wir voraus, dass z.B. Neuronen in einem System vorhanden sein müssen, damit es ein Bewusstsein hat und nicht bloß etwas, das die funktionale Rolle von Neuronen einnimmt (wie z.B. Elektronen), dann ist es ausgeschlossen, dass es jemals ein künstliches Bewusstsein geben kann.

Interessanter Weise ist die Vorstellung, dass nur bestimmte Arten von Materie ein Bewusstsein haben können auch mit dem Substanzdualismus kompatibel. Es ist ohne Weiteres denkbar, dass nur Gehirne mit einer immateriellen Seele verknüpft sein können. Wenn wir nun aber an die Möglichkeit eines künstlichen Bewusstseins glauben wollen, so müssen wir diese Art von Theorie verwerfen, bzw. könnten sie nur mit großen Schwierigkeiten aufrecht erhalten<sup>8</sup>.

Gehen wir also einmal davon aus, Bewusstsein sei tatsächlich eine Eigenschaft, die Systemen, die komplex genug sind, zukommt, egal aus welchem Material diese Systeme

---

<sup>7</sup> Vorausgesetzt, wir meinen ‚künstliches Bewusstsein‘ hier terminologisch und verstehen geklonte Organismen oder ähnliches nicht als Beispiele künstlichen Bewusstseins, auch wenn man das Klonen als etwas Künstliches klassifizieren könnte.

<sup>8</sup> Denkbar wäre dies nur, wenn man eine wirklich verworrene Position einnimmt; z.B. die, dass man in Wirklichkeit selbst eine raffinierte Maschine sei, die ein Bewusstsein besitzt und von bewusstslosen Bio-Organismen geschaffen wurde. In dem Falle wären dann jedoch elektronische Systeme die einzigen, denen man ein Bewusstsein zuschreibt.

bestehen. Diese Annahme belässt uns immer noch mit der Schwierigkeit, ein sinnvolles Test-Kriterium dafür zu finden, ob ein künstliches System ein Bewusstsein hat, oder nicht. Ich hatte im letzten Abschnitt Einwände genannt, warum man den *TT* nicht unbedingt als geeigneten Test für Bewusstseinsfähigkeiten einstufen kann. Doch einen alternativen Test aufzustellen, der ein sicheres Prüfungsverfahren für das Bewusstsein darstellt, ist eine Schwierigkeit, die ich nicht zu lösen vermag.

Eine andere Möglichkeit ist es, einfach den abduktiven Schluss auf ein Bewusstsein zuzulassen, sofern ein bestimmtes Verhalten gezeigt wird und gewisse Grundkriterien erfüllt sind. Eine Kategorie von Kriterien dafür hatten wir bereits kennengelernt, nämlich den strukturell analogen Aufbau eines Systems zu Gehirnen (wie bei konnektionistischen Netzwerken). Einen anderen interessanten Vorschlag, der aus einer klassischeren Richtung der philosophischen Tradition kommt, habe ich noch nicht diskutiert: Susan A.J. Stuart und Chris Dobbyn berufen sich in einem Artikel von 2002 auf kantische Transzendentalien um Kriterien für die Gegebenheit eines Bewusstseins zu finden.

Ihr Vorschlag lautet, die deskriptive Metaphysik Kants in einer präskriptiven Weise umzuinterpretieren und somit eine Vorschrift für die Konstruktion künstlicher Systeme mit Bewusstseinsfähigkeiten zu erhalten:

„Kant is setting out a descriptive metaphysics and is not concerned with its physical instantiation. We have overcome this problem by interpreting Kant’s metaphysics prescriptively so that we can ask in what way his system might be prescriptive of the psychology of an artificial system.” (Stuart und Dobbyn 2002, 409)

Das Entscheidende an diesem Ansatz ist, dass die a priori Formen von raum-zeitlicher Interpretation der Welt, sowie der Ordnung in Kategorien und Konzepte nun zum einen als Vorschriften dafür gelten, wie ein bewusst erlebendes künstliches System grundlegend zu konzipieren ist, zum anderen aber auch philosophische Kriterien dafür gegeben sind, wann wir einem solchen System unter Umständen ein Bewusstsein zusprechen könnten.

Damit haben wir hier also wieder notwendige Bedingungen für Bewusstsein, darüber hinaus aber vielleicht auch insgesamt eine Hinreichende. Denn wenn wir uns vorstellen, dass uns ein künstliches System über sein Konzept von Liebe, Tod oder Gerechtigkeit berichtet, die Menschen, mit denen es konfrontiert wird in Kategorien wie z.B. ‚nett‘ und ‚unfreundlich‘ unterteilt und sich dabei in Raum und Zeit orientieren kann, so hätten wir doch bereits eine starke Verhaltens-Indikation für Bewusstsein gegeben und damit auch dieses (von mir weiter

oben angesprochene) Kriterium erfüllt! Denn was sonst soll gemeint sein, wenn wir uns über Verhalten Gedanken machen, dass als Indikator für ein Bewusstsein dienen soll?

Eine Frage, die darüber hinaus noch im Raume steht und auch von Stuart und Dobbyn aufgeworfen wird, ist die, ob das Bewusstsein einer Maschine – sofern diese denn eines hätte – sich überhaupt mit dem eines Menschen vergleichen ließe. Auch Buttazzo setzt sich mit dieser Frage auseinander und seiner Auffassung nach könnte die unterschiedliche Verarbeitungsgeschwindigkeit einen Unterschied in der Wahrnehmung bedeuten:

“Biological neurons respond within a few milliseconds, but electronic circuits can respond within a few nanoseconds. This observation leads to an interesting question: If consciousness emerges in an artificial machine, what will time perception be like to a simulated brain that thinks millions of times faster than a human brain?” (Buttazzo 2001, 29)

Wie also das Bewusstsein einer Maschine aus einer rein phänomenologischen Warte aussähe, ist nur schwer zu beurteilen. Diverse Faktoren (wie eben die Verarbeitungsgeschwindigkeit) könnten einen grundlegenden, qualitativen Unterschied zum Bewusstsein eines Menschen ausmachen. Aber auch dieses Problem steht in enger Verbindung zu generellen Bewusstseins-Problemen, die wir bereits zu Beginn der Arbeit thematisiert hatten. Denn kein Mensch hat einen genauen Zugriff auf die Phänomenologie, die Innenwelt, die Erlebnisperspektive des anderen. Und aus der Medizin kennen wir etliche Beispiele, in denen deutliche Unterschiede im Gehirn zweier Menschen (durch Krankheiten, Unfälle oder Ähnliches hervorgerufen) einen Unterschied in deren Erleben zu bedingen schienen. Die Frage, die sich im Hinblick auf künstliches Bewusstsein stellt, sollte also sein, *wie* verschieden es eigentlich vom typisch menschlichen Bewusstsein wäre.

## Fazit

Die Resultate, die wir aus unserer Untersuchung ziehen können, lassen sich relativ schnell zusammenfassen. Es ist mit Sicherheit logisch möglich, sowie denk- und vorstellbar, dass es irgendwann einmal Maschinen mit einem Bewusstsein geben wird, auch wenn dies bis jetzt noch nicht der Fall zu sein scheint. Doch um zu erkennen, ob eine Maschine ein Bewusstsein besitzt, müssen wir zunächst Kriterien finden, die wir als notwendige Bedingungen für Bewusstsein annehmen, um dann auf Basis derer ein Prüfungsverfahren – egal ob empirisch oder theoretisch – entwickeln zu können, mit dem wir bestimmen können, ob eine Maschine Bewusstsein hat oder nicht.

Wir hatten gesehen, dass wir zur Findung solcher Kriterien sowohl auf philosophische Überlegungen angewiesen sind (Bestimmung einer Definition von Bewusstsein, Bestimmung von Kriterien auf Basis von philosophischer Analyse des menschlichen Verstandes), als auch auf Kognitionsforschung und Computerwissenschaften (Untersuchung des Gehirns, psychologische Forschung, Konstruktion von konnektionistischen Netzwerken und empirische Untersuchung ihrer Leistungsfähigkeit). Diese Bereiche müssen dabei Hand in Hand gehen; die Philosophie liefert fundamentale Arbeitsgrundlagen, die Forschung untersucht auf Basis dieser die empirisch erfassbare Welt oder bedient sich ihrer z.B. als Konstruktionsvorschriften.

Wir hatten darüber hinaus gesehen, dass es noch nicht unbedingt ein befriedigendes Testverfahren für Bewusstsein gibt und der *TT* beispielsweise vielleicht nicht als hinreichend bewertet werden kann. An dieser Stelle gibt es definitiv Handlungsbedarf! Zudem hat sich in vielen Punkten gezeigt, dass die Problematik des künstlichen Bewusstseins mit den allgemeinen Problemen des Bewusstseins eng verknüpft ist. In beiden Fällen kann letztlich nur philosophische Überlegung Abhilfe schaffen, da hier erst ein theoretisches Fundament existieren muss, damit empirische Untersuchungen angestellt werden können. Es scheint also unerlässlich, dass die Philosophie des Geistes einen Beitrag zur künstlichen Intelligenz- und Bewusstseinsforschung leistet und dass wir auch als Philosophen aktuelle Themen wie das hier untersuchte erforschen und diskutieren.

Es bleibt also zu hoffen, dass in Zukunft auf Basis von Kooperation zwischen Philosophen und Einzelwissenschaftlern eine genauere Klärung der Möglichkeit künstlichen Bewusstseins geschehen wird und somit vielleicht irgendwann sogar der Traum einer bewusst erlebenden

Maschine Realität wird. Vermutlich könnten wir dadurch vieles über uns selbst und die tiefsten Geheimnisse dieser Welt erfahren.

## Verwendete Literatur

- Beckermann, Ansgar. 2000. „Ein Argument für den Physikalismus“, in: Keil und Schnädelbach (Hg.), *Naturalismus*. Frankfurt a.M.: Suhrkamp.
- Buttazzo, Giorgio. 2001. „Artificial Consciousness: Utopia or Real Possibility?“, in: *Computer* 34, 7.
- Descartes, René. 2004 [1641]. „Meditationen“, in: Jens Timmermann, *Sammlung Philosophie*. Übersetzt von Andreas Schmidt. Göttingen: Vandenhoeck & Ruprecht.
- Kirk, Robert 1996 [1995]. „Wie ist Bewusstsein möglich?“, in: Thomas Metzinger (Hrsg.), *Bewusstsein: Beiträge aus der Gegenwartsphilosophie*. Paderborn: mentis.
- Löbel, Benjamin. 2008. *Einführung in die Philosophie des Geistes*. Norderstedt: Grin.
- Nagel, Thomas. 1981 [1974]. „Wie es ist eine Fledermaus zu sein“, in: Peter Bieri (Hrsg.), *Analytische Philosophie des Geistes*. Königstein/Ts.: Hain Verlag.
- Nolan, Daniel. 2005. *David Lewis*. Chesam: Acumen Publishing.
- Schlagel, Richard H. 2004. „Why not Artificial Consciousness or Thought?“, in: *Mind and Machines* 9, 1.
- Stillings, Neil A. et al. 1995. *Cognitive Science: An Introduction*. Cambridge: MIT Press.
- Stuart, Susan A.J. und Dobbyn, Chris. 2002. „A Kantian Prescription for Artificial Conscious Experience“, in: *Leonardo* 35, 4.
- Turing, Alan. 1950. „Computing Machinery and Intelligence“, in: *Mind* 59, 236.

## **Eidesstattliche Versicherung**

Hiermit erkläre ich, Florian Boge, dass ich die Hausarbeit mit dem Titel

### **Über die Möglichkeit künstlichen Bewusstseins - Grundlegende philosophische Überlegungen und Probleme**

selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich unter Angabe der Quelle als Entlehnung kenntlich gemacht.

Düsseldorf, den

---

(Florian Boge)