Gibt es eine Analogie zwischen einer elektronischen Rechenmaschine und dem Gehirn?

Herrn Prof. W. Heitler zum 60. Geburtstag gewidmet

Von

KURT WEISS

Institut für theoretische Physik der Universität Zürich

Einleitung

Es ist üblich geworden, die grossen elektronischen Rechenmaschinen «Elektronengehirne» zu nennen. Durch diese Bezeichnung wird die Meinung nahegelegt, dass das Gehirn nichts weiter als eine besonders komplizierte Rechenmaschine ist. Man ist zwar bereit, zuzugeben, dass eine Rechenmaschine kein Bild malen, kein Gedicht schreiben kann etc.; krass zeigt sich das Unvermögen der Maschine, irgend etwas zu tun, was man ihr nicht explizit beigebracht hat, bei der Verwendung zur Übersetzung von Sprachen. Es ist aber weniger bekannt, dass die Maschine sich selbst auf ihrem ursprünglichen Anwendungsgebiet — der Mathematik — grundlegend vom Gehirn unterscheidet. Dies zu zeigen ist das Hauptthema dieses Aufsatzes. Die Anregung dazu lieferte eine Diskussionsbemerkung von Prof. M. FIERZ im Anschluss an einen Vortrag über die Lernfähigkeit von Maschinen und des Gehirns. In dieser Bemerkung wurde angedeutet, dass die metamathematischen Resultate von GOEDEL und CHURCH vielleicht einen Beitrag zum Verständnis des Verhältnisses von Maschine und Gehirn zu liefern imstande sind. Im Anschluss daran ergab sich bei Diskussionen in unserem Institut, dass der angedeutete Tatbestand eine präzisere Untersuchung und Darstellung erheischt. Es soll in der vorliegenden Arbeit versucht werden, eine solche Untersuchung zu geben.

Im § 1 wird kurz umrissen, warum es möglich ist, überhaupt an eine Analogie zwischen einer elektronischen Rechenmaschine und dem Gehirn zu denken. Basierend auf der Darstellung J. VON NEUMANNS [1] (Zahlen in Klammern beziehen sich auf das Literaturverzeichnis am Schlusse des Aufsatzes), ergänzt durch neuere Daten, wird bereits an dieser Stelle darauf hingewiesen, wie gewaltig die Unterschiede zwischen Gehirn und Maschine sind. Im § 2 werden bekannte naturwissenschaftliche Argumente wiedergegeben, die Indizien zur Verschiedenheit von Rechenmaschine und Gehirn enthalten. Die Bemerkung von Prof. K. AKERT [2]: «Noch hat die moderne Hirnwissenschaft erst eine kurze Strecke zurückgelegt, und sie hat einen weiten Weg bis zum Ziel vor sich. Aber es scheinen genug Anhaltspunkte da zu sein, welche die Richtung der naturwissenschaftlichen Arbeit klar vorzeichnen» wird in dem Sinne erläutert, dass gezeigt wird, dass die naturwissenschaftliche Untersuchung des Gehirns ganz andere Probleme stellt und andere Methoden bedingt, als die Analyse der physikalischen Vorgänge in der Rechenmaschine. Im § 3 wird eine erste Übersicht über den Hauptgegenstand des Aufsatzes gegeben, indem die Logik der Rechenmaschine mit der des Gehirns vorläufig verglichen wird. Im § 4 werden die metamathematischen Resultate von GOEDEL und CHURCH dargestellt (das Hauptresultat von GOEDEL wurde in nichtformalisierter Form schon 5 Jahre vor der Arbeit GOEDELS [15] von P. FINSLER [16] ausgesprochen und durch inhaltliches [im Gegensatz zum formalen] Schliessen bewiesen); im § 5 wird ihr Zusammenhang mit Eigenheiten der menschlichen Sprache aufgezeigt. Der § 6 ist der Relevanz der genannten Methoden für den Unterschied der Logik der Rechenmaschine und derjenigen des Gehirns gewidmet. Es wird sich als Resultat ergeben, wie wenig, falls eine Analogie zwischen Maschine und Gehirn überhaupt angenommen wird, damit über das Gehirn und unser Denken ausgesagt ist.

Es sei noch zum vornherein festgestellt, dass diese Arbeit in keiner Weise den Anspruch stellt, philosophische Aussagen zu machen. Sie ist vielmehr als Versuch aufzufassen, die logistische Strukturanalyse in ihrer begrenzten Zuständigkeit zur Erforschung des Gehirns darzustellen. Es soll aber daraus nicht gefolgert werden, der Vergleich des Gehirns mit einer Maschine sei gänzlich unfruchtbar. Erstaunlich, und zu einem grossen Teil auch nützlich, sind die in einer elektronischen Rechenmaschine realisierten Analoga zu einfachen Denkprozessen. Wir beschränken uns hier grundsätzlich auf mathematische Denkprozesse. Aber selbst dann zeigt sich, dass mit einer solchen Analogie verschwindend wenig über das mathematische Denken selber ausgesagt ist, oder positiv ausgedrückt: Es zeigt sich, dass grundsätzliche qualitative Unterschiede zwischen einer Maschine und dem Gehirn vorhanden sind.

1. Elektronische Rechenmaschinen und das Gehirn

§ 1. Die Analogie Rechenmaschine—Gehirn

In einer digitalen elektronischen Rechenmaschine sind gewisse Funktionselemente (Transistoren etc.), die elektromagnetische Impulse beeinflussen (hemmen, leiten, verkoppeln etc.), verbunden durch eine Verdrahtung, die dazu dient, die Impulse von Funktionselement zu Funktionselement zu transportieren. Ein reiner Analogrechner dagegen verfügt über Vorrichtungen zum Abtasten von Kurven, zum Zeichnen von Kurven etc., indem z. B. zu den Koordinaten-Werten proportionale elektrische Spannungen erzeugt werden; und ausserdem verfügt er über physikalische Anordnungen zur rechnerischen Verarbeitung der gelesenen Daten (z. B. Addition der Drehgeschwindigkeiten von Scheiben etc.). Die rechnerische Verarbeitung erfolgt aber meistens mit einer digitalen Maschine, was den Vorteil viel grösserer Genauigkeit und Rechengeschwindigkeit hat. Da es ausserdem äquivalent ist (wenn auch i. a. zeitraubend und maschinentechnisch ungünstig [begrenzte Speicherkapazität]), die abzutastende Kurve numerisch auf Lochkarten (oder ähnliches) zu übertragen oder sie umgekehrt aus den Lochkarten-Werten aufzuzeichnen, soll die weitere Diskussion auf digitale Maschinen beschränkt bleiben. Eine genauere Diskussion dieser Frage findet man in [1].

Neben den Funktionselementen und der sie verbindenden Verdrahtung tritt bei der Rechenmaschine als weiteres wesentliches Element noch das «Gedächtnis» auf. Es besteht aus einer grossen Zahl von geeigneten physikalischen Einheiten, die binäre Informationen («bits») zu speichern imstande sind (z. B. Ferritkerne mit zwei möglichen Magnetisierungsrichtungen).

Im Gehirn lässt sich eine ähnliche Einteilung treffen. Die Funktionselemente (Neuronen, Synapsen und eventuell andere) sind durch eine «Verdrahtung» (Axone) miteinander verbunden. Ein Verdrahtungselement ist dadurch ausgezeichnet, dass in ihm die materiellen Vorgänge, die die Informationsübertragung von Funktionselement zu Funktionselement bewerkstelligen (im für die Denkvorgänge wesentlichen Gehalt), ungeändert transportiert werden. Das dritte Element, das Gedächtnis, ist im Gehirn durch Mechanismen realisiert, deren Erforschung noch im Anfangszustand steht. (Einerseits wurde gefunden, dass Abläufe, die oft wiederholt werden, mit der Zeit immer leichter ablaufen, solche aber, die lange Zeit nicht auftreten, immer stärkere Hemmnisse überwinden müssen oder gar verunmöglicht werden, indem die benötigten Funktions- oder Verdrahtungselemente absterben. Andererseits wurde die überraschende Entdeckung gemacht, dass Strudelwürmer [Planarien] gewisse einfache Verhaltungsweisen «erlernen» können und — was besonders erstaunlich ist —, falls solche «instruierten» Würmer zerhackt und nicht behandelten Würmern verfüttert werden, die «nicht instruierten» Würmer auf diesem kannibalischen Wege die fragliche Verhaltensweise «erlernen» können [3].)

Ausserdem verarbeitet das Gehirn, genau wie eine digitale Rechenmaschine, alle Informationen binär [2], d. h. in einen zweiwertigen Kode zerlegt.

Natürlich ist nun diese Analogie zwischen der Rechenmaschine und dem Gehirn (vor allem bezüglich der Unterteilung in Funktionselemente und Verdrahtung und der Verarbeitung binärer Informationen) in einem sehr vorsichtigen Sinne zu verstehen. Um einzusehen, wie gewaltig die Unterschiede zwischen Rechenmaschine und Gehirn sind, seien in einer Tabelle einige Zahlen angegeben.

Man ersieht aus dieser Tabelle, wie ungeheuer das Gehirn der Rechenmaschine

Tabelle

	Gehirn	Rechenmaschine
Funktionselemente pro cm^3	$\sim 10^7$ Neuronen	~ 0.01—0.1 Transistoren
Funktionselemente total	$\sim 10^{10}$ Neuronen	$\lesssim 10^4$—10^5 Transistoren
Energiedissipation in Watt	$\sim 10^{-9}$ pro Neuron	$\sim 10^{-1}$ pro Transistor
Speicherkapazität in «bits»	mind. 10^{20}	10^6—10^8
Operationen pro sec.	10^{12}—10^{16}*	$\lesssim 10^6$

* Ein Neuron hat eine Regenerationszeit von 10^{-4}—10^{-2} sec., im Gehirn sind aber vermutlich immer alle $\sim 10^{10}$ Neuronen mit durchschnittlich etwa 100 Synapsen gleichzeitig in Betrieb.

überlegen ist. Im folgenden soll nun gezeigt werden, wie die Analogie zwischen Maschine und Gehirn, auch abgesehen vom Quantitativen, nur eine äusserst dürftige Beschreibung des Gehirns zu liefern imstande ist.

§ 2 Naturwissenschaftliche Argumente für einen wesentlichen Unterschied zwischen Rechenmaschine und Gehirn

Eine Rechenmaschine ist natürlich so konstruiert, dass in ihr alle Vorgänge vollständig determiniert und eindeutig ablaufen. Ein Kernspeicherelement ist in der einen oder in der anderen Richtung magnetisiert, und nichts anderes ist möglich. Ein Transistor leitet oder leitet nicht; in einem Draht fliesst ein Strom oder er fliesst nicht: tertium non datur!

Eine Rechenmaschine besteht also aus makrophysikalischen Elementen und folglich laufen alle Vorgänge streng determiniert ab. Die Messungen in der Maschine, wie das Messen einer Magnetisierung (Ablesen eines Gedächtniselements) etc., müssen natürlich eindeutig und scharf sein.

Im Gehirn dagegen laufen die Vorgänge in einem äusserst komplizierten chemischen Milieu von Makromolekülen ab. Zur Beschreibung dieses Milieus ist die Quantenmechanik zuständig, wobei allerdings zu bemerken ist, dass die ungeheure Komplexität der Situation zum vornherein nur eine sehr beschränkte Beschreibung erwarten lässt. Ausserdem ist bis jetzt so gut wie nichts darüber bekannt, inwiefern typisch quantenmechanische Effekte (Quantensprünge, Indeterminiertheiten, Unschärfen usw.) für die Vorgänge im Gehirn charakteristisch sind; sicher ist nur, dass die letzteren von makrophysikalischen Vorgängen wesentlich verschieden sind.

Hierzu kommt noch ein zweiter sehr wesentlicher Punkt. Nach einer — sehr plausiblen — Hypothese von Bohr [4] sind die chemisch-physikalischen Vorgänge (im Sinne der Quantenmechanik verstanden!) in einem lebenden Organismus selbst noch komplementär zu den spezifisch biologischen Abläufen. Das bedeutet, dass eine *detaillierte* physikalisch-chemische Untersuchung, die wegen des mikroskopischen Charakters der Vorgänge auf molekularer Basis stattfinden müsste, Eingriffe in das biologische Objekt zur Folge hätte, die die Informationen über das eigentlich Biologische zerstören. Es ist also durchaus denkbar — und sogar wahrscheinlich —, dass die Tatsache, dass im Gehirn mikrophysikalische Vorgänge wesentlich sind, in zweierlei Hinsicht eine Rolle spielt: Einerseits ist eine präzise Untersuchung der Gehirnvorgänge durch Messung der physikalischen und chemischen Abläufe unmöglich ohne die Wirksamkeit gewisser biologischer Gesetze zu stören, von denen weiter unten noch die Rede sein wird. Daraus folgt zweitens, dass es sehr zweifelhaft ist, ob die Denkvorgänge im Gehirn dadurch genügend wiedergegeben werden können, dass mit maximaler Detailliertheit eine quantenmechanische Beschreibung der physikalischen Vorgänge im Gehirn gegeben wird. Diese Argumentation wird durch eine neue Arbeit über die Persistenz der Indeterminiertheiten in der Quantenmechanik gestützt [5].

In der Biologie treten mit grosser Wahrscheinlichkeit noch (bis jetzt unbekannte) Gesetze auf — biologische Gesetze —, die nicht auf physikalische Gesetze reduziert werden können [6]. Ein Indiz in dieser Richtung wird z. B. durch die Genetik geliefert

[7]. Dass die physikalischen (chemischen) Gesetze nämlich genügen, um aus den Vererbungsinformationen die Detailstruktur eines Lebewesens deterministisch zu bestimmen, scheint sehr unwahrscheinlich zu sein, indem z. B. die Zahl der dafür benötigten Informationen die Anzahl der in der Erbmasse vorhandenen um sehr viele Grössenordnungen übersteigt. Neue — bis jetzt unbekannte — Gesetze, die z. B. Ausdruck eines Gesamtplans für die Gestalt und Struktur des Lebewesens sind, müssen zusätzlich im Spiel sein [6].

Diese Ausführungen zeigen nun deutlich (wenn auch in äusserst knapper Darstellung; eine ausführlichere Diskussion findet sich in [6]), welch gewaltige Diskrepanz zwischen einer Rechenmaschine und dem Gehirn, oder allgemein, zwischen einem Lebewesen und einer Maschine besteht. Im folgenden soll nun gezeigt werden, dass auch vom formal logischen Standpunkt aus — der Standpunkt, der am meisten zum Vergleich Rechenmaschine—Gehirn verleitet — eine wesentliche Diskrepanz besteht.

§ 3. Logik der Rechenmaschine — Logik des Gehirns

Die Logik einer Rechenmaschine ist sehr einfach. Es ist die BOOLEsche Logik, die auf den drei Grundoperationen «und», «oder» und «nicht» aufgebaut ist [8]. Die Kombination «und» zweier Aussagen ist genau dann wahr, wenn die beiden Aussagen wahr sind, sonst ist sie falsch. Die Kombination «oder» ist genau dann falsch, wenn beide Aussagen falsch sind, sonst ist sie immer wahr. «Nicht» bewirkt, dass eine Aussage «a» dann wahr ist, wenn «nicht a» falsch ist und umgekehrt. Ein wesentlicher Zug dieser Logik ist also ihre Zweiwertigkeit: «wahr» oder «falsch» sind die zwei Werte. (Hier sei angemerkt, dass die Logik der Quantenmechanik nicht vom BOOLEschen Typus ist [5], [9], was ein weiteres Argument zum in § 2 Gesagten beiträgt.) Die Struktur der BOOLEschen Logik ist leicht überblickbar. Die Transistoren und Speicherelemente einer Rechenmaschine sind ebenfalls zweiwertige Elemente oder Kombinationen davon (§ 2). Die Verdrahtung dieser Elemente repräsentiert bestimmte (dem Zweck der Maschine angepasste) Kombinationen der drei Grundoperationen «und», «oder» und «nicht». Die topologische Struktur der Verdrahtung, bei gegebener Anordnung von Transistoren und Speicherelementen, ist ein Bild der zu Grunde gelegten Logik. Damit ist nun die Maschine imstande, geeignet programmierte Probleme mit grosser Geschwindigkeit und Zuverlässigkeit zu lösen. Dabei besteht das Programmieren letzten Endes gerade darin, das zu behandelnde Problem so darzustellen, dass es mit Hilfe der einfachen BOOLEschen Logik behandelt werden kann. Deshalb ist es auch notwendig, mathematische Probleme in viele Einzelschritte aufzulösen.

Ganz anders ist es um die Logik des Gehirns bestellt. Abgesehen davon, dass die Funktionselemente (Neuronen, Synapsen) wegen der möglichen quantenmechanischen Effekte nicht unbedingt die oben beschriebene Zweiwertigkeit aufweisen und abgesehen auch von der Unkenntnis der Funktionsweise des Gedächtnisses, repräsentiert die «Verdrahtung» der Funktionselemente eine Logik, die auf jeden Fall viel komplizierter ist als die BOOLEsche Logik. Wenn überhaupt das Gehirn einer Maschine vergleichend gegenübergestellt wird, so muss davon ausgegangen werden, dass das

Gehirn mindestens diejenige Struktur und «Verdrahtung» hat, die den Denkprozessen entsprechen, die tatsächlich ausgeführt werden. Dabei bleibt diese Untersuchung auf mathematische Denkprozesse beschränkt, und alle andern Gehirnfunktionen werden ausser acht gelassen. Erstere sind aber weit komplizierter als die in einer BOOLEschen Logik möglichen Denkprozesse; sie umfassen eben das ganze Gebiet der Mathematik. Dies wird im zweiten Kapitel ausführlich begründet. Vorher aber soll darauf hingewiesen werden, dass mit der Diskussion der der topologischen Struktur der «Verdrahtung» entsprechenden Logik des Gehirns allein nicht die ganze Diskrepanz zur Logik der Rechenmaschine zur Sprache kommt. Zusätzlich zu den obigen Argumenten, den Einfluss der Quantenmechanik und die Gedächtnismechanismen betreffend, ist es zum Beispiel noch denkbar, dass die Entscheidung darüber, ob ein Neuron in Funktion tritt («feuert») oder nicht, nicht nur davon abhängt, ob genügend Synapsen des Neurons angeregt sind. Es ist z. B. möglich, dass diese Entscheidung auch von der geometrischen Anordnung der angeregten Synapsen abhängt: dass also etwa 15 beliebige Synapsen kein Feuern bewirken, wohl aber 15 in einer speziellen geometrischen Anordnung [1]. Damit wäre also nicht nur die topologische Struktur der Verdrahtung, sondern auch die Lage der Synapsen ein Teil des Bildes für die Logik der Abkäufe im Gehirn. Noch weitere solche Möglichkeiten zu Komplikationen bestehen. (Die Schwierigkeit, solche Modelle maschinell zu simulieren, ist beträchtlich [10].)

Im weitern soll von diesen detaillierten Modellen abgesehen werden. Es soll, ausgehend von den metamathematischen Resultaten von GOEDEL und CHURCH, gezeigt werden, nicht nur, dass die (äusserst plausible) Behauptung, das Gehirn habe eine qualitativ andere und kompliziertere logische Struktur als jede denkbare Rechenmaschine, wahr ist, sondern es wird ausserdem gezeigt, dass dieser Unterschied — zusammen mit den bis hierher angeführten Argumenten — eine Analogiebetrachtung zwischen Rechenmaschine und Gehirn nur noch mit äusserst geringer Bedeutung zulässt; dass das Gehirn nicht auch nur annähernd durch eine Maschine simuliert werden kann.

2. Die metamathematischen Resultate von Goedel und Church und die Logik des Gehirns

§ 4. Die Theoreme von GOEDEL und CHURCH

Der Ausgangspunkt zum Beweis und Verständnis der beiden GOEDELschen Hauptresultate ist die Antinomie von RICHARD [11]. Es scheint nun aber günstiger, in diesem Paragraphen die Theoreme in möglichst einfacher Form darzustellen, um dann in § 5 an der Antinomie nachträglich zu demonstrieren, wie sie mittels einer sprachkritischen Untersuchung paradoxen spachlichen Formen entwachsen. Letzteres ist auch einer der Ausgangspunkte FINSLERS (neben den Antinomien der Mengenlehre) in den erwähnten Arbeiten [16].

Der Gegenstand der Theoreme von GOEDEL und CHURCH ist ein logisches System, nämlich das System der Zahlentheorie und Logik zusammen: System ZL. Dabei liegen der Zahlentheorie die PEANOschen Axiome[1] zu Grunde, wobei zusätzlich die in § 3 beschriebenen drei Grundoperationen der BOOLEschen Logik zum System hinzugenommen werden. (Dazu kommen noch einige Hilfsoperationen, wie Klammernbildung etc.) In ZL sind damit die Zahlentheorie und der bekannte Aussagen- und Prädikatenkalkül zusammengefasst. Wie bei jedem logischen System stellt sich nun die Frage, ob das zu Grunde gelegte Axiomensystem die folgenden Eigenschaften hat: Widerspruchsfreiheit, Vollständigkeit und Minimalität. Die dritte Eigenschaft, ob nämlich Axiome existieren, die von andern Axiomen abhängig und somit überflüssig sind, ist für die GOEDELschen Resultate nicht wesentlich. Die beiden andern Eigenschaften — Widerspruchsfreiheit und Vollständigkeit — sind der Gegenstand der Theoreme. Das erste GOEDELsche Theorem lautet nämlich folgendermassen:

Falls ZL formal widerspruchsfrei ist, so ist es formal unvollständig.

Die Bedeutung davon ist die folgende: Die Voraussetzung des Theorems ist die formale Widerspruchsfreiheit des Systems, d. h. es soll in ZL keine Aussage geben, die gleichzeitig mit ihrer Verneinung aus den Axiomen bewiesen werden kann. Diese Forderung ist aber wesentlich, um überhaupt ein sinnvolles System zu haben. Wäre sie nämlich nicht erfüllt, existierte also ein Satz, der gleichzeitig mit seiner Verneinung bewiesen werden kann, so zeigt man leicht (mittels modus ponens), dass überhaupt jeder Satz im System bewiesen werden kann, folglich auch die Verneinung jedes beliebigen Satzes, und folglich ist überhaupt jeder Satz widersprüchlich, falls nur einer es ist. Somit kann die Voraussetzung der formalen Widerspruchsfreiheit nicht fallen gelassen werden, will man nicht ein triviales (und unsinniges) System vor sich haben. Das Theorem besagt jetzt also, dass damit ZL formal unvollständig ist. Das bedeutet, dass in ZL mindestens ein Satz existiert, der aus den Axiomen des Systems nicht bewiesen werden kann; die Wahrheit (und auch Falschheit) des Satzes ist innerhalb des Systems unentscheidbar. Mit Hilfe der RICHARDschen Antinomie lässt sich ein solcher Satz explizite konstruieren (§ 5). Man denke sich nun einen solchen Satz der Zahlentheorie gefunden, von dem also gezeigt wurde, dass er unbeweisbar (nicht aus den Axiomen herleitbar) ist. Man ist versucht, folgendermassen zu argumentieren: Da natürlich auch die Negation des Satzes nicht bewiesen werden kann (sonst hätte man bewiesen, dass er falsch ist), ist es unmöglich, natürliche Zahlen zu finden, die den Satz explizit falsifizieren (es existiert kein Gegenbeispiel!). Folglich — so könnte man unvorsichtigerweise schliessen — ist der Satz wahr. Das ist aber nicht das, was hier unter Wahrheit oder Falschheit verstanden wird. Die Wahrheit oder Falschheit des Satzes ist wegen seiner bewiesenen Unbeweisbarkeit eben innerhalb des Systems auf keine Art und Weise verifizierbar. Es

---

[1] Die Axiome von PEANO lauten:

I 1 ist eine natürliche Zahl (nZ).

II Jede nZ a hat einen bestimmten Nachfolger a^+ in der Menge der nZ.

III Für alle nZ a ist $a^+ \neq 1$. (Es gibt keine nZ mit Nachfolger 1.)

IV Aus $a^+ = b^+$ folgt $a = b$.

V Jede Menge von nZ, die die 1 enthält, und zu jeder nZ a, die sie enthält, auch a^+ enthält, enthält alle nZ. (Prinzip der vollständigen Induktion.)

ist m. a. W. sinnlos, die Wahrheit eines Satzes in einem System zu behaupten, wenn der Satz im System als nicht beweisbar erkannt ist; seine Wahrheit (und seine Falschheit) ist unentscheidbar. (Diese Situation ist wesentlich für die Beweistheorie [12].) Ein, im Sinne dieser Theorie, beweisbarer Satz kann in endlich vielen Schritten aus den Axiomen gefolgert werden. Ein unbeweisbarer Satz kann dagegen auf keinem logischen Wege des Systems aus den Axiomen in endlich vielen Schritten erreicht werden. Das System zerfällt sozusagen in zwei getrennte Gebiete, wobei das eine aus dem andern nicht durch endliche logische Bahnen erreicht werden kann.

Ein weiteres Licht wirft das zweite GOEDELsche Theorem auf diese Situation:

Falls ZL formal widerspruchsfrei ist, so gibt es keinen Widerspruchsfreiheitsbeweis, der im System ZL selber erbracht werden kann.

Anders ausgedrückt ist der Satz «ZL ist widerspruchsfrei» in ZL unbeweisbar. Die Widerspruchsfreiheit muss also vorausgesetzt werden, kann aber in ZL nicht bewiesen werden. Die Situation ist dann so, dass durch Vergrösserung des Systems die Widerspruchsfreiheit von ZL im neuen System bewiesen werden kann. In der Analysis lässt sich dies erreichen; weiter kann in der Mengentheorie die Widerspruchsfreiheit der Analysis gezeigt werden etc. Es entsteht so eine Hierarchie von logischen Systemen, wobei das nächsthöhere System das darunter liegende vervollständigt, so dass z. B. auch der Satz von der Widerspruchsfreiheit eines bestimmten Systems in einem höheren System bewiesen werden kann. Da wir aber nur eine endliche Auswahl von Systemen haben können, so kommt dieses Prozedere nie zu einem Abschluss. Das höchste betrachtete System enthält immer unbeweisbare Sätze, und die Widerspruchsfreiheit kann in ihm nicht bewiesen werden.

Schliesslich ist für das Folgende noch das Theorem von CHURCH von grundlegender Bedeutung:

Falls ZL formal widerspruchsfrei ist, so gibt es keine effektive Methode, um festzustellen, ob eine Formel aus ZL in ZL beweisbar ist; das Entscheidungsproblem ist in ZL unlösbar.

Hier springt die Bedeutung des Theorems im Hinblick auf das Thema dieses Aufsatzes in die Augen: Das Theorem besagt, dass in ZL kein (formelunabhängiges, d. h. für jede beliebige Formel wirksames) Programm aufgestellt werden kann, das gestattet, für eine beliebige vorgegebene Formel zu entscheiden, ob sie beweisbar ist oder nicht. Es kann keine Maschine gebaut werden, die in ZL so programmiert werden kann, dass sie für jeden beliebigen Satz entscheidet, ob er beweisbar ist oder nicht.

Eine ausführlichere Diskussion der Bedeutung dieser Theoreme bezüglich des Vergleiches der Logik der Rechenmaschine und des Gehirns wird in § 6 gegeben. Vorher soll noch in § 5 anhand der Antinomie von RICHARD gezeigt werden, dass die besprochenen metamathematischen Resultate eng mit Eigenschaften der menschlichen Sprache zusammenhängen. Die Sprache der Rechenmaschine dagegen (die Sprache der BOOLEschen Logik), gibt zu keinen solchen Resultaten Anlass. Die BOOLEsche Logik ist trivialerweise vollständig und widerspruchsfrei; das Entscheidungsproblem ist trivial: jede Formel ist entweder beweisbar, oder ihre Verneinung ist beweisbar, und nicht beides.

§ 5. Die Antinomie von RICHARD

Die Antinomie kommt dadurch zustande, dass innerhalb der (deutschen) Sprache Aussagen über die natürlichen Zahlen gemacht werden. Es lässt sich — das wird sich zeigen — eine Eigenschaft der natürlichen Zahlen sprachlich so fixieren, dass die Eigenschaft für eine bestimmte Zahl zutrifft und gleichzeitig nicht zutrifft. Um die Bedeutung dieser Antinomie im rechten Licht zu sehen, muss sie mit anderen, «harmloseren» Antinomien oder Paradoxa verglichen werden. Die meisten dieser Paradoxa lassen sich leicht auflösen. Drei typische Beispiele sind die folgenden:

a) $x=1$ folglich $x^2-1=x-1$.
Durch Division durch $x-1$ ergibt sich $x+1=1$.
Aber, da $x=1$ folgt $\underline{2=1}$.
Offensichtlich liegt hier ganz einfach ein im System nicht regelgemässer Schluss vor, indem die Division durch $x-1=0$ nicht erlaubt ist.

b) In einem Dorf ist der Barbier derjenige Einwohner, der alle rasiert, die sich nicht selber rasieren. Fragt man nun, ob sich der Barbier selber rasiert oder nicht rasiert, so sieht man, dass beide Annahmen zu einem Widerspruch führen (nämlich jeweils zur andern Annahme). Die Auflösung des Paradoxons erfolgt durch reductio ad absurdum: Da (stillschweigend) vorausgesetzt wurde, dass im Dorf überhaupt ein solcher Barbier wohnt und da sich aus dieser Voraussetzung ein Widerspruch ergibt, muss die Voraussetzung fallen gelassen werden: Im Dorf wohnt kein Barbier, der der obigen Definition genügt.

c) Auf einer Karte steht: «Der Satz auf der Rückseite ist wahr.» Auf der Rückseite steht: «Der Satz auf der Rückseite ist falsch.» Die Frage, ob diese Sätze wahr oder falsch sind, führt offensichtlich zu Widersprüchen, die sich aber durch eine sorgfältige Verwendung der Begriffe «wahr» und «falsch» vermeiden lassen. Für Details sei auf [13] verwiesen.

Für die Antinomie von RICHARD versagen aber sämtliche Auflösungsmethoden. Dies sieht man folgendermassen ein: Eine zur Antinomie führende Zahleigenschaft ist: «Die kleinste Zahl, deren Name mindestens vierzig Silben enthält.» Diese Zahl existiert und ist durch Aufzählen eindeutig bestimmt. Die Eigenschaft «....» ist aber auch ein Name für eben diese Zahl, der weniger als vierzig Silben umfasst. Folglich ist die durch Aufzählen gefundene Zahl die durch die Auffindvorschrift geforderte (falls die Zahlwörter zur Benennung der Zahlen verwendet werden), und sie ist es gleichzeitig nicht, denn sie ist durch die Eigenschaft «....» selber durch weniger als vierzig Silben bestimmt. Man kann die Antinomie für dieses spezielle Beispiel durch eine sorgfältige Handhabung der Sprache auf zwei Arten auflösen:

a) Die Eigenschaft «....» werde abgeändert in:
«Die kleinste *durch ein Zahlwort bezeichnete* Zahl, deren Name mindestens vierzig Silben enthält.»

Dabei ist mit «Zahlwort» der grammatikalische Term gemeint, somit ist «....» kein zulässiger Name der Zahl mehr; die Antinomie ist aufgelöst.

b) Die abgeänderte Zahleigenschaft sei: «Die kleinste Zahl, deren Name mindestens zehn Silben enthält.» Nun hat der Name «....» mehr als 10 Silben; die Antinomie verschwindet.

Die Methode der Auflösung besteht in beiden Fällen offensichtlich darin, die kritische Zahleigenschaft durch eine andere zu ersetzen, die nicht mehr zur Antinomie führt. Nun zeigt sich aber, dass die RICHARDsche Antinomie gleichwohl konstruiert werden kann, sobald alle (abzählbar vielen) Zahlbezeichnungen beliebiger Sprachen betrachtet werden. Es versagen dann beide Auflösungsmethoden. Es lässt sich dann beweisen, dass in der Menge aller Zahlbezeichnungen sicher eine existiert, die zur Antinomie führt. Zusätzliche Spezifizierung einer gegebenen Zahleigenschaft, oder eine Veränderung ihres sprachlichen Ausdrucks, ändert also an dieser allgemeinen Sachlage nichts. Die Antinomie ist konstruierbar, ohne auf irgend eine spezielle Zahleigenschaft Bezug zu nehmen; sie kann also durch vorsichtige Handhabung einer solchen speziellen Zahleigenschaft nicht aufgelöst werden; sie kann nur auf eine andere Zahleigenschaft abgewälzt werden. Es ist folglich auch nicht möglich, alle Zahleigenschaften, die zur Antinomie führen, aus der Menge der zulässigen Zahlbezeichnungen zu eliminieren. Das würde zum absurden Resultat führen, dass überhaupt keine sinnvollen Zahlbezeichnungen existieren können. Eine präzise Darstellung dieses im Detail komplizierten Sachverhaltes findet sich in [11]. (Als Randbemerkung sei darauf hinzuweisen, dass aus diesen Überlegungen deutlich wird, dass es unmöglich ist, alle Zahlen im Rahmen eines endlichen sprachlichen Alphabets mit endlicher Wortlänge zu benennen, da offensichtlich die Anzahl der in einem solchen Alphabet möglichen Zahlbenennungen nur endlich ist.)

GOEDELS Idee bestand nun darin, die Aussagen über die Zahlentheorie und BOOLEsche Logik (Metasprache[2] bezüglich ZL) zu einem Bestandteil von ZL selber zu machen. Dies wird erreicht durch die sogenannte Arithmetisierung der Metasprache (auch GOEDELisierung), indem jedem metasprachlichen Ausdruck mittels eines eindeutigen Verfahrens eine natürliche Zahl — die GOEDELzahl — zugeordnet wird [14]. Mit dem gleichen Verfahren wird jedem Element von ZL (Formel, Theorem, Regel, Axiom etc.) eine GOEDELzahl zugeordnet. Somit entspricht jeder metasprachlichen Aussage über ZL eindeutig ein Element von ZL selber. Umgekehrt entsprechen gewissen Elementen von ZL eindeutig Metaaussagen über ZL. (Natürlich wird im allgemeinen einer GOEDELzahl nicht immer eine sinnvolle Metaaussage entsprechen.) Das Verfahren GOEDELS, das zur Konstruktion eines unbeweisbaren Satzes führt, besteht nun darin, dass ein zahlentheoretischer Satz konstruiert wird, der in der «metasprachlichen Übersetzung» seine eigene Unbeweisbarkeit behauptet. Diese metasprachliche Aussage geht durch Präzisierung der Begriffe nun gerade aus der RICHARDschen Antinomie hervor. Aus dieser Sachlage folgt, dass der so konstruierte Satz wirklich unbeweisbar ist, wenn das System als widerspruchsfrei voraus-

[2] Die Metasprache bezüglich eines Systems wird durch die Aussagen über die Sprache des Systems gebildet; also im vorliegenden Fall durch Aussagen über die Formeln aus ZL.

gesetzt ist, denn sonst könnten ja in ZL inhaltlich falsche Sätze bewiesen werden oder inhaltlich richtige Sätze falsifiziert werden, oder es wäre möglich, einen Satz und seine Verneinung gleichzeitig zu beweisen. In § 4 wurde aber gezeigt, dass Widerspruchsfreiheit vorausgesetzt werden muss.

Damit ist es gelungen, explizite einen unbeweisbaren Satz in ZL zu konstruieren. Wesentlich ist aus dieser Konstruktion ersichtlich, wie die Existenz solcher Sätze mit der menschlichen Sprache verknüpft ist. Dies wirft ein grelles Licht auf den gewaltigen Unterschied zwischen der Logik einer Rechenmaschine und derjenigen des Gehirns. In § 6 sollen die Folgerungen gezogen werden.

§ 6. Logik der Rechenmaschine — Logik des Gehirns: Schlussfolge

Abgesehen von den Argumenten, die in den Paragraphen 2 und 3 angedeutet worden sind, hat sich ergeben, dass die Verdrahtung des Gehirns eine viel kompliziertere Logik repräsentiert als die einfache BOOLEsche Logik einer heutigen elektronischen Rechenmaschine. Dabei wird angenommen, dass die logische Struktur des Gehirns mindestens so kompliziert ist wie das Denken, das es leistet. Andernfalls würden Denkprozesse möglich sein, für die keine genügende materielle Struktur im Gehirn vorhanden ist. Unter diesen Umständen würde es überhaupt nicht mehr sinnvoll sein, von einer Analogie zwischen der Maschine und dem Gehirn zu sprechen. (Ob diese Situation vorliegt oder nicht, kann bis jetzt keineswegs entschieden werden.) Es soll jetzt also angenommen werden, dass das Gehirn im Hinblick auf die mathematischen Denkprozesse, auf deren Betrachtung sich die Untersuchung beschränkt, mindestens eine Struktur besitzt, die so kompliziert ist, wie die in § 4 erwähnte mathematische Hierarchie. Es gelten also für diese Gehirnprozesse die Resultate von GOEDEL und CHURCH.

Ein typischer Zug von Systemen, für die diese metamathematischen Resultate gelten, ist nun die Eigenschaft, innerhalb des Systems Aussagen über das System zuzulassen. Im mathematischen Denken tritt dieser Zug offensichtlich auf; in der vollständigen und widerspruchsfreien BOOLEschen Logik einer heutigen Maschine dagegen nicht, da eben in diesem System die erwähnten Theoreme nicht gelten. Prinzipiell kann aber von vornherein nicht ausgeschlossen werden, dass einmal eine Maschine gebaut werden kann, die z. B. nach ZL verdrahtet ist und folglich eine solche Rückkoppelung von Metasprache und Objektsprache zulässt. Bezüglich einer solchen hypothetischen Maschine ist nun aber zweierlei zu beachten. Erstens ist eine im § 4 nicht erwähnte Voraussetzung für die Gültigkeit der Theoreme von GOEDEL und CHURCH die Forderung, dass das betrachtete System einen genügenden Umfang hat. Man kann zeigen, dass das System ZL diese Voraussetzung erfüllt. (Dass diese Voraussetzung im Gehirn erfüllt ist, ist kaum zu bezweifeln, denn falls von den 10^{10} Neuronen nur 10^6 paarweise verknüpft sind, so existieren bereits $10^{2\,783\,000}$ solche Verbindungen [2]. In [2] wird auch gezeigt, dass diese Zahl bestimmt die Komplexität der Struktur nur zu einem kleinen Bruchteil repräsentiert.) Ob es gelingt, eine genügend grosse Maschine zu bauen, ist nicht sicher, indem der benötigte Umfang leicht astronomische Grössenordnungen erreicht.

Zweitens muss beachtet werden, dass eine hypothetische, nach ZL verdrahtete

Maschine niemals die Korrelation zwischen Metasprache (Menge der Aussagen über Theoreme, Axiome etc. aus ZL) und Objektsprache (Menge der Theoreme etc. aus ZL) hinreichend herstellen kann. Sie kann bestenfalls nach Speicherung eines entsprechenden Kataloges zu jeder GOEDELzahl das entsprechende zahlentheoretische oder aussagenlogische Theorem (Axiom etc.) und die zur gleichen GOEDELzahl gehörige Metaaussage auffinden. Sie ist aber erstens nicht imstande zu untersuchen, ob die aufgefundene Metaaussage sinnvoll ist, denn dazu müsste die Maschine zusätzliche Kriterien über die Grammatik der Sprache etc. enthalten[3]. Zweitens ist es natürlich nicht möglich, in der Maschine einen unendlich grossen Katalog zu speichern. Es gibt aber unendlich viele GOEDELzahlen. Deshalb wird sich die zu einer willkürlichen GOEDELzahl gehörende Metaaussage, falls sie überhaupt sinnvoll ist, im allgemeinen nicht zufällig auf eine Objektaussage aus ZL beziehen, die im Katalog, der der Maschine zur Verfügung steht, enthalten ist.

Das menschliche Gehirn ist in einer solchen Situation in der Lage, seinen «Katalog» selbständig zu erweitern, eine Leistung, zu der keine Maschine imstande ist.

Aus dem Satz von CHURCH lässt sich nun noch eine weit drastischere und völlig eindeutige Schlussfolge ziehen. Da es in ZL keine systematische Methode geben kann, um in ZL einen beliebig vorgegebenen Satz aus ZL zu entscheiden, ist es überhaupt unmöglich eine ZL-verdrahtete Maschine so zu konstruieren und zu programmieren, dass sie einen beliebigen Satz entscheidet. Eine Maschine ist eben per definitionem etwas, was nach einer «systematischen Methode» arbeitet. Das gilt auch für Maschinen, die statistische Elemente enthalten [10]. — Die Mathematiker haben natürlich mittels ihrer Ideen schon zahllose Sätze aus ZL bewiesen (und damit entschieden), mit Hilfe von mehr oder weniger intuitiv verfolgten Gedankengängen. Das Gehirn ist also selbst (und gerade) im eigentlichen Anwendungsgebiet elektronischer Rechenmaschinen zu Denkabläufen fähig, die prinzipiell von keiner noch so grossen und komplizierten Maschine geleistet werden können. Dazu kommt natürlich noch die triviale Tatsache, dass eine Maschine überhaupt nichts tun kann, ohne vorher programmiert worden zu sein. Das menschliche Gehirn dagegen schliesst seine «Programmierung» in sich ein.

Die Analogie Gehirn—Maschine beschränkt sich also auf die Denkprozesse der BOOLEschen Logik, in der keine GOEDEL-CHURCHschen Sätze gelten. Die Analogie bringt also nur einen verschwindend kleinen Teil der Vorgänge im Gehirn mit der Gesamtheit der Vorgänge in einer Maschine in Verbindung, auch bei Beschränkung auf das rein mathematische Denken. Das Gehirn ist auch in dieser Hinsicht niemals durch eine Maschine auch nur annähernd darstellbar. Formal logistische Argumente zeigen, dass diese Tatsache keine Funktion des technischen Fortschritts ist, sondern durch das Wesen der Gehirnfunktionen selber bedingt ist.

* * *

[3] Es ist denkbar, dass durch eine gewaltige Vergrösserung des Katalogs die Maschine befähigt wird, diese Fragen zu entscheiden. Es ist aber ohne weiteres ersichtlich, dass ein solcher Katalog beinahe alle sinnvollen Metaaussagen zum vornherein enthalten müsste, so dass durch Vergleich entschieden werden könnte, ob eine bestimmte gegebene Metaaussage sinnvoll ist. Die Leistung der Maschine, verglichen mit der des Gehirns, zeigt sich hier in besonders kümmerlichem Lichte.

Ich möchte Herrn Prof. Heitler herzlichst danken für die vielen Diskussionen und die nützlichen Ratschläge bei der Abfassung des Manuskripts. Mein Dank geht aber noch weiter: Herr Prof. Heitler hat es gestattet und Interesse dafür geweckt, dass an seinem Institut auch Fragen diskutiert und verarbeitet werden, die den Themenkreis der theoretischen Physik im heutigen engern Sinn überschreiten, die aber durchaus zur «Natural Philosophy» (engl. Bezeichnung für theor. Physik), d. h. der theoretischen Naturwissenschaft, gehören.

Mein weiterer Dank gebührt Dr. G. Rasche und Dr. N. Straumann vom Institut für die vielen Diskussionen und die wertvollen Bemerkungen zu diesem Aufsatz.

Herrn Dr. med. H. Isler schulde ich Dank für eine Diskussion über physiologische und historische Aspekte der Gehirnforschung.

Die Arbeit wurde dank der finanziellen Unterstützung des Schweizerischen Nationalfonds für wissenschaftliche Forschung ermöglicht.

Literatur

1. JOHN VON NEUMANN: The Computer and the Brain. (Yale University Press, New Haven, 1958.)
2. K. AKERT: Antrittsrede an der Universität Zürich vom 27. 10. 62, publ. in der «Neuen Zürcher Zeitung» vom 16. 12. 1962, Nr. 5045 (134), Blatt 5.
3. J. B. BEST: Sc. American, Febr. 1963, 54.
4. W. HEISENBERG: Physik und Philosophie, Kap. VI. (Ullsteinbuch 249, 1959.)
5. J. M. JAUCH and C. PIRON: Can Hidden Variables be Excluded in Quantum Mechanics? To be published in H.P.A., *36*, 827, 1963.
   JOHN VON NEUMANN: Mathematische Grundlagen der Quantenmechanik. (Dover Publ., New York, 1943.)
6. W. HEITLER: Der Mensch und die naturwissenschaftliche Erkenntnis, Kap. 4. (Viehweg und Sohn, Braunschweig, 2. Aufl., 1962.)
7. M. W. NIRENBERG: Sc. American, March 1963, 80.
   F. H. C. CRICK: Sc. American, Oct. 1962, 66.
8. C. I. LEWIS: A Survey of Symbolic Logic. (Dover Publ. Inc., New York, 1960.)
9. H. REICHENBACH: Philosophische Grundlagen der Quantenmechanik. (Birkhäuser, Basel, 1949.)
10. A. GAMBA et al.: N.C. Suppl. at Vol. XX, 112, 146, 221, 1961.
11. W. STEGMÜLLER: Unvollständigkeit und Unentscheidbarkeit. (Springer-Verlag, Wien, 1959.) Mit ausführlichem Literaturverzeichnis zum GOEDELschen Themenkreis.
12. K. SCHÜTTE: Beweistheorie. (Springer-Verlag, Göttingen, 1960.)
13. W. V. QUINE: Sc. American, April 1962, 84.
14. E. NAGEL and J. R. NEUMAN: Sc. American, June 1956, 71, mit einfachen Beispielen zur GOEDELisierung.
15. K. GOEDEL: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatsheft f. Mathematik und Physik *38*, 173, 1931.
16. P. FINSLER: Formale Beweise und die Entscheidbarkeit. Math. Zeitschr. *25*, 676, 1926.
    Gibt es Widersprüche in der Mathematik? Jahresbericht der Deutschen Math.-Ver. *34*, 143, 1925.

In den «Proceedings of the Symposium on Theory of Models» wird eine Arbeit von P. J. COHEN erscheinen — «The Independence of the Axiom of Choice» —, in der die Unbeweisbarkeit des Auswahlaxioms bewiesen wird. Es ergibt sich somit außer dem aus der RICHARDschen Antinomie etwas künstlich konstruierten Satz noch ein weiteres explizites Beispiel einer unbeweisbaren Aussage von eminenter mathematischer Bedeutung.