Dirk Evers

DIE GÖDELSCHEN THEOREME UND DIE FRAGE NACH DER KÜNSTLICHEN INTELLIGENZ IN THEOLOGISCHER SICHT[1]

1. Das Hilbertsche Programm im historischen Kontext

Die historischen Wurzeln, die Gödels Unvollständigkeitssätze mit der Frage verbinden, ob der menschliche Geist als Maschine zu verstehen sei, reichen zurück bis in das 19. Jahrhundert. In einem seinerzeit und in der Folge viel beachteten öffentlichen Vortrag äußerte sich der Berliner Physiologe Emil Du Bois-Reymond 1872 „Über die Grenzen des Naturerkennens". Du Bois-Reymond hatte selber den Vitalismus vehement und erfolgreich bekämpft und in Berlin die rein materialistisch orientierte Schule der mechanischen Physiologie gegründet. Ihm war es 1848 gelungen, die elektrischen Ströme in Nervenzellen nachzuweisen, die so genannten „Aktionspotentiale". Naturerkennen, so stellte Du Bois-Reymond in seinem Vortrag als Fazit der Erfolgsgeschichte der Naturwissenschaften fest,

„ist Zurückführen der Veränderungen in der Körperwelt auf Bewegungen von Atomen, die durch deren von der Zeit unabhängigen Centralkräfte bewirkt werden", sie findet ihr Ziel im „Auflösen der Naturvorgänge in Mechanik der Atome."[2]

Die Gesetze der Mechanik sind mathematisch darstellbar und tragen denselben apodiktischen Charakter wie die Mathematik. Deshalb stellt der Zustand der Welt während eines Zeitabschnitts sich als unmittelbare Wirkung des vorigen und als unmittelbare Ursache des nachfolgenden Zeitabschnitts dar, ja, es lässt sich denken, dass das Weltganze

„durch Eine mathematische Formel vorgestellt würde, durch Ein [sic] unermessliches System simultaner Differentialgleichungen, aus dem sich Ort, Bewegungsrichtung und Geschwindigkeit jedes Atoms im Weltall zu jeder Zeit ergäbe".[3]

[1] Eine etwas veränderte Fassung dieses Vortrags ist unter dem Titel „Der Mensch als Turing-Maschine? Die Frage nach der künstlichen Intelligenz in philosophischer und theologischer Sicht" inzwischen erschienen in: Neue Zeitschrift für Systematische Theologie und Religionsphilosophie 47 (2005), 101–118.

[2] E. Du Bois-Reymond, Über die Grenzen des Naturerkennens, [6]1884, 12.

[3] AaO., 13.

Du Bois-Reymond prägte in diesem Vortrag auch die inzwischen geläufige Bezeichnung ‚Laplacescher Geist' für die auf Laplace zurückgehende Fiktion einer übermenschlichen Vernunft, die

„für einen gegebenen Augenblick alle in der Natur wirkenden Kräfte sowie die gegenseitige Lage der sie zusammensetzenden Elemente kennte, und überdies umfassend genug wäre, um diese gegebenen Größen der Analysis zu unterwerfen".

Eine solche Intelligenz „würde in derselben Formel die Bewegungen der größten Weltkörper wie die des leichtesten Atoms umschließen; nichts würde ihr ungewiss sein und Zukunft wie Vergangenheit würden ihr offen vor Augen liegen."[4]

Der menschliche Geist wiederum ist dem Laplaceschen prinzipiell ebenbürtig und nur gradweise von ihm verschieden:

„Wir gleichen diesem Geist, denn wir begreifen ihn."[5]

Zwar sind die Grenzen des menschlichen Erkennens offensichtlich, übersteigt doch die Komplexität der Phänomene allein durch ihre große Zahl schon bei weitem das Fassungsvermögen menschlicher Vernunft. Doch ist dies nicht in der Natur der Dinge begründet, sondern nur in der begrenzten Kapazität menschlichen Zugangs zu den Phänomenen, der weder alle Bestimmungsstücke mit der nötigen Präzision feststellen noch die dazugehörigen Differentialgleichungen in ihrer Komplexität bewältigen kann. Prinzipiell aber kann die Grenze unseres Nicht-Wissens, unseres ‚ignoramus' immer weiter hinausgetrieben werden und nichts in der materiellen Welt sich seinem Zugriff grundsätzlich verweigern.

Zwei grundsätzliche Grenzen jedoch markiert Du Bois-Reymond, die dem berechnenden Zugriff der Naturwissenschaften auf immer entzogen sein werden. Es ist die Frage nach dem *Wesen* von Materie und Kraft sowie die Frage nach dem Zusammenhang von *Bewusstsein* und Materie. In beiden Fällen ist nicht etwa bloß ein durch mögliche zukünftige Erkenntnis zu überwindendes ‚ignoramus' zu konstatieren, sondern ein jegliche Erkenntnis verweigerdes ‚ignorabimus' hinzuzufügen: Wir wissen hier nicht, und wir werden hier nie wissen können, denn es sind keine Umstände denkbar, unter denen ein Wissen in der Form, wie es ein Laplacesche Geist produzieren könnte, in diesen Fragen hervorgebracht werden könnte. Der Übergang vom Kausalzusammenhang, vom bloß Relationalen des Materiellen, wie es die Naturwissenschaft erfasst, zu seinem Wesen ist ebenso unableitbar wie der Übergang von der Anordnung und Bewegung materieller Teilchen in das Reich des Bewusstseins. Bei genauer Betrachtung

[4] P. S. de Laplace, Philosophischer Versuch über die Wahrscheinlichkeiten (Ostwald's Klassiker der exakten Wissenschaften 233), 1932, 4.

[5] E. Du Bois-Reymond, aaO., 18.

zeigt sich jeweils die „transcendente Natur des Hindernisses",[6] die nicht eine vorläufige Grenze auf dem Gebiet der Wissens, sondern eine wirkliche Schranke für die Möglichkeit von Naturerkenntnis überhaupt darstellt. Das sagte, wohl gemerkt, ein materialistisch orientierter Hirnforscher.

Es ist dieses berühmt gewordene ignorabimus, gegen das sich der Mathematiker David Hilbert mit seiner Beweistheorie wandte. In der Mathematik jedenfalls als der Sphäre des reinen Denkens sollte es gelingen, alle sinnvollen Fragen auch einer Beantwortung zuzuführen: „In der Mathematik gibt es kein Ignorabimus"[7], und von da aus schien ihm das Ergebnis auch auf die Naturwissenschaften übertragbar. Hilbert gab 1930, dem Jahr, in dem Gödel seine Unvollständigkeitssätze zuerst vorstellte, vor eben der Gesellschaft Deutscher Naturforscher und Ärzte, vor der auch Du Bois-Reymond gesprochen hatte, in seiner Eröffnungsansprache nun die gegenteilige Losung aus:

„Für den Mathematiker gibt es kein *Ignorabimus*, und meiner Meinung nach auch für die Naturwissenschaften nicht ... Der wahre Grund, warum es nicht gelang, ein unlösbares Problem zu finden, besteht meiner Meinung nach darin, dass es unlösbare Probleme überhaupt nicht gibt. Statt des törichten Ignorabimus heiße im Gegenteil unsere Losung: Wir müssen wissen, wir werden wissen."[8]

Auf diese Weise sollte eine als widerspruchsfrei erwiesene Mathematik der sichere Grund werden, auf dem alle anderen Wissenschaften aufbauen. Die Widerspruchsfreiheit und Vollständigkeit der Mathematik sei deshalb

„wie ein Heiligtum zu hüten, damit einst *alles* menschliche Wissen überhaupt der gleichen Präzision und Klarheit teilhaftig wird"[9].

Hilberts Projekt stellt also nichts anderes als den Gegenentwurf zu Du Bois-Reymonds Ignorabimus dar, insofern er mittels des Aufweisens von Widerspruchsfreiheit und Vollständigkeit der mathematischen Grundlagen die Widerspruchsfreiheit und Vollständigkeit von Wissen überhaupt garantieren möchte.

[6] AaO., 23. Vgl. ders., Die sieben Welträtsel, aaO., 77: „*Transcendent* nenne ich darunter die [Schwierigkeiten], welche mir unüberwindlich erscheinen, auch wenn ich mir die in der aufsteigenden Entwicklung ihnen voraufgehenden gelöst denke." Auf diese Welträtsel nimmt Ernst Haeckel Bezug, wenn er beansprucht, sie mit Hilfe seiner popularisierten ‚monistischen Philosophie' gelöst zu haben, vgl. E. Haeckel, Die Welträthsel. Gemeinverständliche Studien über Monistische Philosophie, 1899.

[7] D. Hilbert, Probleme der Grundlegung der Mathematik, Math. Annalen 102 (1925), 1–9, 9.

[8] D. Hilbert, Naturerkennen und Logik, Ges. Abh. III, 387. Der letzte Satz steht auch auf Hilberts Grabstein.

[9] D. Hilbert, Grundlegung der elementaren Zahlenlehre, Mathematische Annalen 104 (1931), 485–494, 494. In der Widerspruchsfreiheit gründet auch Hilberts mathematischer Existenzbegriff: was widerspruchsfrei ist, existiert. Vgl. seine Auseinandersetzung darüber mit Frege in: Gottlob Freges Briefwechsel (PhB 321), G. Gabriel et al. (Hg.), 1980, 12.

2. Künstliche Intelligenz und Turing-Maschinen

Die Frage nach der künstlichen Intelligenz lässt sich formulieren als die Frage, ob mit Hilfe formaler, nach mathematischen Regeln operierender Systeme Funktionen menschlicher Intelligenz nachgebildet und dann auf künstlich hergestellten Maschinen realisiert werden können, so dass diesen Maschinen selbst Intelligenz zugeschrieben werden muss. Insofern damit behauptet wird, dass auf diese Weise Eigenschaften kontrolliert nachkonstruiert werden können, die traditionellerweise dem Menschen zugeschrieben werden, meldet sich auch der Anspruch, dass nun auf objektive und an der unumstößlichen Gewissheit mathematischer Erkenntnis teilhabende Weise nachvollzogen wird, was denn Intelligenz überhaupt sei. Theorie und Pragmatik der technischen Systeme, so dann der weitergehende Schluss, umfassen auch Theorie und Praxis menschlicher Intelligenz, ja bringen sie gerade wegen ihrer mathematisierten Form aus dem bloßen Herumtappen in den sicheren Gang einer Wissenschaft.

So wie die klassische Mechanik die Bewegungsgesetze zurückführte auf ideale störungsfreie Trägheitsbewegungen, so führt die Theorie der künstlichen Intelligenz das Denken zurück auf die ideale Maschine. Als das Modell einer idealen Denk-Maschine gilt die universelle Turing-Maschine[10], eine diskret, d. h schrittweise gesteuerte Maschine mit unendlicher Speicherkapazität, die elektromechanisch realisiert oder auch mit Computern simuliert werden kann – natürlich ohne die ihrer Theorie zugrunde gelegten idealen Eigenschaften wie unendlicher Speicherkapazität und unbegrenzt zur Verfügung stehender Zeit. Es lässt sich zeigen, dass alle algorithmischen, rekursiven Verfahren auf der universellen Turing-Maschine simuliert werden können.

Inzwischen sind viele Maschinen realisiert, die der Theorie der Turing-Maschinen gehorchen und von Algorithmen gesteuert werden, die in Verbindung mit Rezeptoren, die optische, akustische, olfaktorische und andere Reize in verarbeitbare Daten

[10] Alan M. Turing (1912–1954), britischer Mathematiker, Logiker und Computerwissenschaftler, entwickelte 1936–37 die Theorie der nach ihm benannten Turing-Maschinen (A. M. Turing, On Computable Numbers, With An Application to the Entscheidungsproblem (1937), in: Ders., Intelligence Service. Schriften, hg. v. B. Dotzler/F. Kittler, 1987, 18–60). Eine Turing-Maschine verfügt über eine endliche Anzahl diskreter innerer Zustände und einen äußeren Speicher in Form eines Magnetbandes. Dieses prinzipiell unendlich lange Band ist in Felder unterteilt, die an einem Lese-/Schreibkopf vorbeilaufen. Dieser ist in der Lage, ein Zeichen, das in das jeweilige Feld eingetragen ist, zu lesen, ein Zeichen in ein leeres Feld zu schreiben oder ein vorhandenes Zeichen zu löschen. Der Lese-/Schreibkopf ebenso wie die schrittweise Vorwärts- bzw. Rückwärtsbewegung des Bandes werden nach Maßgabe der inneren Zustände der Maschine gesteuert, die außerdem festlegen, wie von einem inneren Zustand in einen anderen übergegangen werden soll, je nachdem, welche Zeichen vom Band gelesen werden. Turing war in der Lage zu zeigen, dass Turing-Maschinen im Prinzip alles tun können, was jede beliebige Rechenmaschine können muss, welcher Bauart sie auch sei. Er zeigte auch, dass es kein endliches Entscheidungsverfahren gibt, das für alle möglichen Turingmaschinen und alle möglichen Eingaben entscheidet, ob die Turingmaschine mit dieser Eingabe nach endlich vielen Schritten anhält oder endlos weiterläuft, das so genannte Halteproblem.

übersetzen und damit zu überaus differenziertem Problemlösungsverhalten in der Lage sind. Dazu gehören komplexe Entscheidungsprozesse, wie sie in Schachcomputern realisiert sind, aber auch in Geräten zur chemischen Analyse und medizinischen Diagnose, und wie sie sich vom frühen General Problem Solver (GPS)[11] bis zu heutigen regelgeleiteten Expertensystemen entwickelt haben, aber auch immer ausgefeiltere Kommunikationssysteme und selbständig agierende und auf ihre Umwelt reagierende Roboter wären hier zu nennen.

3. Künstliche und natürliche Intelligenz

Auch aus theologischer Sicht dürfte es einigermaßen unproblematisch sein, solchen Maschinen Intelligenz zuzuschreiben und diese als „künstliche Intelligenz" zu apostrophieren.[12] Denn diese Maschinen erbringen Leistungen, die wir als „intelligent" bezeichnen würden, wenn ein Mensch sie erbringen würde. Insofern sie aber kein Mensch, sondern eine Maschine erbringt, ist es sinnvoll, dieser Maschine eine nicht natürlich entstandene, sondern eben eine *künstliche*, also durch von Menschenhand konstruierte Maschinen erbrachte Intelligenz zuzuschreiben. *Intelligenz* ist wohl auch nicht missverstanden, wenn man sie als einen regelgeleiteten Prozess versteht, der in Verfolgung vorgegebener Ziele auf externe, in symbolisierter Form prozessierte Daten mit angemessenen Reaktionen antwortet und womöglich die auf die erteilte Antwort erfolgenden Reaktionen noch einmal in Bezug auf die verfolgten Ziele analysiert, um die ihn steuernde Regel zu optimieren. In diesem Sinne behauptet der Ausdruck „Künstliche Intelligenz" nicht zu viel.

Doch insofern nun konstruierte *Maschinen* diese Leistung erbringen, wird diese Form der Intelligenz eben *künstlich* genannt. Die prinzipielle und weiterführende Frage, die sich nun stellt, ist die, ob die Prinzipien, die diesen Maschinen zugrunde liegen, nur die *notwendigen* oder aber auch die *hinreichenden* Bedingungen von Intelligenz überhaupt darstellen. Letzteres legt sich im Rahmen einer naturwissenschaftlich orientierten und als Naturalismus bezeichneten Grundanschauung nahe. Unter *Naturalismus* verstehe ich in diesem Zusammenhang die These, dass alle realen Systeme, seien es Sterne, Planeten, Steine, Pflanzen, Tiere oder Menschen, „aus mehr oder weniger komplexen Anordnungen von materiellen Konstituentien" bestehen, „die durch eine kleine Zahl von

11 Vgl. dazu H.A. Simon/A. Newell, Informationsverarbeitung in Computer und Mensch, in: Künstliche Intelligenz. Philosophische Probleme, W. Ch. Zimmerli/St. Wolf (Hg.), 1994, 112–145, bes. 135 ff.

12 So auch E. Herms, Künstliche Intelligenz, in: ders., Gesellschaft gestalten. Beiträge zur evangelischen Sozialethik, 1991, 284–295, 288.

dynamischen Wechselwirkungen zusammengehalten werden"[13], und zwar genau denjenigen Wechselwirkungen, die die mathematisierten Naturwissenschaften zu beschreiben in der Lage sind. Auch das Verhalten großer und komplexer Systeme in seiner ganzen Vielfalt ließe sich dann zumindest prinzipiell zurückführen auf die Eigenschaften der elementaren Bausteine und ihre Wechselwirkungen. Der Naturalismus kann in seinem Anspruch als die Vollendung des durch den Laplaceschen Geist repräsentierten Programms von Du Bois-Reymond verstanden werden, jetzt aber unter Einschluss der psychischen Prozesse.

Nun ist für alle kognitiven Prozesse menschlicher Intelligenz dasjenige System bekannt, das von Rezeptoren symbolisierte Daten prozessiert und regelgeleitet in angemessenes Verhalten umsetzt, nämlich unser Gehirn. Einzelne mentale Operationen lassen sich heute bestimmten Gehirnarealen zuordnen, emotive und kognitive Leistungen werden durch Schädigungen von Gehirnformationen spezifisch beeinträchtigt, und auch die einheitlichen Bausteine des Gehirns, die Nervenzellen sind bekannt und in einigen Aspekten ihrer Funktionalität analysiert. Diese Neuronen scheinen dabei im Prinzip wie elektro-chemische Schalter zu funktionieren und sollten durch berechenbare Funktionen darstellbar sein. Darüber hinaus ist der ganze kognitive Apparat im Rahmen eines naturalistischen Weltbilds als Produkt evolutionärer Entwicklung und Optimierung anzusehen, so dass die ihm extern vorgegebenen Ziele in umwelt-adäquatem Verhalten und überlebensfördernder Effizienz bestehen, sowohl im Sinne der individuellen Fitness des Individuums wie auch der Gesamtfitness der Gattung.

Damit legt sich der Schluss nahe, dass auch der menschliche kognitive Apparat nichts anderes ist als ein General Problem Solver und damit ein Spezialfall einer Turing-Maschine. Genau auf dieser Linie argumentierte Turing selber, um die Intelligenz von Maschinen als Paradigma von Intelligenz an sich zu behaupten. Zwar gesteht Turing zu, dass das Nervensystem mit Sicherheit keine diskrete Maschine ist[14]. Dennoch hält er fest, dass Gehirne diskreten Systemen sehr nahe stehen,

„und es scheint allen Grund für die Annahme zu geben, sie könnten auch so beschaffen sein, daß sie gänzlich darunter fallen, ohne irgendeine Veränderung ihrer wesentlichen Eigenschaften."[15]

Deshalb sollten jedenfalls die Zustände, in denen das Gehirn sich dann befindet, wenn es Rechenoperationen durchführt, vollkommen durch entsprechende Zustände einer Turing-Maschine repräsentiert werden können:

[13] B. Kanitscheider, Vorwort, in: Hermeneutik und Naturalismus, ders./F.J. Wetz (Hg.), 1998, V.

[14] A. Turing, Computing Machinery and Intelligence (1950), in: ders., Intelligence Service. Schriften, hg. von B. Dotzler/F. Kittler, 1987, 148–182, 171.

[15] A. Turing, Intelligent Machinery (1969), in: aaO., 82–113, 87. Vgl. auch aaO., 96: „Die Schaltkreise, die in elektronischen Rechenmaschinen verwendet werden, scheinen die wesentlichen Eigenschaften von Nerven zu haben."

„Jedem Geisteszustand des Rechnenden entspricht ein ‚*m*-Zustand' der Maschine."[16]

Daran anschließend hat Alan Turing den bekannten und später nach ihm benannten Turing-Test entwickelt. Es handelt sich um ein Spiel, an dem ein Mensch, eine Maschine und ein Fragesteller teilnehmen. Alle drei finden sich in getrennten Räumen, und der Fragesteller befragt Mensch und Maschine gleichermaßen, um herauszufinden, in welchem Raum die Maschine sich befindet. Die Maschine soll so programmiert werden, dass sie den Fragesteller zu täuschen versucht. Wenn er nach einer ausreichenden Anzahl von Durchgängen in nicht wesentlich mehr als der Hälfte der Versuche den Menschen richtig identifizierte, hat die Maschine den Test bestanden, so dass ihr ein dem menschlichen analoges Denkvermögen zuzuschreiben ist.

Der einzige von Turing zugestandene Unterschied zwischen der maschinellen und der menschlichen Hardware ist der, dass sich im Falle einer stetig, also nicht-diskret arbeitenden Maschine kleinste Fehler und Änderungen in großen Effekten äußern können[17]. Das kann jedoch den Turing-Test nicht beeinflussen, denn es „wird der Fragesteller nicht in der Lage sein, aus diesem Unterschied irgendeinen Vorteil zu ziehen"[18], da solche Effekte leicht nachzumachen und von programmiertem Zufall nicht zu unterscheiden wären. Überhaupt würde es keine Schwierigkeit machen, die Fehleranfälligkeit und Irrtumfähigkeit menschlichen Denkens auf Maschinen zu simulieren, so dass auch die Perfektion, mit der eine künstliche Maschine arbeitet, sie im Turing-Test nicht notwendigerweise entlarven müsste. In der „Behauptung, ‚Maschinen können keine Fehler machen'" und das unterscheide sie von Menschen, sieht Turing jedenfalls keinen wirklich schlagenden Einwand:

16 A. Turing, On Computable Numbers, 43.

17 Im Gegensatz dazu tritt bei einer diskreten Maschine dieser Effekt, den die Chaostheorie heute als eine häufige Eigenschaft natürlicher Systeme ansieht, nicht auf: „Es will scheinen, als ob es bei gegebenem Anfangszustand der Maschine und gegebenen Eingabesignalen immer möglich sei, alle zukünftigen Zustände vorherzusagen. Das erinnert an die Laplacesche Ansicht, daß es möglich sein müßte, aus dem vollständigen Zustand des Universums zu einem bestimmten Zeitpunkt, beschrieben durch Lage und Geschwindigkeiten sämtlicher Partikel, alle zukünftigen Zustände vorherzusagen. Die von uns hier betrachtete Vorhersage ist jedoch praktikabler als die von Laplace erwogene. Das System des ‚Universums als ganzem' ist so beschaffen, daß minimale Fehler in den Anfangsbedingungen zu einem späteren Zeitpunkt einen überwältigenden Einfluß haben können. Die Verschiebung eines einzigen Elektrons um einen billionstel Zentimeter in einem Augenblick könnte ein Jahr später darüber entscheiden, ob ein Mensch von einer Lawine getötet wird oder ihr entkommt. Es ist eine wesentliche Eigenschaft der mechanischen System, die wir ‚diskrete Maschinen' genannt haben, daß dieses Phänomen nicht auftritt. Selbst wenn wir die tatsächlichen, physikalischen Maschinen anstelle der idealisierten Maschinen betrachten, ergibt sich aus einer verhältnismäßig genauen Kenntnis des jeweiligen Zustandes eine verhältnismäßig genaue Kenntnis aller späteren Schritte" (A. Turing, Computing Machinery and Intelligence (1950), 157 f.).

18 AaO., 172.

„Man ist versucht zu erwidern: ‚Sind sie deswegen weniger wert?'"[19]

4. Zur Differenz zwischen natürlicher und künstlicher Intelligenz

Mit seinem Testverfahren für die Zuschreibung von Intelligenz hat Turing im Grunde ein unschlagbares Argument geliefert: Sage mir klar und deutlich, worin der Unterschied zwischen Maschine und Mensch in ihrem Verhalten besteht, und ich baue eine Maschine, die auch diese Differenz noch simuliert. Denn sobald du etwas klar und deutlich angeben kannst, kann es algorithmisch nachvollzogen werden.

Ausgehend von dieser schlecht von der Hand zu weisenden Argumentation, dass, sobald wir eine Differenz im Verhalten von Mensch und Maschine nicht bloß intuitiv behaupten oder raten, sondern bestimmt angeben können, worin sie besteht, diese Differenz auch wieder simulierbar ist, ergeben sich zwei mögliche Argumentationsmuster. Zum *einen* kann man nun behaupten, die natürliche Intelligenz sei im Grunde auch nichts anderes als die künstliche. Naturalisten könnten berechtigt fühlen, die Behauptung, dass Computer ebenso denken können wie Menschen, so lange aufrecht zu erhalten, wie das *eigentümliche* Wesen von natürlicher Intelligenz nicht aufgedeckt ist. Doch so bald dieses klar und deutlich bestimmt wird, kann man sogleich eine neue Maschine jedenfalls als Möglichkeit entwerfen, die auch diese vorgebliche Wesenseigenschaft natürlicher Intelligenz noch simuliert.

Zum *anderen* können aber auch die Gegner des Reduktionismus immer darauf hinweisen, dass die Differenz dann eben im Nicht-Empirischen liegen müsse, in dem, was nicht klar und deutlich bestimmt werden kann, sondern intuitiv erfasst werden muss. Eine diskret arbeitende Turing-Maschine, so kann umgekehrt argumentiert werden, ist dann prinzipiell nicht in der Lage, durch Simulation das einzuholen, was nur intuitiv und nicht nach formalen Regeln erkannt werden kann und was nur natürlicher Intelligenz eigen ist, die deshalb mehr und anderes sein muss, als was funktional modelliert werden kann.

Ich möchte an dieser Stelle für den zweiten Argumentationsgang plädieren, wohl wissend, dass dieses Plädoyer auf menschliche Selbsterfahrung und menschliches Selbstbewusstsein verweisen muss, das zwar phänomenologisch aufgezeigt, nicht aber als empirische Größe gemessen werden kann. Es wird jedenfalls nicht gelingen, eine Differenz zwischen einer maschinell-funktionalen und der natürlichen Intelligenz aufzuweisen, die ihrerseits so präzise empirisch erfasst werden kann, dass sie auch wieder simulierbar wäre. Der nächste Abschnitt soll darüber hinaus dann noch zeigen, dass

[19] AaO., 167 f.

auch die Gödelschen Theoreme für das hier vorgetragene Plädoyer in Anschlag gebracht werden können.

Doch zunächst sei auf die zentrale phänomenologische Differenz zwischen bloß funktionaler und der aus unserem Selbsterleben bekannten natürlichen Intelligenz hingewiesen. Eine Turing-Maschine ist Sklavin ihres Programms, so offen und variabel dieses auch programmiert sein mag. Die Eigenart einer Turing-Maschine, so würden wir heute sagen, besteht in ihrer Software, die Hardware ist beliebig. Wir können deshalb ihr Programm, ihre Funktion von gänzlich verschiedenen Maschinen realisieren lassen. Darunter sind auch solche Maschinen, denen wir intuitiv gerade keine Fähigkeit des Denkens zuschreiben würden. So könnte jede Turing-Maschine als mechanisch-pneumatische Maschine realisiert werden. Doch dass Ventilen und Kolben künstliche Intelligenz im Sinne von Denken zukommt, wird man nicht so leicht behaupten wollen[20]. Umgekehrt kann man die Funktionen der Turing-Maschine aber auch gerade durch ein bewusstes und auf jeden Fall denkendes Wesen wie einen Menschen ausführen lassen, ohne dass dieser versteht, was er tut. Dies ist das bekannte Argument des chinesischen Zimmers von John Searle[21]. In diesem Gedankenexperiment sitzt ein Mensch, der des Deutschen, aber nicht des Chinesischen mächtig ist, in einem abgeschlossenen Zimmer und hat vor sich eine Aufstellung von Regeln, wie welche deutschen Worte durch welche chinesische Schriftzeichen, die ihm vorgefertigt zur Verfügung stehen, zu ersetzen sind. Wann immer man einen Zettel mit einem deutschen Text von außen durch einen Briefkastenschlitz ins Zimmer schiebt, wird man nach einiger Zeit von dem Menschen im Zimmer eine Zusammenstellung von chinesischen Schriftzeichen zurück erhalten, die eine chinesische Übersetzung des deutschen Textes darstellen. Vorausgesetzt, die rein mechanisch auszuführenden Regeln sind gut und der Mensch verfolgt sie gewissenhaft, wird auch die Übersetzung gut

[20] Die starke KI gründet auf der Voraussetzung der Unterscheidung von Hardware und Software, doch ist fraglich, ob diese Differenz immer konsequent durchdacht wird. Entscheidend ist ja, dass es unerheblich sein soll, ob neuronale Strukturen oder elektronische Schaltkreise in derselben Weise funktionieren und ununterscheidbare Effekte hervorrufen. Wenn wir jedoch sagen, das Entscheidende sei einzig und allein das Programm, dann wären Denken, Bewusstsein etc. zumindest im Prinzip auch auf *mechanischen* Maschinen simulierbar, denn jeder elektronische Computer könnte mit mechanischen Maschinen wiederum simuliert werden. Die Überzeugungskraft der starken KI beruht sicher mit auf der geheimnisvollen Kraft der Elektrizität, die Geisthaftigkeit suggeriert und die zugleich mit ungeheurer Geschwindigkeit nahezu verlustfrei arbeitet, damit aber auch für uns unbeobachtbar bleibt. Intuitiv erscheint es uns aber nicht beliebig, ob Neuronen, elektronische Schaltkreise oder pneumatische Ventile „Denkprozesse" durchführen. Wir trauen dem ersten Baustein mehr zu als dem zweiten und diesem wiederum mehr als dem dritten. Vielleicht sollte man wirklich deutlicher machen, dass die biochemischen und elektrischen Funktionen, mit denen Neuronen arbeiten, längst noch nicht wirklich verstanden sind.

[21] J. Searle, Geist, Gehirn, Computer, in: W. Zimmerli/St. Wolf (Hg.), Künstliche Intelligenz. Philosophische Probleme, 1994, 232–265.

sein und man könnte von außen den Eindruck bekommen, im Zimmer sitze jemand, der des Chinesischen mächtig ist. Doch der die Regeln Ausführende tut dies rein mechanisch, ohne das geringste Verständnis für das, was er da tut, auch wenn er von seinen geistigen Fähigkeiten her prinzipiell in der Lage wäre, Chinesisch zu beherrschen. Daraus ergibt sich der Schluss: Ein Programm, eine Funktion, ein regelgesteuertes Verhalten muss nicht selbst denken und verstehen können, um zu funktionieren.

Wenn also Maschinen, die aus Bauelementen bestehen, denen wir die Fähigkeit wirklichen Denkens und bewusster Intelligenz grundsätzlich absprechen, in ihrer Funktionalität jedem Computer gleichzusetzen sind, und wenn umgekehrt ein denkendes Wesen alles das ausführen kann, was ein Computer tut, ohne ein bewusstes Verständnis für die Sache zu entwickeln, dann ist deutlich, was künstliche Intelligenz von natürlicher Intelligenz unterscheidet: Künstliche Intelligenz befolgt Regeln, natürliche Intelligenz versteht Bedeutungen. Auch wenn das wahrnehmbare Verhalten zwischen beiden Systemen nicht grundsätzlich, nicht regelmäßig und deshalb in bestimmten Situationen gar nicht phänomenal zu unterscheiden ist, ist diese Differenz festzuhalten. Jeder rein formal operierenden Maschine ist dann in einem eingeschränkten Sinne nur „künstliche", das heißt nicht Bedeutungen erzeugende und verstehende Intelligenz zuzuschreiben. Oder umgekehrt, natürliche intelligente Systeme können nicht bloß formal operierende Maschinen sein.

5. Semantische Offenheit und die Gödelschen Theoreme

Welche Bedeutung haben dann die Gödelschen Theoreme? Sie machen jedenfalls dieses deutlich, dass der Begriff der mathematischen Wahrheit nicht innerhalb desselben Systems mit konstruiert werden kann. Zwar hat Gödel gerade gezeigt, dass die Differenz zwischen Objekt- und Metasprache durch sein Verfahren der Gödelisierung aufgehoben werden kann. Der Begriff der Beweisbarkeit kann mit Mitteln des Systems selbst dargestellt werden. Doch gerade dann, wenn das der Fall ist, lässt sich nicht mehr sicherstellen, dass das System vollständig und widerspruchsfrei ist. Es ist demnach auch nicht die bloße Selbstbezüglichkeit, die durch selbstreferentielle Schließungen die Unvollständigkeit des Systems hervorruft, es ist die Tatsache, dass ein System seine eigenen Aussagen semantisch bewertet, es ist, was man seine *semantische Geschlossenheit* nennen kann.

Es scheint mir damit gezeigt zu sein, dass ein rein syntaktischer Begriff von Wahrheit in einer semantisch geschlossenen Sprache zu Widersprüchen führen kann. Wahrheit, so ist dann umgekehrt zu sagen, ist ein Beziehungsbegriff, der nur dann Sinn macht, wenn er außerhalb und relativ zu einem System formuliert wird. Und diese Relationalität von Wahrheit ist nicht abschließbar, sie kann sozusagen nicht durch eine

umfassende, abschließende formale Theorie gedeckelt werden. Gödel lässt also nicht den Formalismus als solchen scheitern, sondern stellt so etwas wie eine konsequente „Selbstkritik der Hilbertschen Grundlagenforschung"[22] dar, und nicht zuletzt wegen ihres strengen Beweischarakters wurde Gödels Erkenntnis im Lager der Formalisten so schnell akzeptiert[23]. Heinrich Scholz hat deshalb Gödels Resultat in den „Rang einer zweiten nachkantischen Vernunftkritik"[24] erhoben, die unüberschreitbare Grenzen aufzeigt, zugleich aber alles Denken, das innerhalb dieser Grenzen bleibt, in das ihm zukommende Recht setzt.

Gödel hat nach dieser Interpretation

„der Fortsetzung des formalistischen Programms den Weg geebnet, indem er die Konzeption einer Hierarchie von Systemen präsentierte, in denen unentscheidbare Aussagen auf einer höheren Ebene entscheidbar werden, die zwar wiederum nicht vollständig und widerspruchsfrei im Sinne Hilberts sein kann, die aber auf einer noch höheren Eben ergänzungsfähig ist, und so fort."[25]

Insofern ist jedenfalls hinter Hilberts Losung „Wir müssen wissen, wir werden wissen", ein Fragezeichen zu setzen, wenn eine vollständig bestimmte, abgeschlossene Totalität von Wahrheiten gemeint sein soll. Ganz in diesem Sinne hat dann auch Gödel selbst die Konsequenzen seiner Entdeckung formuliert. Der Fortschritt der Mathematik wie überhaupt aller menschlichen Erkenntnis ist nach seiner Auffassung in einer Sinnklärung zu suchen und nicht in immer bestimmteren Definitionen. Das menschliche Denken kann und muss sich immer wieder selbst in Richtung auf die Gewinnung neuer möglicher Wahrheit und Erkenntnis hin überschreiten:

„Es zeigt sich nämlich, daß bei einem systematischen Aufstellen der Axiome der Mathematik immer wieder neue und neue[[re]] Axiome evident werden, die nicht formallogisch aus den bisher aufgestellten folgen ... eben dieses Evidentwerden immer neuerer Axiome auf Grund des Sinnes der Grundbegriffe ist etwas, was eine Maschine nicht nachahmen kann."[26]

22 H. Scholz, David Hilbert, Altmeister der mathematischen Grundlagenforschung (1942), in: ders., Mathesis universalis, ²1969, 279–290, 289.

23 Vgl. J. W. Dawson, The Reception of Gödel's Incompleteness Theorem, in: S.G. Shankar (Hg.), Gödel's Theorem in Focus, London 1988, 74–95.

24 H. Scholz, David Hilbert, Altmeister der mathematischen Grundlagenforschung (1942), 289.

25 H. Mehrtens, Moderne – Sprache – Mathematik, 1990, 298.

26 Vgl. den Vortragsentwurf *The modern development of the foundations of mathematics in the light of philosophy* von 1961 in: K. Gödel, Collected Works vol. III, hg. von S. Fefermann et al., New York/Oxford 1995, 372–387, 384.

6. Künstliche Intelligenz und Unentscheidbarkeit in theologischer Perspektive

Aus dem bis jetzt Vorgeführten lassen sich in mancherlei Hinsicht Konsequenzen und Perspektiven entwickeln, die für die Theologie und Ethik von einiger Bedeutung sind. Ich möchte dazu einige relevante Überlegungen entwickeln, die sich den folgenden drei Fragestellungen zuordnen lassen:

1. Was bedeutet künstliche Intelligenz für unser Menschenbild?
2. Welche ethischen Fragen und Problembereiche in Bezug auf Systeme künstlicher Intelligenz schließen sich an?
3. Was bedeuten diese Überlegungen für ein theologisches Gottesbild und die Erkenntnisleistung bzw. Erkenntnisgrenzen von Theologie?

6.1 Zum Menschenbild

Zunächst einmal legen die Gödelschen Theoreme nahe, dass eine vollständige objektive Theorie des menschlichen Bewusstseins nicht möglich ist. Selbst wenn das Gehirn vollkommen der bekannten Physik gehorchte, brächte doch eine Eigenschaftsbeschreibung mit Hilfe finiter Analyse nicht alle Aspekte des menschlichen Geistes in einer Theorie erschöpfend zur Darstellung, vor allem nicht die so wichtigen und entscheidenden Aspekte der selbstbezüglichen Subjektivität. Aus empirischen Daten allein könnte dann nicht vollständig auf subjektive Erlebniszustände in finiter Zeit und mit endlichen Mitteln geschlossen werden.

Aber auch eine Erklärung des Bewusstseins im Rahmen eines evolutionistischen Weltbildes im Sinne eines Fitness steigernden Problemlösungsapparates ist dann als unvollständig anzusehen. Seit der Etablierung der Evolutionstheorie als dem durchgängigen Paradigma zur genetischen Erklärung biologischer Systeme wurde auch der kognitive Apparat des Menschen in Analogie zu rein physiologischen Körperorganen als Instrument zur Sicherung eines Überlebensvorteils gesehen. Als eine Stimme unter vielen mag hier der Physiker Ludwig Boltzmann zu Wort kommen:

„Das Gehirn betrachten wir als den Apparat, das Organ zur Herstellung der Weltbilder, welches sich wegen der großen Nützlichkeit dieser Weltbilder für die Erhaltung der Art entsprechend der Darwinschen Theorie beim Menschen geradeso zur besonderen Vollkommenheit herausbildete, wie bei der Giraffe der Hals, beim Storch der Schnabel zu ungewöhnlicher Länge."[27]

Das ist zunächst als genetische Erklärung nicht falsch, doch als hinreichende Beschreibung unzulänglich. Der Überlebensvorteil der eigentlichen Verstehensdimension des Bewusstseins wird dann jedenfalls obsolet, wenn man es auf bloße Steuerungs- und Regelungsfunktionen reduziert. Dazu hätte ein bewusstloser zentraler kybernetischer

27 L. Boltzmann, Über die Frage nach der objektiven Existenz der Vorgänge in der unbelebten Natur, in: ders., Populäre Schriften, 1905, 162–187, 179.

Apparat ausgereicht. Doch ein solcher Apparat wäre nicht in der Lage gewesen, wirkliches Verstehen und semantische Offenheit zu erzeugen. Das Fragen nach und das Verstehen von Bedeutungen aber sind notwendige Bedingungen für die Entstehung von Sprache, Kommunikation, Sozialität und Wissenschaft, aber auch von Religion.

Grundlegend für die verstehende Wahrnehmung von Welt ist ja die Differenz zwischen wahrnehmendem Subjekt und wahrgenommener Wirklichkeit. Nur durch die Unhintergehbarkeit dieser Unterscheidung, die ihrerseits nicht wieder durch das erkennende System rekonstruierend eingeholt werden kann, ist z. B. auch der Aufwand unseres Wahrnehmungsapparates erklärbar, eine einheitliche Erfahrungswelt zu konstruieren, in der selbst der blinde Fleck unseres Gesichtsfelds noch wegretuschiert wird und uns eine kohärente Welt gegenübertritt. Als bewusste Wesen haben wir eine *Welt*, in der wir sind und der wir zugleich gegenüber stehen, ein General Problem Solver zur Regelung von Prozessen bräuchte nur eine Menge von *Informationen*. Und ebenso *haben* wir nicht bloß ein Selbstmodell als interne formale Repräsentation unseres Weltbezugs[28], sondern *sind* wir ein wirkliches Ich. Wir brauchen das Gegenüber von Welt und Subjekt, um die semantische Geschlossenheit eines nur formalen Problemlösens durchbrechen zu können.

Unsere Subjektivität ist allerdings wiederum enthalten in einer anderen, sie mit umfassenden Hierarchieebene, in der auch die Differenz zwischen Subjekt und Welt wieder eingebettet ist, in die soziale Kommunikationsgemeinschaft von Subjekten. Ein Mensch wird erst am Du zum Ich, lautet die schlichte Form, in die Martin Buber diese Einsicht gebracht hat. Wir müssen erst durch die Beziehungen, mit denen wir aufwachsen, und durch die Sprache, mit der wir uns in Auseinandersetzung mit anderen entwerfen, lernen, dass wir ein Subjekt sind. Wir brauchen den anderen, um ein Selbstverhältnis, um Subjektivität und Personalität entwickeln zu können. Zugleich aber ist auch deutlich, warum wir aus der Ich-Perspektive nicht aussteigen können. Dies hat Douglas Hofstadter auch als Konsequenz der Gödelschen Theoreme zur Geltung gebracht:

> Sie stellen „eine mathematische Analogie zu dem Umstand dar, daß ein Verständnis davon, wie es ist, wenn man ... eine Fledermaus ist, nur mittels einer unendlichen Reihe von immer genaueren Simulationsvorgängen möglich ist ... Das Gödelsche Theorem ist eine Folgerung aus diesem allgemeinen Sachverhalt: Ich bin mein eigener Gefangener und kann deshalb nicht sehen, wie andere Systeme mich sehen."[29]

[28] So versteht Thomas Metzinger z. B. menschliche Gehirne als „*General Problem Solvers*" und vergleicht sie mit Flugsimulatoren, die jedoch zugleich unser Selbstbewusstsein als ein internes Selbstmodell zusammen mit den Szenarien der Außenwelt generieren: „*Menschliche Gehirne simulieren den Piloten gleich mit*" (Th. Metzinger, Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation, 1993, 243).

[29] D. R. Hofstadter/D. C. Dennett, Einsicht ins Ich, 1986, 398.

Doch auch die Pluralität der Semantiken, der Interpretationen, die wir kommunikativ generieren, kann ihrerseits nicht wieder in einem Super-System vollständig zusammengefasst werden. Denn auch und gerade das kommunikative Handeln von Subjekten mit privatem Selbstbewusstsein kann nicht rein funktional formalisiert werden, denn das Ich des anderen bleibt auf direkte Weise unzugänglich. Das meint nicht, dass es nicht funktionalen Zwecken dient, es lässt sich nur nicht auf solche reduzieren, ohne seine Eigenart der Generierung und Zuschreibung von Bedeutung und Sinn zu verlieren. Auch soziales und kommunikatives Handeln lässt sich nicht in bloß instrumentelles Handeln auflösen.

Insofern wir nicht nur funktional und regelgeleitet operieren, gehört eine auch theologisch höchst relevante Kategorie mit zur Charakterisierung von Personalität und Menschsein: die in technischen Zusammenhängen zumeist nur als Störung zu verstehende Kategorie der *Unterbrechung*[30]. Wir können über uns erschrecken, wenn wir bemerken, dass wir in unserem Leben nur ein Programm abspulen, dass wir in ein dumpfes, regelgeleitetes *Funktionieren* abgeglitten sind. Ob auch eine auf Optimierung hin programmierte kybernetische Steuerungsmaschine über sich erschrecken und aufgrund eines empfundenen Defizits in ethischer oder geistiger Hinsicht sich selbst unterbrechen und fragen könnte: Was machst du da eigentlich? Warum funktionierst du bloß? Es ist jedenfalls ein wichtiger Aspekt von Religion, dass sie solche Unterbrechungen herbeiführen will, im Gottesdienst, in Gebet und Meditation und anderem mehr. Aber auch weltlich macht sich ein spontaner Einfall oder eine intuitive Wahrnehmung als Unterbrechung von Routinen bemerkbar.

Bloße Funktionalität jedenfalls wäre Grundlage eines dürftigen Menschenbildes. Sie ignoriert die letzte Unverfügbarkeit, in der wir uns als handelnde Subjekte gegenüberstehen, und sie ignoriert, dass wir uns Beziehungen verdanken, die nicht formal beherrschbar sind, ohne dass man sie zerstört. An diese Grundstruktur des Menschseins knüpft die theologische Rechtfertigungslehre an: wir sind nicht unser eigenes Werk, auch nicht das Machwerk eines uns hergestellt habenden Schöpfers, und wir dürfen und sollen weder unserem herstellenden Handeln unterwerfen, was durch Beziehung konstituiert wird, noch umgekehrt das, was durch Beziehung konstituiert ist, durch hergestellte Objekte ersetzen.

Die Gefahr der Maschinenmetapher ist dann die, dass angesichts der Tatsache, dass der Mensch in vielen Zusammenhängen immer mehr und immer effektiver verzweckt wird, die Behauptung der Simulation humaner Eigenschaften den Trend verstärkt, dass extern vorgegebene Effizienzkriterien den Maßstab bilden für gelingendes Dasein.

30 Vgl. E. Jüngel, Wertlose Wahrheit. Christliche Wahrheitserfahrung im Streit gegen die ‚Tyrannei der Werte', in: Ders., Wertlose Wahrheit. Zur Identität und Relevanz des christlichen Glaubens (Theologische Erörterungen III), 90–109, 100 ff.

Das Problem in der Auseinandersetzung mit Robotik und künstlicher Intelligenz scheint mir deshalb nicht so sehr zu sein, dass Roboter und Computer sich *uns* immer mehr angleichen, so dass wir in unserer Einzigartigkeit gekränkt würden, sondern dass wir uns genötigt sehen oder versucht sind, uns *ihnen* immer mehr anzugleichen. Maschinen, so sahen wir, arbeiten heteronom nach extern vorgegebenen Effizienzkriterien, die sie passiv und, wenn meine Interpretation richtig ist, auch ohne Verstehen erfüllen. Diese Verzweckung scheint auf den ersten Blick zu kongruieren mit einer Sicht der Evolution, die als Überleben des Tüchtigeren verstanden wird und sich dadurch als das Gesetz der Natur schlechthin imponiert. Und zugleich ist dies das Gesetz der globalisierten Ökonomie, dem immer mehr Bereiche menschlicher Existenz unterworfen werden.

„Wo Natur und Gesellschaft gemeinsam mitwirken, ein auf bloßes Effizienzdenken reduziertes Menschenbild zu zementieren, setzt sich eine weltanschauliche Grundstimmung durch, die die These plausibel erscheinen lässt, auch der Mensch sei schließlich nichts anderes als ein Roboter."[31]

Hier scheinen mir die eigentlichen Bedenken gegen die Einebnung des Unterschiedes zwischen natürlicher und künstlicher Intelligenz zu liegen, und nicht in der allzu abstrakten Furcht davor, dass der Mensch in seiner Einzigartigkeit gekränkt werden könnte. Damit kommt die eigentlich ethische Dimension unseres Themas in den Blick.

6.2 Ethische Fragen im Umgang mit künstlicher Intelligenz

Für die folgenden Erörterungen müssen wir noch einmal zu der Frage zurückkommen, ob und wie wir denn künstliche Intelligenz herzustellen in der Lage sind. Ich hatte als eine Konsequenz aus den Gödelschen Theoremen festgehalten, dass bei rein formal arbeitenden Entscheidungssystemen ein semantisches Verstehen im Rahmen der eigenen Prozesse nicht anzunehmen ist. Doch wenn das nicht durch technische Konstruktion direkt hergestellt werden kann, dann kann es vielleicht, und so hatte auch schon Turing argumentiert, durch lernende Systeme erzeugt werden[32], die vielleicht noch in einen sozialen und kommunikativen Kontext mit Menschen eingebettet sind und denen möglicherweise funktionale Einheiten zugrunde liegen, die nicht vollkommen diskret arbeiten.

31 H.-D. Mutschler, Ist der Mensch ein Roboter?, in: M. Kossler/R. Zecher (Hg.), Von der Perspektive der Philosophie. Beiträge zur Bestimmung eines philosophischen Standpunkts in einer von den Naturwissenschaften geprägten Zeit, 2002, 291–308, 306.

32 Vgl. A. Turing, Computing Machinery and Intelligence (1950), 177: „Warum sollte man nicht versuchen, statt ein Programm zur Nachahmung des Verstandes eines Erwachsenen eines zur Nachahmung des Verstandes eines Kindes herzustellen? Unterzöge man dieses dann einem geeigneten Erziehungsprozeß, erhielte man den Verstand eines Erwachsenen."

Dass dies möglich sein könnte, lässt sich sicher nicht ausschließen. Mit anderer Hardware und offeneren Programmstrukturen mag es möglich sein, Systeme zu entwickeln, denen wir nicht Bewusstsein einprogrammiert haben, sondern bei denen so etwas wie Bewusstsein im Sinne von semantischem Verstehen im Laufe von Lernprozessen sich einstellen kann. Dann aber stellt sich nicht mehr nur die Frage, ob und wie wir solche Maschinen konstruieren *können*, sondern auch, ob wir das *wollen*.

Auch von solchen Maschinen würde gelten, was wir über menschliches Bewusstsein selbst gesagt haben, dass ihr innerer verstehender Zustand von außen in direkter Weise unzugänglich wäre. Wieder hat schon Turing dies festgehalten:

„Ein wichtiges Kennzeichen einer lernenden Maschine ist, daß ihr Lehrer oft reichlich wenig von dem weiß, was genau in ihr vorgeht, wenn er auch bis zu einem gewissen Grad dennoch in der Lage sein mag, das Verhalten seines Schülers vorauszusagen."[33]

Wenn wir das Vermögen von semantischer Differenz und Offenheit auf einer Maschine realisieren wollen, dann müssen wir zulassen, dass die Maschine selbständig agiert und auch die sie leitenden Zwecke modifiziert. Sie könnte sich möglicherweise eigene Zwecke setzen, gesetzte Zwecke aufheben. Ein Roboter, dem wir Bewusstsein und verstehende Intelligenz zuschreiben könnten, müsste sich also in einem fundamentalen Sinne sein Gesetz selbst geben können. Er wäre auch in der Lage, sich zu unterbrechen und sich zu fragen, was er da eigentlich macht und warum, er könnte streiken. Kurz, ein solches Wesen würde dem Begriff der Maschine widersprechen, die ja auf externe Zwecke hin optimierte Funktionen ausführen soll.

Die Frage ist also, ob wir *solche* Roboter bauen wollen. Doch warum bauen wir Systeme mit möglichst großer künstlicher Intelligenz eigentlich überhaupt? Drei Gründe scheinen mir vor allen Dingen bedeutsam:

1. Wir können objektivierende, berechnende Instrumente zur Entscheidungsvorbereitung gut gebrauchen, gerade weil wir selbst nicht rein instrumentell funktionieren (Expertensysteme).
2. ‚Intelligent' gesteuerte mechanische Maschinen können vielfältige Aufnahmen übernehmen und Dinge tun, die uns körperlich schwer fallen, bei denen aber komplexe Entscheidungen notwendig sind (Roboter in schwieriger Mission im Weltraum, bei der Verbrechensbekämpfung, in Produktionsprozessen), und die uns intellektuell schwer fallen, weil sie in ihrer Komplexität für uns heuristisch schwer zu bewältigen sind (Internet).
3. Wir wollen wissen, „wie es geht", und damit auch darauf schließen können, wie wir selbst funktionieren. Bei den Robotern, mit denen wir direkt und ausschließlich menschenähnliches Verhalten nachkonstruieren, dürfte die Neugier auf die konstru-

[33] AaO., 181.

ierte Vernunft im Vordergrund stehen. Und da hinein mischen sich sicher auch Visionen, die uns künstliche Systeme geradezu als die besseren Menschen vorführen.

Einige Sätze von Hans Moravec, einem Roboterbauer, mögen dies illustrieren:

„Ich sehe diese Maschinen als unsere Nachkommen. Im Augenblick glaubt man das kaum, weil sie eben nur so intelligent sind wie Insekten. Aber mit der Zeit werden wir das große Potential erkennen, das in ihnen steckt. Und wir werden unsere neuen Roboterkinder gern haben, denn sie werden angenehmer sein als Menschen. Man muß ja nicht all die negativen menschlichen Eigenschaften, die es seit der Steinzeit gibt, in diese Maschinen einbauen. Damals waren diese Eigenschaften für den Menschen wichtig. Aggressionen etwa brauchte er, um zu überleben. Heute, in unseren großen zivilisierten Gesellschaften machen diese Instinkte keinen Sinn mehr. Diese Dinge kann man einfach weglassen – genauso wie den Wesenszug der Menschen, daß sie ihr Leben auf Kosten anderer sichern wollen. Ein Roboter hat das alles nicht. Er ist ein reines Geschöpf unserer Kultur und sein Erfolg hängt davon ab, wie diese Kultur sich weiterentwickelt. Er wird sich also sehr viel besser eingliedern als viele Menschen das tun. Wir werden sie also mögen und wir werden uns mit ihnen identifizieren. Wir werden sie als Kinder annehmen – als Kinder, die nicht durch unsere Gene geprägt sind, sondern die wir mit unseren Händen und mit unserem Geist gebaut haben."[34]

Auch wenn solche Visionen reine Fiktion sind und hier vieles nüchtern abgewartet werden kann, sollte man vielleicht einige Leitlinien bei der Beantwortung der Frage, welche intelligenten Maschinen wir denn zu welchen Zwecken haben wollen, durchaus entwickeln. Ich will dies in einigen wenigen Punkten versuchen.

1. Ein Hauptproblem ist die Frage nach der *Delegation von Verantwortung* an Maschinen. Effizienzdenken und Fragen der Haftung fördern die Versuchung, persönliche Verantwortung zu delegieren an Expertensysteme. Eine deutliche Grenze aber setzen hier wohl *qualifizierte ethische Entscheidungen* mit einer Tragweite, die menschliches Leben direkt betreffen. Es verbietet sich, ein Expertensystem z. B. über das Abschalten einer Herz-Lungen-Maschine entscheiden zu lassen oder ein Waffensystem über den Einsatz von lebensvernichtenden Waffen oder ein Strategieprogramm über politische Entscheidungen.
2. Aus einem ähnlichen Grund ist auch die *Jurisprudenz* als eine Kunst zu würdigen, die geübt sein will, die humane Urteilskraft und Verantwortlichkeit erfordert und die nicht das Resultat eines mechanisch anwendbaren Verfahrens sein kann, aber auch nicht an Expertensysteme delegiert werden darf. Es ist sicher kein Zufall, dass der Kantsche Begriff der ‚Urteilskraft' ursprünglich aus der Rechtssphäre stammt. Wer Urteile fällt, braucht diese ‚Urteilskraft', die nicht vollständig formalisierbar ist und auch allein in persönlicher Verantwortung stehende Billigkeits-Erwägungen mit einschließt. Aber auch viele andere Gebiete menschlichen Erkennens und Handelns kommen ohne Urteilskraft nicht aus. Sie vollziehen sich zwar nicht chaotisch, sondern durchaus nach Regeln, diese aber sind so offen gestaltet, dass sie persönlicher Verantwortung und Interpretation einen weiten Spielraum lassen, wie z. B. *Pädago-*

34 H. Moravec, Robot: Mere Machine to Transcend Mind, Oxford 1998; dt. Übersetzung: Computer übernehmen die Macht: Vom Siegeszug der künstlichen Intelligenz, 1999, 136.

gik, *Ästhetik* oder literarische *Hermeneutik*. Auch hier ist Skepsis geboten, wenn allzu viel formalisierte Maschinenintelligenz darin Einzug hält.

3. Ein weiterer Punkt ist die Frage nach der *sozialen Komponente* der Computer-Mensch-Beziehung. Festzuhalten ist, dass eine Interaktion mit Systemen künstlicher Intelligenz nicht die Interaktionen zwischen Menschen ersetzen kann und darf. Die Unterhaltung mit einem vielleicht bald realisierten Alten-Pflegeroboter kann die menschliche Begegnung nicht ersetzen.

Gegen diese Möglichkeiten gilt es, nicht einfach zu behaupten, dass nicht sein kann, was nicht sein darf, sondern die Frage wach halten, ob das sein muss, was sein kann. Es gilt also auch die Eigendynamik der technischen Entwicklungen zu unterbrechen und die Frage „Wozu?" zu stellen, damit nicht nur unter Absehung von Endzwecken Teilmechanismen optimiert werden, die dann zusammen mit politischen und ökonomischen Zwängen problematische Trends freisetzen, auch wenn wir auf die Frage nach dem Endzweck des Menschen unterschiedliche und selbst immer wieder nur vorläufige Antworten erhalten. Dann ist es sehr bedenklich, erscheint unter technischen und ökonomischen Gesichtspunkten aber auch als überflüssig, bewusste künstliche Wesen zu erzeugen, die nicht als bloße Maschinen anzusehen wären, die rein äußeren Zwecken dienen. Und unsere Neugier, wie wir denn funktionieren (und wir funktionieren Gott sei Dank ja auch nicht nur!) würde erst recht nicht befriedigt, denn wir könnten diese Maschinen wohl auch nicht besser verstehen als uns selbst.

6.3 Glauben und Wissen

Werfen wir zunächst noch einmal kurz einen Blick zurück auf das Menschenbild, das wir oben entwarfen, und entfalten wir dann den Begriff von Gott als Schöpfer, so dass er darauf bezogen werden kann. Wenn der Mensch nicht ein evolutionär optimierter Problemlösungsapparat ist, in seinem Sein und Wesen nicht externe Zwecke zur Geltung kommen, sondern der Mensch als semantisch offenes Wesen sich selbst und seine Welt zu verstehen sucht und nur als relationales Wesen existieren kann, dann ist das Bekenntnis zu Gott als seinem Schöpfer nicht der Hinweis auf einen transmundanen Ingenieur oder Programmierer. Der sich in den ungeheuren Räumen und Zeiten des Kosmos vollziehende Entwicklungsprozess der Schöpfung ist nicht Ausdruck eines herstellenden Handelns. Wenn in einer besonders ausgezeichneten Nische aus überaus verhaltenen Anfängen und mit und aus einer überschwänglichen Fülle von Gestalten auch der Mensch als ein Wesen entsteht, das nach sich selbst fragt, dann ist der Mensch ebenso wenig wie alle anderen Kreaturen das Machwerk eines göttlichen Konstrukteurs. Wir sind, wie alle Schöpfung, zunächst um unserer selbst willen da. Und im Blick auf unser Dasein und Sosein ist Gott nicht als unser Hersteller zu preisen, sondern als die Quelle, der Grund und der Antrieb der Fülle von Möglichkeiten, der seiner Schöpfung gegenwärtig bleibt und uns als solchermaßen ent-

stehende und vergehende Geschöpfe bejaht und in die Eigentlichkeit unserer beziehungsreichen Existenz ruft.

Das heißt dann aber auch, dass der Gedanke eines unabhängig von der Wirklichkeit der sich vollziehenden Schöpfung über das Ganze des Geschehens „voll informierten Schöpfers"[35] abzuweisen ist. Oder genauer, und also mit Gödel gesagt, aus der Annahme, dass der Inbegriff aller Wahrheiten in einem einzigen System erfasst werden könnte, lässt sich ein Widerspruch herleiten, wenn vorausgesetzt ist, dass das System widerspruchsfrei ist. An Wittgensteins bekanntem Diktum, das die Welt die Gesamtheit der Tatsachen ist, ist dies problematisch, dass die wahren Tatsachen keine Gesamtheit, keine widerspruchsfreie und vollständige Totalität darstellen[36]. Es ist auch in einer semantisch offenen Sprache kein Subjekt widerspruchsfrei denkbar, das „einen epistemisch ausgezeichneten externen Standpunkt einnehmen" könnte, „von dem aus es zu allen wahren Aussagen bzw. Sachverhalten epistemischen Zugang besitzt"[37]. Der theologische Begriff von Gottes „Wahrheit" dürfte deshalb wesentlich vom Begriff der „Wahrhaftigkeit" Gottes her zu bestimmen sein und nicht von der Vorstellung einer Totalität von Satzwahrheiten[38].

Deshalb ist umgekehrt auch kein Gottesbeweis möglich, denn ein Begriff unüberbietbarer Totalität ist für uns unerreichbar. Jeder Beweis ist nur relativ zu einem System von Voraussetzungen zu führen. Einen voraussetzungslosen Beweis gibt es nicht. Und das Absolute ist von relativen Voraussetzungen her gerade unerreichbar. Die Behauptung, dass Gott existiert, ist deshalb zwar eine notwendige Implikation und Voraussetzung des Glaubens, zugleich aber ist die Frage nach ihrer Wahrheit eine in absolutem Sinn unentscheidbare Frage.

Der Glaube verlangt aber auch nicht nach Beweisen, nicht nach Vollständigkeit und Widerspruchsfreiheit von Argumenten oder Systemen. Denn auch das beste Wissen könnte den Glauben unter Bezug auf den derzeitigen Stand des Wissens nur wahrscheinlich machen – doch „das Christentum als das Wahrscheinliche – ... dann ist das Christentum abgeschafft"[39]. Eher verlangt der Glaube seinerseits nach einer Unterbre-

35 M. Eigen/R. Winkler, Das Spiel, ^{5}1983, 224.

36 Vgl. P. Grim, The Incomplete Universe, 1991, 2 f. und 6.

37 E. Brendel, Wahrheit und Wissen, 1999, 165.

38 Das dürfte auch in Joh 14,6 gemeint sein, wo der Gottessohn von sich sagt, er sei der Weg, die Wahrheit und das Leben. Das Semitische kennt kein Abstraktum, und so gehören alle drei Begriffe zusammen und sind auf der gleichen semantischen Ebene anzuordnen. Der Bezugpunkt unseres Denkens, Glaubens und Meinens sind dann nicht die ewigen, im Verstande Gottes residierenden Vernunftwahrheiten, sondern es ist die Wahrhaftigkeit Gottes, die wir erfahren in seiner Schöpfung, in seiner Rechtfertigung, auf die wir uns verlassen in konkreten Situationen des Lebens, auf die wir hoffen in Bezug auf die Bejahung unseres Daseins.

39 S. Kierkegaard, Søren Kierkegaards Papirer, Kopenhagen ^{2}1968 ff., V, Bd. X–4, A 633.

chung, nach einer Suspension der diskursiven, der formalisierten Vernunft, die Raum gibt dafür, vertrauensvoll zu leben. Es geht der Religion gerade um die Dimension menschlichen Daseins, die sich nicht beweisend einfordern oder formal herstellen lässt und die doch fundamental ist für unsere beziehungsreiche Existenz:

„Sympathie, Vertrauen, wechselseitige Achtung und Anerkennung, Freundlichkeit, Gütigkeit, oder, im Kontext christlicher Glaubenskommunikation, Liebe oder Barmherzigkeit."[40]

Religion versucht diese Dimension des Lebens zugänglich und kommunizierbar zu machen, ohne sie selbst wieder zu formalisieren und so unter Kontrolle zu bringen. Das versuchen allenfalls Aberglauben und Magie.

Aber auch die weltliche und wissenschaftliche Vernunft, das scheinen mir die Gödelschen Theoreme nahe zu legen, nimmt immer schon Vertrauen in Anspruch. Verständnis ist ohne vorgängiges Einverständnis nicht möglich, eine Rationalität ohne Vertrauen ist unvernünftig, eine Wahrheit ohne Wahrhaftigkeit ist leer.

Damit ergibt sich, was auch der Wissenschaftstheoretiker und Philosoph Wolfgang Stegmüller als Konsequenz aus der Analyse der Logik und der Gödelschen Theoreme festgestellt hat:

„Eine ‚Selbstgarantie' des menschlichen Denkens ist, auf welchem Gebiete auch immer, ausgeschlossen. Man kann nicht vollkommen ‚voraussetzungslos' ein positives Resultat gewinnen. Man muß bereits an etwas glauben, um etwas anderes rechtfertigen zu können. Mehr könnte sinnvollerweise nur dann verlangt werden, wenn wir die Endlichkeit unseres Seins zu überspringen vermöchten. Aber der archimedische Punkt außerhalb unserer endlichen Realität bleibt, zumindest für uns, eine Fiktion."[41]

Die christliche Theologie ist deshalb primär zu verstehen als eine Heuristik zur Orientierung im Offenen und Endlichen, nicht als abgeschlossene Theorie über transzendente Gegenstände. Und auch für die Theologie gilt, „auf Letztbegründungsversuche und damit auf die Suche nach unhintergehbaren Gegebenheiten, letzten Identitäten und unhinterfragbaren außertheologischen Sachverhalten ganz zu verzichten".[42] Auch ihr stehen keine zwingenden letzten Gründe, „sondern nur relative, interne und lokale Kriterien"[43] zur Verfügung.

[40] J. Fischer, Pluralismus, Wahrheit und die Krise der Dogmatik, ZThK 91 (1994), 487–539, 496.

[41] W. Stegmüller, Metaphysik Skepsis Wissenschaft, ²1969, 307.

[42] I. U. Dalferth, Subjektivität und Glaube. Zur Problematik der theologischen Verwendung einer phi-losophischen Kategorie, NZSTh 36 (1991), 18–58, 49. Dalferth hat diese und die folgenden Äußerungen gegen subjektivitätstheoretische Fundierungsversuche der Theologie eingewandt, sie sind aber in analoger Weise auch gegen jede „Neuauflage kosmologischer oder absolutheitstheoretischer Letztbegründungen" (ebd.) zur Geltung zu bringen, die etwa über das anthropische Prinzip eine kosmisch fundierte ‚Theorie über alles' einschließlich der menschlichen Existenz und ihrer Lebensfragen entwerfen.

[43] Ebd.

Damit aber steht sie als Erkenntnisbemühung nicht isoliert da im Zusammenhang verstehenden Denkens des Menschen überhaupt, mit dem er auch und besonders nach sich selbst fragt. Und so möchte ich schließen mit einem gewagten Satz, der aber aus unverdächtigem Munde oder besser unverdächtiger Feder, nämlich noch einmal von Wolfgang Stegmüller stammt:

„Denken, Lieben und Beten haben also *etwas* gemeinsam."[44]

[44] W. Stegmüller, Metaphysik Skepsis Wissenschaft, 456.