

RECHENMASCHINEN UND INTELLIGENZ

1. DAS IMITATIONSSPIEL

ICH schlage die Betrachtung der Frage vor: »Können Maschinen denken?« Sie sollte mit Definitionen der Begriffe »Maschine« und »denken« beginnen. Man könnte diese Definitionen so formulieren, daß sie so weit wie möglich den allgemeinen Sprachgebrauch wiedergeben, aber diese Einstellung ist gefährlich. Wenn die Bedeutung der Begriffe »Maschine« und »denken« herausgefunden werden soll, indem man untersucht, wie sie gemeinhin gebraucht werden, ist es schwer, dem Schluß zu entkommen, daß Sinn und Beantwortung der Frage »Können Maschinen denken?« durch eine statistische Untersuchung wie eine Gallup-Befragung ermittelt werden sollten. Dies ist jedoch absurd. Anstatt eine derartige Definition zu versuchen, werde ich die Frage durch eine andere ersetzen, die eng mit ihr verwandt und in verhältnismäßig unzweideutigen Begriffen ausgedrückt ist.

Die neue Form des Problems läßt sich in Terms eines Spiels beschreiben, das wir das »Imitationsspiel« nennen wollen. Es wird von drei Personen gespielt, einem Mann (A), einer Frau (B) und einem Fragesteller (C), der von beiderlei Geschlecht sein kann. Der Fragesteller sei allein in einem Raum. Das Ziel des Fragestellers ist es zu entscheiden, welche der beiden anderen Personen der Mann bzw. die Frau ist. Er kennt sie zunächst als X bzw. Y und am Ende des Spiels sagt er entweder »X ist A und Y ist B« oder »X ist B und Y ist A«. Der Fragesteller darf an A und B Fragen stellen wie:

C: Würde mir X bitte die Länge seines oder ihres Haars sagen? Angenommen X sei A, so muß A antworten. A's Ziel bei diesem Spiel besteht nun darin, C möglichst zur falschen Identifizierung zu veranlassen. Seine Antwort könnte demnach lauten:

»Mein Haar ist kurz geschnitten, und die längsten Strähnen sind ungefähr neun Zoll lang.«

Dabei sollten die jeweiligen Antworten schriftlich, am besten maschinengeschrieben, abgegeben werden, damit der Ton der Stimme dem Fragesteller nicht hilft. Ein Fernschreiber wäre das ideale Verständigungsmittel zwischen beiden Räumen. Andernfalls können Frage und Antworten durch eine Mittelsperson wiederholt werden. Das Ziel der dritten Spielerin (B) besteht darin, dem Fragesteller zu helfen. Die beste Strategie für sie ist wahrscheinlich, wahrheitsgetreu zu antworten. Sie kann ihren Antworten solche Bemerkungen hinzu-

fügen wie »Ich bin die Frau, höre nicht auf ihn«, aber das wird nichts nützen, da der Mann ähnliche Dinge sagen kann.

Wir stellen nun die Frage: »Was passiert, wenn eine Maschine die Rolle von A in diesem Spiel übernimmt?« Wird der Fragesteller sich in diesem Fall ebenso oft falsch entscheiden wie dann, wenn das Spiel von einem Mann und einer Frau gespielt wird? Diese Fragen treten an die Stelle unserer ursprünglichen: »Können Maschinen denken?«

2. KRITISCHE BEMERKUNGEN ZUM NEUEN PROBLEM

Ebenso gut, wie die Frage zu stellen: »Wie lautet die Antwort auf diese neue Problemstellung?«, kann man auch fragen; »Ist diese neue Frage eine Untersuchung wert?« Die letztere Frage untersuchen wir ohne weitere Umstände, wodurch wir einen regressus ad infinitum einfach unterbinden.

Das neue Problem besitzt den Vorteil, eine ziemlich scharfe Trennungslinie zwischen den physischen und den intellektuellen Fähigkeiten eines Menschen zu ziehen. Kein Ingenieur oder Chemiker wird von sich behaupten, er könne ein Material produzieren, das von der menschlichen Haut nicht zu unterscheiden ist. Möglich, daß dies irgendwann einmal gelingt, aber selbst wenn man annimmt, eine derartige Erfindung sei zu verwirklichen, hätte man doch das Gefühl, daß es wenig sinnvoll wäre, eine »denkende Maschine« dadurch menschlicher gestalten zu wollen, daß man sie mit solch künstlichem Fleisch umgibt. Die Form, auf die wir das Problem gebracht haben, trägt dieser Tatsache insofern Rechnung, als sie den Fragesteller daran hindert, die anderen Beteiligten zu sehen oder zu berühren bzw. ihre Stimmen zu hören. Einige andere Vorteile des vorgeschlagenen Kriteriums lassen sich an Hand von Musterfragen und Musterantworten aufzeigen, z. B.:

F: Schreiben Sie mir bitte ein Gedicht über die Firth of Forth-Brücke.

A: In diesem Punkt muß ich passen. Ich könnte nie ein Gedicht schreiben.

F: Addieren sie 34957 zu 70764.

A (wartet ca. 30 Sekunden und gibt dann zur Antwort): 105621.

F: Spielen Sie Schach?

A: Ja.

F: Mein König steht auf e8; sonst habe ich keine Figuren mehr. Sie haben nur noch Ihren König auf e6 und einen Turm auf h1. Sie sind am Zug. Wie ziehen Sie?

A(nach einer Pause von 15 Sekunden): Th1-h8, matt.

Die Frage-und-Antwortmethode scheint geeignet zu sein, fast jeden Bereich menschlichen Bemühens einzubringen, den wir einbeziehen möchten. Wir wollen weder die Maschine für ihre Unfähigkeit bestrafen, in einem Schönheitswettbewerb zu brillieren, noch einen Menschen bestrafen, wenn er im Wettlauf mit einem Flugzeug unterliegt. Diese Unvermögen werden im Hinblick auf unsere Spielbedingungen irrelevant. Die »Zeugen« mögen, wenn es ihnen ratsam erscheint, mit ihren Reizen, ihren Kräften oder ihrer Heldenhaftigkeit prahlen, soviel sie wollen, aber der Fragesteller kann keine praktischen Beweise verlangen.

Das Spiel könnte vielleicht im Hinblick darauf kritisiert werden, daß die Maschine zu sehr im Nachteil sei. Wenn der Mensch versuchen sollte, so zu tun, als wäre er die Maschine, würde er zweifellos einen armseligen Eindruck machen. Er würde sich wegen seiner Langsamkeit und seiner rechnerischen Ungenauigkeit sofort verraten. Können nicht Maschinen etwas ausführen, das als Denken bezeichnet werden müßte, das sich jedoch stark von dem unterscheidet, was ein Mensch tut? Dies ist ein schwerwiegender Einwand, aber wir können zumindest sagen, daß er uns nicht zu beunruhigen braucht, wenn sich trotzdem eine Maschine konstruieren läßt, die das Imitationsspiel befriedigend spielt.

Man könnte vorbringen, daß bei dem »Imitationsspiel« die für die Maschine beste Strategie möglicherweise etwas anderes ist als die Nachahmung menschlichen Verhaltens. Das mag sein, doch halte ich einen größeren Effekt dieser Art für unwahrscheinlich. Jedenfalls ist hier nicht beabsichtigt, die Theorie des Spiels zu untersuchen, und es wird vorausgesetzt, daß die beste Strategie in dem Versuch besteht, Antworten zu geben, die ein Mensch normalerweise geben würde.

3. DIE AM SPIEL BETEILIGTEN MASCHINEN

Die in Abschnitt 1 gestellte Frage wird erst dann klar definiert sein, wenn wir spezifiziert haben, was wir mit dem Wort »Maschine« meinen. Natürlich möchten wir, daß die Anwendung jeder Art maschinenbautechnischer Hilfsmittel erlaubt ist. Wir wollen auch die Möglichkeit zulassen, daß ein Ingenieur oder ein Ingenieurteam eine funktionstüchtige Maschine konstruiert, deren Operationsweise jedoch von ihren Konstrukteuren nicht befriedigend beschrieben werden

kann, weil sie eine Methode anwandten, die weitgehend experimentell ist. Schließlich wollen wir Menschen, die auf die übliche Weise zur Welt kamen, von den Maschinen ausschließen. Es ist schwierig, die Definitionen so zu formulieren, daß sie diese drei Bedingungen erfüllen. Man könnte z. B. darauf bestehen, daß das Ingenieurteam aus Personen einerlei Geschlechts besteht, doch dies wäre nicht wirklich befriedigend, da es wahrscheinlich möglich ist, aus einer einzigen Hautzelle, etwa eines Mannes, ein vollständiges Individuum zu züchten. Das zu tun wäre eine biotechnische Meisterleistung, die allergrößtes Lob verdiente, jedoch wären wir nicht geneigt, es als einen Fall der »Konstruktion einer denkenden Maschine« anzusehen. Dies zwingt uns, die Forderung fallenzulassen, daß jede Art von Technik zugelassen sein soll. Wir sind dazu um so mehr bereit angesichts der Tatsache, daß das gegenwärtige Interesse an »denkenden Maschinen« durch eine besondere Art von Maschinen geweckt wurde, die man gewöhnlich als »Elektronenrechner« oder »Digitalrechner« bezeichnet. Dieser Anregung folgend lassen wir nur Digitalrechner an unserem Spiel teilnehmen.

Diese Einschränkung erscheint auf den ersten Blick sehr drastisch. Ich versuche zu zeigen, daß dem in Wirklichkeit nicht so ist. Dies erfordert eine kurze Darlegung der Natur und der Eigenschaften dieser Rechner.

Es darf auch gesagt werden, daß diese Identifizierung von Maschinen mit Digitalrechnern ebenso wie unser Kriterium für »Denken«, nur dann unbefriedigend sein wird, wenn es sich herausstellt (was ich nicht glaube), daß Digitalrechner nicht fähig sind, sich bei dem Spiel zu bewähren.

Es gibt schon eine Anzahl Digitalrechner im Einsatz, und man könnte fragen: »Warum versucht man nicht, das Experiment einfach durchzuführen? Es würde leichtfallen, die Bedingungen des Spiels zu erfüllen. Man könnte eine Anzahl Fragesteller heranziehen und Statistiken zusammentragen, um aufzuzeigen, wie oft die richtige Identifizierung gelang.« Die kurze Antwort hierauf ist, daß wir weder danach fragen, ob alle Digitalrechner bei dem Spiel Erfolg haben, noch danach, ob die gegenwärtig verfügbaren Rechner Erfolg haben, sondern ob Rechner vorstellbar sind, die Erfolg hätten. Aber das ist nur die kurzgefaßte Antwort. Wir werden diese Frage später in anderem Licht sehen.

4. DIGITALRECHNER

Die Idee, die Digitalrechnern zugrunde liegt, kann erklärt werden, indem man sagt, daß diese Maschinen dazu bestimmt sind, jede Operation auszuführen, die von einem menschlichen Rechner durchgeführt werden könnte. Vom menschlichen Rechner nimmt man an, daß er festen Regeln folgt; er darf in keinem Detail davon abweichen. Wir mögen annehmen, daß diese Regeln in einem Buch festgehalten sind, welches jedesmal abgeändert wird, wenn er eine neue Aufgabe bekommt. Er besitzt außerdem einen unbeschränkten Papiervorrat, auf dem er seine Rechnungen ausführt. Seine Multiplikationen und Additionen darf er auch auf einer »Tischrechenmaschine« ausführen, das ist jedoch unwesentlich.

Wenn wir die obige Erklärung als Definition verwenden, laufen wir Gefahr, einen Zirkelschluß zu ziehen. Das vermeiden wir, indem wir die Mittel beschreiben, durch welche der gewünschte Effekt erreicht wird. Ein Digitalrechner kann gewöhnlich als aus drei Teilen bestehend angesehen werden:

- (i) Speicher,
- (ii) ausführende Einheit,
- (iii) Kontrollwerk.

Der Speicher ist ein Informationsspeicher und entspricht dem Papier des menschlichen Rechners, ob es das Papier ist, auf dem er rechnet, oder jenes, auf dem das Buch, in dem die Regeln festgehalten sind, gedruckt ist. Insoweit der menschliche Rechner einige Berechnungen im Kopf ausführt, wird ein Teil des Speichers seinem Gedächtnis entsprechen.

Die ausführende Einheit ist der Teil, der die verschiedenen bei einer Rechnung auftretenden individuellen Operationen ausführt. Worin diese individuellen Operationen bestehen, variiert von Maschine zu Maschine. Gewöhnlich sind ziemlich lange Operationen möglich, z. B. »Multipliziere 3 540 675 445 mit 7 076 345 687«, bei einigen Maschinen dagegen nur ganz einfache Operationen, z. B. »Notiere Null«.

Wir haben erwähnt, daß das »Regelbuch«, mit dem der Rechner ausgestattet ist, in der Maschine durch einen Teil des Speichers ersetzt wird. Man nennt es dann die »Befehlsliste«. Das Kontrollwerk hat

dafür zu sorgen, daß die Befehle korrekt und in der richtigen Reihenfolge ausgeführt werden. Das Kontrollwerk ist so konstruiert, daß dies notwendigerweise geschieht.

Die im Speicher enthaltene Information ist gewöhnlich in einigermaßen kleine Bündel aufgeteilt. In einer Maschine z. B. könnte ein Bündel aus zehn Dezimalziffern bestehen. Die Teile des Speichers, in denen die verschiedenen Informationsbündel gespeichert werden, sind nach einem bestimmten System durchnummeriert. Ein typischer Befehl könnte lauten:

»Addiere die in Platz 6809 gespeicherte Zahl zu der in 4302 gespeicherten und bringe das Ergebnis an den letzteren Speicherplatz zurück.«

Es ist wohl nicht nötig, darauf hinzuweisen, daß ein solcher Befehl in der Maschine nicht auf Englisch erscheinen würde. Er würde vielmehr in einer Form verschlüsselt, wie etwa 6809430217. Hierbei bestimmt 17 diejenige der verschiedenen möglichen Operationen, die über den beiden Zahlen ausgeführt werden soll. In diesem Fall ist die Operation jene oben genannte, nämlich »Addiere die Zahl ...« Wie man sieht, besteht der Befehl aus 10 Ziffern und bildet somit ein sehr praktisches Informationsbündel. Normalerweise wird das Kontrollwerk die zu befolgenden Befehle in der Reihenfolge der Positionen aufgreifen, in denen sie gespeichert sind, gelegentlich trifft man jedoch auch auf einen Befehl wie:

»Führe jetzt den in Platz 5606 gespeicherten Befehl aus, und mache von dort aus weiter«,

oder

»Falls Platz 4505 0 enthält, führe als nächsten Befehl den in 6707 stehenden aus, andernfalls fahre direkt fort.«

Befehle dieser letzteren Art sind sehr wichtig, da sie es ermöglichen, eine Folge von Operationen solange zu wiederholen, bis eine bestimmte Bedingung erfüllt ist, und zwar so, daß für jede Wiederholung keine neuen Befehle, sondern immer wieder dieselben beachtet werden müssen. Zur Illustration ein Vergleich mit dem täglichen Leben. Angenommen die Mutter wünscht, daß Thomas jeden Morgen auf seinem Schulweg beim Schuster nachfragt, ob ihre Schuhe fertig sind, so kann sie ihm dies jeden Morgen aufs neue sagen. Sie kann aber auch ein für allemal eine entsprechende Notiz in den Flur hängen, die ihn beim Verlassen des Hauses daran erinnert, nach den

Schuhen zu fragen, und ihn auffordert, die Notiz zu vernichten, wenn er die Schuhe bekommen hat.

Der Leser muß es als eine Tatsache hinnehmen, daß Digitalrechner konstruiert werden können und tatsächlich konstruiert wurden, die im Einklang mit den beschriebenen Prinzipien stehen, und daß sie wirklich in der Lage sind, die Tätigkeit eines menschlichen Rechners sehr genau nachzuahmen.

Das Regelbuch, dessen Gebrauch durch den menschlichen Rechner wir geschildert haben, ist natürlich eine zweckdienliche Fiktion. Wirklich menschliche Rechner haben im Kopf, was sie tun müssen. Will man eine Maschine dazu bringen, bei einer schwierigen Operation das Verhalten des menschlichen Rechners nachzuahmen, muß man ihn fragen, wie sie ausgeführt wird, und die Antwort dann in die Form einer Befehlsliste übertragen. Die Aufstellung von Befehlslisten bezeichnet man gewöhnlich als »programmieren«. »Eine Maschine für die Ausführung einer Operation A programmieren« heißt, der Maschine die geeignete Befehlsliste eingeben, daß sie A ausführen wird.

Eine interessante Variante der Idee des Digitalrechners ist ein »Digitalrechner mit einem Zufallselement«. Diese besitzen Befehle, die das Werfen eines Würfels oder einen äquivalenten elektronischen Vorgang einschließen; z. B. könnte ein solcher Befehl lauten: »Würfle und bringe die gewürfelte Zahl in die Speicherzelle 1000«. Manchmal beschreibt man eine solche Maschine dadurch, daß man sagt, sie habe einen freien Willen (obwohl ich selbst diesen Ausdruck nicht verwenden würde). Gewöhnlich ist es nicht möglich, durch Beobachtung einer Maschine festzustellen, ob sie ein Zufallselement besitzt, da ein ähnlicher Effekt durch solche Vorrichtungen erreicht werden kann, daß man die Wahlen von den Dezimalstellen von π abhängen läßt.

Die meisten wirklichen Digitalrechner besitzen nur einen endlichen Speicher. Theoretisch ist es nicht schwierig, sich einen Rechner mit einem unbegrenzten Speicher vorzustellen. Natürlich kann jedesmal nur ein endlicher Teil des Speichers benützt werden. Ebenso kann immer nur ein endlicher Umfang konstruiert sein, wir können uns aber je nach den Erfordernissen mehr und mehr hinzugefügt denken. Solche Rechenmaschinen sind von besonderem theoretischen Interesse und heißen Rechenmaschinen mit unbegrenzter Kapazität.

Die Idee eines Digitalrechners ist alt. Charles Babbage, Mathema-

tik-Professor in Cambridge von 1828 bis 1839, entwarf eine solche Maschine, »Analytische Maschine« genannt, die jedoch nie vollendet wurde. Obwohl Babbage alle wesentlichen Gedanken hatte, war seine Maschine zu jener Zeit keine sehr attraktive, zukunftsweisende Idee. Die erreichbare Geschwindigkeit hätte zwar die eines menschlichen Rechners bestimmt übertroffen, wäre jedoch ungefähr hundertmal unter der Geschwindigkeit der Manchester-Maschine geblieben, die ihrerseits eine der langsameren unter den modernen Maschinen ist. Der Speicher sollte, unter Verwendung von Zahnrädern und Lochkarten, rein mechanisch sein.

Die Tatsache, daß Babbages Analytische Maschine vollständig mechanisch sein sollte, wird uns von einem Aberglauben befreien. Oft mißt man der Tatsache Bedeutung bei, daß moderne Digitalrechner elektrisch sind und daß auch das Nervensystem elektrisch ist. Da Babbages Maschine nicht elektrisch war und da alle Digitalrechner in einem gewissen Sinn äquivalent sind, sehen wir, daß die Verwendung der Elektrizität nicht von theoretischer Bedeutung sein kann. Natürlich tritt Elektrizität immer dort auf, wo man Signale schnell übertragen will, so daß es nicht weiter überrascht, sie in diesen beiden Zusammenhängen anzutreffen. Im Nervensystem sind chemische Phänomene mindestens ebenso wichtig wie elektrische. In einigen Rechenmaschinen ist das Speichersystem in der Hauptsache akustisch. Die Verwendung von Elektrizität erscheint somit nur mehr als sehr oberflächliche Ähnlichkeit. Wenn wir Ähnlichkeiten finden wollen, sollten wir eher nach mathematischen Funktionsanalogien sehen.

5. UNIVERSALITÄT DER DIGITALRECHNER

Die im letzten Abschnitt betrachteten Digitalrechner lassen sich unter die »diskreten Maschinen« (discrete-state machines) einreihen. Diese sind die Maschinen, bei denen der Übergang von einem ganz bestimmten Zustand in einen anderen in plötzlichen Sprüngen erfolgt. Die Zustände sind dabei hinreichend verschieden, so daß man die Möglichkeit von Verwechslungen außer acht lassen kann. Streng genommen gibt es keine Maschinen dieser Art. In Wirklichkeit verläuft alles stetig. Es gibt jedoch viele Maschinenarten, die zweckmäßigerweise als diskrete Maschinen *gedacht* werden können. Betrachtet

man z. B. die Schalter einer Beleuchtungsanlage, dann ist es eine passende Fiktion, daß jeder Schalter eindeutig ein- oder eindeutig ausgeschaltet sein muß. Wohl muß es Zwischenstellungen geben, die man jedoch in den meisten Fällen vernachlässigen kann. Als Beispiel einer diskreten Maschine könnten wir ein Rad betrachten, das sich in Abständen von einer Sekunde jeweils um 120° ruckartig weiterdreht und durch einen Hebel angehalten werden kann, der sich von außen bedienen läßt; außerdem soll bei einer bestimmten Radstellung eine Lampe aufleuchten. Diese Maschine ließe sich wie folgt abstrakt beschreiben. Der interne Zustand der Maschine (durch die Radstellung beschrieben) kann q_1 , q_2 oder q_3 sein. Es gibt ein Eingabesignal i_0 oder i_1 (Hebelstellung). Der interne Zustand in einem beliebigen Augenblick wird durch den letzten Zustand und das Eingabesignal gemäß folgender Tabelle bestimmt:

		Letzter Zustand		
		q_1	q_2	q_3
Eingabe	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

Die Ausgabesignale, die einzige äußerlich sichtbare Anzeige des internen Zustands (das Licht), werden durch die folgende Tabelle beschrieben

Zustand	q_1	q_2	q_3
Ausgabe	o_0	o_0	o_1

Dieses Beispiel ist typisch für diskrete Maschinen. Sie lassen sich durch solche Tabellen beschreiben, vorausgesetzt, daß sie nur endlich viele Zustände annehmen können.

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 Systeme, 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.

Wie erwähnt, gehören Digitalrechner zur Klasse der diskreten Maschinen. Die Anzahl der verschiedenen Zustände, die eine solche Maschine annehmen kann, ist jedoch gewöhnlich ungeheuer groß. Zum Beispiel beträgt sie bei der zur Zeit in Manchester arbeitenden Maschine ungefähr $2^{165\,000}$, d. h. ungefähr $10^{50\,000}$. Man vergleiche dies mit unserem Beispiel des sich ruckartig bewegendes Rades, welches drei Zustände annehmen konnte. Es ist nicht schwer einzusehen, weshalb die Anzahl der Zustände so groß sein muß. Die Rechenmaschine besitzt einen Speicher, welcher dem vom menschlichen Rechner verwendeten Papier entspricht. So muß es möglich sein, in den Speicher jede beliebige Zeichenkombination zu schreiben, die man auch auf Papier hätte schreiben können. Der Einfachheit halber nehme man an, daß als Zeichen nur die Ziffern von 0 bis 9 verwendet werden. Handschriftliche Variationen werden ignoriert. Angenommen, dem Rechner stünden 100 Blatt Papier zur Verfügung, von denen jedes 50 Zeilen für jeweils 30 Ziffern enthält. Dann ist die Anzahl der Zustände gleich $10^{100 \times 50 \times 30}$, d. h. gleich $10^{150\,000}$. Dies ist ungefähr die Anzahl der Zustände von drei Manchester-Maschinen zusammengekommen. Den Logarithmus zur Basis 2 der Anzahl der Zustände nennt man gewöhnlich die »Speicherkapazität« der Maschine. Demzufolge besitzt die Manchester-Maschine eine Speicherkapazität von ungefähr 165 000 und die Radmaschine unseres Beispiels eine solche von ca. 1,6. Setzt man zwei Maschinen zusammen, so muß man deren Kapazitäten miteinander addieren, um die Kapazität der daraus resultierenden Maschine zu erhalten. Dies führt auf die Möglich-

keit solcher Feststellungen wie: »Die Manchester-Maschine enthält 64 Magnetbänder, deren jedes eine Kapazität von 2560 besitzt, acht Elektronenröhren mit einer Kapazität von 1280. Die Kapazität gemischter Speicher beläuft sich auf ca. 300, so daß sich eine Gesamtkapazität von 174 380 ergibt.«

Kennt man die zu einer diskreten Maschine gehörige Tabelle, so ist es möglich, vorherzusagen, was sie tun wird. Es gibt keinen Grund, weshalb diese Vorhersage nicht mit Hilfe eines Digitalrechners ausgeführt werden sollte. Vorausgesetzt, daß sie hinreichend schnell ausgeführt werden könnte, wäre der Digitalrechner in der Lage, das Verhalten jeder diskreten Maschine nachzuahmen. Das Imitationsspiel ließe sich dann mit der betreffenden Maschine (als B) und dem sie nachahmenden Digitalrechner (als A) spielen, und der Fragesteller wäre außerstande, sie zu unterscheiden. Natürlich muß der Digitalrechner eine ausreichende Speicherkapazität haben und auch schnell genug arbeiten. Darüber hinaus muß er für jede Maschine, die er nachahmen soll, neu programmiert werden.

Diese spezielle Eigenschaft von Digitalrechnern, daß sie jede beliebige diskrete Maschine nachahmen können, beschreibt man dadurch, daß man sagt, sie seien *universale* Maschinen. Aus der Existenz von Maschinen mit dieser Eigenschaft ergibt sich als wichtige Folgerung, daß es bei Vernachlässigung von Geschwindigkeitsbetrachtungen nicht erforderlich ist, für verschiedene Rechenprozesse jeweils verschiedene Maschinen zu entwerfen. Sie alle können von ein und demselben, für den jeweiligen Fall programmierten Digitalrechner ausgeführt werden. Wir werden sehen, daß sich hieraus in gewissem Sinn die Gleichwertigkeit aller Digitalrechner ergibt.

Wir können nun nochmals an die am Ende von Abschnitt 3 vorgebrachte Überlegung anknüpfen. Es wurde versuchsweise angeregt, die Frage »Können Maschinen denken?« durch die Frage »Sind Digitalrechner vorstellbar, welche sich beim Imitationsspiel bewähren?« zu ersetzen. Wenn wir wollen, können wir dies oberflächlich verallgemeinern und fragen: »Gibt es diskrete Maschinen, die sich in diesem Sinne bewähren?« Im Hinblick auf die Universalitätseigenschaft sehen wir jedoch, daß beide Fragen mit der folgenden äquivalent sind: »Wir wollen unsere Aufmerksamkeit auf einen bestimmten Digitalrechner C heften. Ist es wahr, daß dieser Rechner durch geeignete Modifizierung seines Speichervermögens und seiner Aktionsge-

schwindigkeit sowie nach angemessener Programmierung in die Lage versetzt werden kann, die Rolle von A im Imitationsspiel befriedigend zu spielen, während B's Rolle von einem Menschen übernommen wird?»

6. GEGENSÄTZLICHE ANSICHTEN ÜBER DIE ZENTRALE FRAGE

Wir können nun das Fundament als geklärt betrachten und sind bereit, zur Diskussion unserer Frage »Können Maschinen denken?« und ihrer am Ende des letzten Abschnitts angeführten Variante überzugehen. Die ursprüngliche Form des Problems können wir nicht gänzlich fallen lassen, denn die Meinungen darüber, ob diese Ersetzung angemessen ist, gehen auseinander, und wir müssen uns zumindest das, was in diesem Zusammenhang vorzubringen ist, anhören.

Es wird die Sache für den Leser vereinfachen, wenn ich zunächst meine eigenen Ansichten hierzu darlege. Wir wollen zuerst die genauere Form der Frage ins Auge fassen. Meiner Meinung nach wird es in ca. 50 Jahren möglich sein, Rechenmaschinen mit einer Speicherkapazität von ungefähr 10^9 zu programmieren, die das Imitationsspiel so vollendet spielen, daß die Chancen für einen durchschnittlichen Fragesteller, nach einer fünfminütigen Fragezeit die richtige Identifizierung herauszufinden, nicht höher als sieben zu zehn stehen. Die ursprüngliche Frage »Können Maschinen denken?« halte ich für zu bedeutungslos, als daß sie ernsthaft diskutiert werden sollte. Nichtsdestoweniger glaube ich, daß am Ende unseres Jahrhunderts der Sprachgebrauch und die allgemeine gebildete Meinung sich so stark gewandelt haben werden, daß man widerspruchsfrei von denkenden Maschinen reden kann, ohne mit Widerspruch rechnen zu müssen. Auch erachte ich es für zwecklos, diese Ansichten zu verheimlichen. Die landläufige Meinung, daß Wissenschaftler unerbittlich von einer wohlbegründeten Tatsache zur anderen fortschreiten, ohne sich durch unbewiesene Vermutungen beeinflussen zu lassen, ist ganz und gar irrig. Vorausgesetzt, es ist klargelegt, was bewiesene Tatsachen und was Vermutungen sind, entsteht hieraus kein Schaden. Vermutungen sind sehr wichtig, da sie für die Forschung richtungsweisend sein können.

Ich möchte nun Meinungen betrachten, die meiner eigenen entgegenstehen.

(1) *Der theologische Einwand.* Denken ist eine Funktion der unsterblichen Seele des Menschen. Gott gab jedem Mann und jeder Frau eine unsterbliche Seele, aber keinen anderen Lebewesen oder Maschinen. Somit kann weder Tier noch Maschine denken. Ich sehe mich außerstande, auch nur einen Teil hiervon anzuerkennen, will jedoch versuchen, in theologischen Terms zu antworten. Ich hielte das Argument für überzeugender, würde man Menschen und Tiere in eine Klasse zusammenfassen, da meiner Meinung nach zwischen dem typisch lebendigen und dem leblosen Wesen ein größerer Unterschied besteht als zwischen Menschen und den übrigen Lebewesen. Der willkürliche Charakter der orthodoxen Ansicht tritt klarer hervor, wenn wir betrachten, wie sie dem Anhänger einer anderen religiösen Gemeinschaft erscheinen mag. Wie stehen Christen zu der Ansicht der Moslems, daß Frauen keine Seele haben? Wir wollen diesen Punkt jedoch beiseite lassen und zu dem Hauptargument zurückkehren. Mir scheint, daß das oben erwähnte Argument eine ernsthafte Einschränkung der Allmacht des Allmächtigen einschließt.¹ Zugegeben, es gibt gewisse Dinge, die Er nicht tun kann, z. B. eins gleich zwei machen, aber sollten wir nicht glauben, daß Er die Freiheit besitzt, Elefanten eine Seele zu verleihen, wenn Er es für passend hielte? Wir könnten erwarten, daß Er diese Macht nur im Zusammenhang mit einer Mutation ausübt, durch die der Elefant ein entsprechend verbessertes Gehirn erhält, das den Bedürfnissen dieser Seele gewachsen ist. Ein Argument genau der gleichen Form kann man für Maschinen ins Feld führen. Es mag uns anders vorkommen, da es schwieriger zu »schlucken« ist. Eigentlich bedeutet das jedoch nur, daß wir es für weniger wahrscheinlich halten, daß Ihm diese Umstände für die Verleihung einer Seele geeignet erscheinen könnten. Die fraglichen Umstände werden im folgenden behandelt werden. Beim Versuch, solche Maschinen zu konstruieren, sollten wir uns

¹ Diese Ansicht ist vielleicht häretisch. Thomas von Aquin (*Summa Theologica*, zitiert von Bertrand Russell in *History of Western Philosophy*, London 1940, S. 480) stellt fest, daß Gott keinen Menschen ohne Seele erschaffen kann. Aber das ist vielleicht keine Einschränkung der Macht Gottes, sondern einfach eine Folge der Unsterblichkeit und damit der Unzerstörbarkeit der menschlichen Seele.

nicht ehrfurchtslos Seine Macht anmaßen, Seelen zu erschaffen, ebenso wenig wie bei der Erzeugung von Kindern: eher sind wir, in beiden Fällen, Instrumente seines Willens und stellen für die von ihm erschaffenen Seelen eine Heimstatt bereit.

Dies ist jedoch reine Spekulation. Von theologischen Argumenten bin ich nicht sehr beeindruckt, wozu sie auch als Stütze dienen mögen. Solche Argumente haben sich in der Vergangenheit oft als unbefriedigend herausgestellt. Zu Galileis Zeiten argumentierte man, daß die Bibelstellen »Da stand die Sonne ... still ... und verzog unterzugehen beinahe einen ganzen Tag« (Josua 10, 13) und »der du das Erdreich gegründet hast auf seinen Boden, daß es bleibt immer und ewiglich« (Psalm 104,5) zur Widerlegung der Kopernikanischen Theorie ausreichen. Bei unserem gegenwärtigen Wissen erscheint ein solches Argument nichtig. Als dieses Wissen noch nicht vorhanden war, hinterließ es einen ganz anderen Eindruck.

(2) *Der »Vogel-Strauß«-Einwand.* »Die Konsequenzen, die sich aus denkenden Maschinen ergäben, wären zu schrecklich. Wir wollen hoffen und glauben, daß sie es nicht tun.«

Dieses Argument wird selten so offen ausgesprochen wie in der obigen Form. Aber es bewegt die meisten von uns, die überhaupt über das Problem nachdenken. Wir möchten gerne glauben, daß der Mensch in irgendeiner subtilen Weise der übrigen Schöpfung überlegen sei. Am besten wäre es, man könnte zeigen, daß er notwendigerweise überlegen ist, da für ihn dann keine Gefahr mehr besteht, seine Führungsrolle zu verlieren. Die Beliebtheit des theologischen Arguments steht eindeutig mit diesem Empfinden in Zusammenhang. Wahrscheinlich ist es bei Intellektuellen sehr stark ausgeprägt, da sie die Kraft des Denkens weit höher bewerten als andere und mehr geneigt sind, ihren Glauben an die Überlegenheit des Menschen auf diese Kraft zu gründen.

Ich denke nicht, daß dieses Argument so substantiell ist, daß es der Widerlegung bedürfte. Trost wäre hier besser am Platze: Vielleicht sollte man ihn in der Seelenwanderung suchen.

(3) *Der mathematische Einwand.* Es gibt eine Reihe von Resultaten der mathematischen Logik, die man benützen kann, um zu zeigen, daß die Fähigkeiten der diskreten Maschinen Beschränkungen unterliegen. Das bekannteste unter diesen Resultaten kennt man als das

Gödelsche Theorem², wonach in jedem genügend reichhaltigen logischen System Sätze formuliert werden können, die man innerhalb des Systems weder beweisen noch widerlegen kann, es sei denn, das System ist selbst möglicherweise inkonsistent. Es gibt andere, in gewisser Hinsicht ähnliche Resultate von Church, Kleene, Rosser und Turing³. Das letztere ist das für die Betrachtung geeignetste, da es sich direkt auf Maschinen bezieht, wohingegen sich die anderen nur für ein vergleichsweise indirektes Argument benützen lassen: wenn z. B. das Gödelsche Theorem angewendet werden soll, benötigen wir zusätzliche Hilfsmittel zur Beschreibung logischer Systeme durch Maschinen und umgekehrt von Maschinen durch logische Systeme. Das betreffende Resultat bezieht sich auf einen Maschinentyp, der im wesentlichen ein Digitalrechner mit unendlichem Speicher ist. Es besagt, daß es gewisse Dinge gibt, die diese Maschine nicht tun kann. Wenn sie so ausgerüstet wird, daß sie, wie im Imitationsspiel, Fragen beantworten kann, wird es einige Fragen geben, auf die sie entweder eine falsche oder überhaupt keine Antwort gibt, wieviel Zeit man ihr auch immer läßt. Es mag freilich viele solche Fragen geben, und Fragen, die von der einen Maschine nicht beantwortet werden können, mag eine andere zufriedenstellend beantworten. Vorläufig nehmen wir natürlich an, daß die Fragen solcher Art sind, auf die die Antwort »Ja« oder »Nein« paßt, und nicht Fragen wie: »Was halten Sie von Picasso?« Eine Frage, bei der die Maschine versagt, ist z. B.: »Betrachte die wie folgt spezifizierte Maschine ... Wird diese Maschine je auf eine Frage mit »Ja« antworten?« Die Punkte müssen durch die Beschreibung einer bestimmten Maschine in standardisierter Form ersetzt werden, die etwa von der in Abschnitt 5 benutzten Art sein könnte. Steht die beschriebene Maschine in einer gewissen, verhältnismäßig einfachen Beziehung zu der befragten Maschine, so läßt sich zeigen, daß die Antwort entweder falsch ist oder gar nicht gege-

2 K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, *Monatshefte für Math. und Phys.* (1931), 173–189.

3 Alonzo Church, An Unsolvability Problem of Elementary Number Theory, *American J. of Math.*, 58 (1936), 345–363.

S.C. Kleene, General Recursive Functions of Natural Numbers, *American J. of Math.*, 57 (1935), 153–173 und 219–244.

(A.M. Turing, On Computable Numbers, with an Application to the Entscheidungsproblem. In diesem Band S. 17–60.

ben wird. Dies ist das mathematische Ergebnis: es wird argumentiert, daß es eine Unfähigkeit der Maschinen beweist, der der menschliche Verstand nicht unterworfen ist.

Die kurze Antwort hierauf lautet, daß zwar nachgewiesen ist, daß jeder speziellen Maschine in ihren Fähigkeiten Grenzen gesetzt sind, hingegen nur behauptet wurde, und zwar ohne jeden Beweis, daß keine derartigen Einschränkungen für den menschlichen Verstand gelten. Aber ich glaube nicht, daß man diese Ansicht so leicht abtun kann. Wann immer einer dieser Maschinen die entsprechende kritische Frage gestellt wird und sie eine bestimmte Antwort gibt, wissen wir, daß diese Antwort falsch sein muß, und das gibt uns ein gewisses Gefühl der Überlegenheit. Ist dieses Gefühl illusorisch? Es ist zweifellos ganz echt, doch sollte ihm meiner Meinung nach keine zu große Bedeutung beigemessen werden. Wir selbst geben zu oft falsche Antworten auf Fragen, als daß die augenfällige Fehlbarkeit auf Seiten der Maschinen ein angenehmes Gefühl auf unserer Seite rechtfertigen würde. Außerdem können wir diese Überlegenheit nur in einer solchen Gelegenheit gegenüber der einen Maschine, über die wir unseren winzigen Triumph errungen haben, empfinden. Es kommt gar nicht in Frage, gleichzeitig über *alle* Maschinen zu triumphieren. Kurz gesagt: Es mag Menschen geben, die klüger sind als irgendeine gegebene Maschine, aber dann könnte es wiederum noch klügere Maschinen geben, usw.

Diejenigen, die an der mathematischen Argumentation festhalten, sind – denke ich – wohl am ehesten bereit, das Imitationsspiel als Diskussionsgrundlage zu akzeptieren. Diejenigen hingegen, die an die beiden vorhergehenden Einwände glauben, sind wahrscheinlich an keinerlei Kriterien interessiert.

(4) *Das Bewußtseinsargument*. Dieses Argument kommt sehr gut in Professor Jeffersons *Lister Oration for 1949*⁴ zum Ausdruck, aus der ich zitiere: »Erst wenn eine Maschine ein Sonett schreiben oder ein Konzert komponieren kann, und zwar aufgrund empfundener Gedanken und Gefühle und nicht als Zufallsverteilung von Symbolen, könnten wir zustimmen, daß die Maschine dem Gehirn gleicht – d. h., daß sie es nicht nur schreibt, sondern auch weiß, daß sie es geschrie-

ben hat. Kein Mechanismus kann sich über seine Erfolge freuen (und nicht nur künstlich signalisieren, was ein leichter Kunstgriff wäre), Kummer empfinden, wenn seine Röhren durchbrennen, durch schöne Worte sich geschmeichelt, durch Fehler sich niedergeschlagen fühlen, vom Sex bezaubert werden, zornig oder traurig sein, wenn er nicht bekommt, was er will.«

Dieses Argument läßt unseren Test geradezu als wertlos erscheinen. Gemäß der extremsten Form dieser Ansicht besteht die einzige Möglichkeit, sicher zu sein, daß eine Maschine denkt, darin, selbst die Maschine zu *sein* und sich selbst denken zu fühlen. Man könnte diese Gefühle dann der Welt mitteilen, doch hätte niemand Veranlassung, dem irgend Beachtung zu schenken. Entsprechend könnte man gemäß dieser Ansicht nur dann wissen, daß ein *Mensch* denkt, wenn man dieser betreffende Mensch ist. In der Tat ist das der solipsistische Standpunkt. Es mag der am meisten logische Standpunkt sein, jedoch macht er einen Gedankenaustausch recht schwierig. A kann leicht glauben: »A denkt, B jedoch nicht«, während B glaubt: »B denkt, aber A nicht«. Statt diesen Punkt unaufhörlich zu erörtern, ist es üblich, die höfliche Übereinkunft zu wählen, daß jedermann denkt.

Ich bin sicher, daß Professor Jefferson sich die extreme und solipsistische Ansicht nicht zu eigen machen möchte. Wahrscheinlich wäre er durchaus willens, das Imitationsspiel als Test zu akzeptieren. Das Spiel (bei dem man den Spieler B wegläßt) wird in der Praxis häufig unter dem Namen *viva voce* verwendet, um herauszufinden, ob jemand etwas wirklich versteht oder nur »gleich einem Papagei gelernt« hat. Wir wollen einmal Auszüge aus einem solchen *viva voce* anhören:

Fragesteller: In der ersten Zeile Ihres Sonetts, die lautet »Soll ich dich mit einem Sommertag vergleichen« würde nicht »Frühlingsmorgen« genauso oder besser passen?

Zeuge: Es würde sich nicht skandieren lassen.

Fragesteller: Wie wäre es mit »Wintertag«; das würde sich gewiß skandieren lassen.

Zeuge: Ja, doch möchte niemand mit einem Wintertag verglichen werden.

Fragesteller: Würden Sie sagen, daß Herr Pickwick Sie an Weihnachten erinnert?

Zeuge: In gewisser Weise.

Fragesteller: Weihnachten ist aber ein Wintertag, und ich glaube nicht, daß Herr Pickwick sich an dem Vergleich stören würde.

Zeuge: Ich glaube, daß kann nicht Ihr Ernst sein. Unter einem Wintertag versteht man einen typischen Wintertag, also keinen speziellen wie Weihnachten.

Und so weiter. Was würde Professor Jefferson sagen, wenn die

4 G. Jefferson, *The Mind of Mechanical Man. Lister Oration for 1949. British Medical Journal*, vol. 1 (1949), 1105–1121.

sonett-schreibende Maschine in der Lage wäre, in dieser Weise bei dem *viva voce* zu antworten? Ich weiß nicht, ob er der Ansicht wäre, daß die Maschine diese Antworten »nur künstlich signalisiert«, aber wenn die Antworten so zufriedenstellend und entschieden lauten wie oben, glaube ich nicht, daß er das als »leichten Kunstgriff« beschreiben würde. Diese Ausdrücke, denke ich, sollen sich wohl auf solche Vorrichtungen beziehen wie eine in die Maschine eingebaute Schallplatte, auf der einer ein Sonett liest, mit einem eingebauten Schalter, um sie von Zeit zu Zeit anzuschalten.

Kurz, ich glaube, die Mehrzahl derer, die das Bewußtseinsargument verfechten, könnte eher zu seiner Aufgabe überredet als in eine solipsistische Position gedrängt werden. Wahrscheinlich werden sie dann bereit sein, unseren Test zu akzeptieren.

Ich möchte nicht den Eindruck erwecken, als gäbe es für mich kein Geheimnis um das Bewußtsein. Zum Beispiel tritt im Zusammenhang mit jedem Versuch, es zu lokalisieren, eine Art Paradox auf. Aber ich denke nicht, daß diese Rätsel notwendig gelöst sein müssen, bevor wir die Frage beantworten können, mit der wir in diesem Aufsatz beschäftigt sind.

(5) *Argumente, die verschiedene Unfähigkeiten betreffen.* Diese Argumente sind von der Form: »Ich gestehe Ihnen zu, daß Sie Maschinen all die Dinge tun lassen können, die Sie erwähnt haben, es wird Ihnen jedoch nie gelingen, sie dahin zu bringen, X zu tun.« In diesem Zusammenhang werden zahlreiche Möglichkeiten für X vorgeschlagen. Ich gebe eine Auswahl:

Liebevoll, einfallsreich, schön, freundlich sein (S. 167), Initiative besitzen, Sinn für Humor haben, Recht von Unrecht unterscheiden, Fehler machen (S. 167f.), sich verlieben, gerne Erdbeeren mit Schlagsahne essen (S. 167), jemanden in sich verliebt machen, aus Erfahrung lernen (S. 177f.), Wörter richtig gebrauchen, Gegenstand seiner eigenen Gedanken sein (S. 168f.), über ebenso viele Verhaltensmöglichkeiten verfügen wie ein Mensch, etwas wirklich Neues tun (S. 170). (Einigen dieser Unfähigkeiten wird – wie durch die Seitenzahlen angezeigt – eine besondere Betrachtung gewidmet.)

Gewöhnlich werden solche Feststellungen durch nichts untermauert. Ich glaube, sie basieren zum größten Teil auf dem Prinzip wissenschaftlicher Induktion. Ein Mensch sieht im Laufe seines Lebens Tausende von Maschinen. Aus dem, was er von ihnen wahrnimmt, zieht er eine Reihe allgemeiner Schlüsse. Sie sind häßlich, jede ist für einen eng umrissenen Zweck entworfen und wertlos, wenn man

sie für einen geringfügig abweichenden Zweck braucht, die Anzahl der Verhaltensweisen ihrer aller ist sehr klein usw., usw. Er zieht den natürlichen Schluß, daß es sich hierbei um notwendige Eigenschaften der Maschinen im allgemeinen handelt. Viele dieser Einschränkungen stehen im Zusammenhang mit der sehr kleinen Speicherkapazität der meisten Maschinen. (Ich setze voraus, daß die Idee der Speicherkapazität sich in irgendeiner Weise auch auf andere Maschinen als nur Maschinen mit diskreten Zuständen erstreckt. Auf die exakte Definition kommt es hierbei nicht an, da in der vorliegenden Diskussion kein Anspruch auf mathematische Genauigkeit erhoben wird.) Vor ein paar Jahren, als man noch sehr wenig über Digitalrechner hörte, war es möglich, große Skepsis ihnen gegenüber hervorzurufen, wenn man ihre Eigenschaften erwähnte, ohne dabei ihre Konstruktion zu beschreiben. Das kam vermutlich von einer ähnlichen Anwendung des Prinzips der wissenschaftlichen Induktion. Diese Anwendungen des Prinzips sind sicher weitgehend unbewußt. Wenn ein gebranntes Kind das Feuer scheut und dies dadurch zum Ausdruck bringt, daß es das Feuer meidet, würde ich sagen, daß es die wissenschaftliche Induktion angewandt hat. (Ich könnte sein Verhalten natürlich auch auf manch andere Weise beschreiben.) Das Tun und die Gepflogenheiten der Menschen scheinen kein besonders geeignetes Objekt für die Anwendung wissenschaftlicher Induktion zu sein. Umfangreiche raum-zeitliche Untersuchungen sind erforderlich, um zuverlässige Resultate zu erhalten. Andernfalls könnten wir (wie die meisten englischen Kinder) zu dem Schluß kommen, daß jedermann Englisch spricht und es töricht ist, Französisch zu lernen.

Zu vielen der erwähnten Unfähigkeiten sind jedoch besondere Bemerkungen nötig. Die Unfähigkeit, für Erdbeeren mit Schlagsahne zu schwärmen, mag der Leser als frivol empfunden haben. Möglicherweise könnte man eine Maschine dazu bringen, sich an dieser köstlichen Speise zu erfreuen, doch jeder Versuch in dieser Richtung wäre idiotisch. Das Wesentliche an dieser Unfähigkeit ist, daß sie bei einigen der anderen Unfähigkeiten mitwirkt, z. B. bei der Schwierigkeit, die die Maschine hat, dem Menschen gegenüber ähnlich freundschaftliche Gefühle zu hegen, wie es sie zwischen Weißen und Weißen oder Schwarzen und Schwarzen gibt.

Die Behauptung, »Maschinen können keine Fehler machen«, scheint seltsam. Man ist versucht zu erwidern: »Sind sie deswegen

weniger wert?« Wir wollen jedoch eine wohlwollendere Haltung einnehmen und zu verstehen suchen, was wirklich gemeint ist. Ich glaube, daß sich diese Kritik in Terms des Imitationsspiels erklären läßt. Es wird behauptet, daß der Fragesteller die Maschine von einem Menschen dadurch unterscheiden kann, daß er beiden einige Rechenaufgaben vorlegt. Die Maschine würde sich durch ihre tödliche Genauigkeit entlarven. Die Antwort darauf ist einfach. Die (für das Spiel geeignet programmierte) Maschine wird nicht versuchen, die richtigen Antworten auf die Rechenprobleme zu geben. Sie würde absichtlich Fehler einführen, auf eine Weise berechnet, die den Fragesteller verwirren soll. Ein mechanisches Versagen würde sich wahrscheinlich in einer ungeeigneten Entscheidung äußern, welcher Fehler in der Rechnung zu machen ist. Doch selbst diese Auslegung der Kritik ist nicht wohlwollend genug. Aus Platzgründen können wir es uns jedoch nicht leisten, das Thema weiter zu verfolgen. Mir scheint, daß diese Kritik auf der Verwechslung zweier Arten von Fehlern beruht. Wir können sie »Funktionsfehler« und »Fehlschlüsse« nennen. Funktionsfehler sind durch ein mechanisches oder elektrisches Versagen bedingt, das die Maschine veranlaßt, sich anders als geplant zu verhalten. Bei philosophischen Diskussionen neigt man dazu, die Möglichkeit solcher Fehler zu ignorieren; man diskutiert deshalb »abstrakte Maschinen«. Diese abstrakten Maschinen sind eher mathematische Fiktionen als physikalische Objekte. Per definitionem sind Funktionsfehler ausgeschlossen. In diesem Sinne läßt sich in der Tat sagen, daß »Maschinen nie Fehler machen können«. Fehlschlüsse können nur dann entstehen, wenn den Ausgabesignalen der Maschine eine bestimmte Bedeutung gegeben wird. Die Maschine könnte z. B. mathematische Gleichungen oder Sätze auf Englisch ausdrücken. Wenn eine falsche Behauptung ausgedruckt wird, sagen wir, daß der Maschine ein Fehlschluß unterlaufen ist. Selbstverständlich gibt es keinerlei Grund zu sagen, daß eine Maschine keine Fehler dieser Art machen kann. Sie könnte z. B. nichts als immer wieder »0 = 1« ausdrücken. Um ein weniger abartiges Beispiel zu nehmen: sie könnte eine bestimmte Methode besitzen, durch wissenschaftliche Induktionen Schlüsse zu ziehen. Wir müssen erwarten, daß eine solche Methode gelegentlich zu falschen Ergebnissen führt.

Der Einwand, eine Maschine könne nicht Gegenstand ihres eigenen Denkens sein, läßt sich natürlich nur dann beantworten, wenn

gezeigt werden kann, daß sich die Maschine überhaupt irgendwelche Gedanken über *etwas* macht. Nichtsdestoweniger scheint »der Gegenstand der Operationen einer Maschine« tatsächlich etwas zu bedeuten, zumindest für die Leute, die mit ihr zu tun haben. Versuchte z. B. eine Maschine, eine Lösung der Gleichung $x^2 - 40x - 11 = 0$ zu finden, so wäre man versucht, diese Gleichung als Teil des momentanen Gegenstandes der Maschine zu beschreiben. In diesem Sinn kann eine Maschine zweifellos sich selbst zum Gegenstand haben. Man kann sie zur Ausarbeitung ihrer eigenen Programme verwenden oder zur Vorhersage des Effekts von Veränderungen ihrer eigenen Struktur. Durch Beobachtung der Ergebnisse ihres eigenen Verhaltens ist sie in der Lage, ihre eigenen Programme so zu modifizieren, daß sie ein bestimmtes Ziel effektiver erreicht. Das sind keine utopischen Träume, sondern Möglichkeiten der nahen Zukunft.

Der Einwand, eine Maschine habe keine große Verhaltensbreite, läuft auf die Behauptung hinaus, sie könne über keine große Speicherkapazität verfügen. Bis vor kurzem war eine Speicherkapazität von nur 1000 Ziffern sehr selten.

Die hier betrachteten kritischen Bemerkungen sind oft verkleidete Formen des Bewußtseinsarguments. Hält man daran fest, daß eine Maschine eines dieser Dinge ausführen *kann*, und beschreibt man die Art der Methode, welche sie dabei benutzen könnte, wird man gewöhnlich keinen großen Eindruck machen. Man ist überzeugt, daß die Methode (worin sie auch immer besteht, denn sie muß jedenfalls mechanisch sein) in Wirklichkeit ziemlich primitiv sei. Vergleiche die Parenthese in der Stellungnahme Jeffersons auf S. 165.

(6) *Der Einwand von Lady Lovelace*. Unsere detaillierteste Information über Babbages Analytische Maschine stammt aus einem Bericht von Lady Lovelace.⁵ Sie stellt darin fest: »Die Analytische Maschine erhebt keinen Anspruch, irgend etwas zu *erzeugen*. Sie kann ausführen, *was immer wir ihr zu befehlen wissen*« (ihre Hervorhebungen). Diese Feststellung wird von Hartree zitiert, der hinzufügt: »Dies impliziert nicht, daß es nicht doch möglich sein könnte, ein elektronisches Gerät zu konstruieren, welches »für sich selbst denken« wird oder in welches sich, biologisch ausgedrückt, ein bedingter Reflex

⁵ Countess of Lovelace, Translator's notes to an article on Babbage's Analytical Engine, Scientific Memoirs (ed. R. Taylor), vol. 3 (1842), 691–731.

einbauen ließe, der als Grundlage für einen ›Lernprozeß‹ dienen könnte. Ob dies grundsätzlich möglich ist oder nicht, ist eine an- und aufregende Frage, die sich durch einige neuere Entwicklungen aufdrängt. Es hatte jedoch nicht den Anschein, als verfügten die damals konstruierten oder geplanten Maschinen über diese Eigenschaft.«⁶

Hierin stimme ich völlig mit Hartree überein. Wie man sieht, behauptet er nicht, daß die fraglichen Maschinen die Eigenschaft nicht besaßen, sondern vielmehr, daß das, was Lady Lovelace sehen konnte, sie nicht zu dem Glauben ermutigte, sie hätten sie. Es ist sehr gut möglich, daß die fraglichen Maschinen in einem gewissen Sinn diese Eigenschaft hatten. Denn angenommen, eine diskrete Maschine besitzt diese Eigenschaft. Die Analytische Maschine war ein universaler Digitalrechner, so daß sie bei ausreichender Speicherkapazität und Geschwindigkeit und durch geeignete Programmierung in die Lage hätte versetzt werden können, die fragliche Maschine nachzuahmen. Wahrscheinlich kam der Gräfin oder Herrn Babbage dieses Argument nicht in den Sinn. Jedenfalls bestand für sie keine Verpflichtung, alles zu beanspruchen, was hätte beansprucht werden können.

Der ganze Fragenkomplex wird nochmals im Abschnitt über lernende Maschinen erörtert werden.

Eine Variante des Einwands von Lady Lovelace besagt, daß eine Maschine »nie etwas wirklich Neues ausführen kann«. Dem mag für einen Augenblick mit dem Sprichwort entgegnet werden: »Es gibt nichts Neues unter der Sonne.« Wer kann sicher sein, daß eine »schöpferische Arbeit«, die er geleistet hat, nicht nur die Frucht des Samens war, der ihm durch Unterricht eingepflanzt wurde, oder das Ergebnis eines Nachvollzugs wohlbekannter allgemeiner Prinzipien? Eine bessere Variante des Einwands besagt, daß eine Maschine uns nie »überraschen« kann. Diese Feststellung ist eine direktere Herausforderung, der man ebenso direkt begegnen kann. Maschinen überraschen mich sehr häufig. Das liegt größtenteils daran, daß ich keine ausreichenden Kalkulationen anstelle, um zu entscheiden, was von ihnen zu erwarten ist, oder vielmehr, obwohl ich Berechnungen anstelle, daß ich sie übereilt und nachlässig ausführe, Risiken auf mich nehmend. Ich sage mir vielleicht, »ich vermute, die Spannung müßte

hier wie dort dieselbe sein: jedenfalls wollen wir annehmen, daß es so ist«. Selbstverständlich irre ich mich oft, und das Ergebnis ist eine Überraschung für mich, denn bis das Experiment durchgeführt ist, sind diese Annahmen bereits vergessen. Dieses Eingeständnis mag mich Belehrungen hinsichtlich meiner mangelhaften Methoden aussetzen, zieht jedoch nicht meine Glaubwürdigkeit in Zweifel, wenn ich die Überraschungen bezeuge, die ich erlebe.

Ich erwarte von dieser Erwiderung nicht, daß sie meinen Kritiker zum Schweigen bringt. Er wird wahrscheinlich einwenden, solche Überraschungen seien die Folge eines schöpferischen Denkvorgangs meinerseits und nicht ein Verdienst der Maschine. Dies führt uns wieder zurück zu dem Bewußtseinsargument und weit weg von dem Moment der Überraschung. Damit müssen wir unsere Argumentation als abgeschlossen betrachten, aber vielleicht ist es noch wert anzumerken, daß es, um etwas als Überraschung zu empfinden, stets genauso sehr eines »schöpferischen Denkvorgangs« bedarf, gleichgültig ob die Überraschung nun von einem Menschen, einem Buch, einer Maschine oder von sonst etwas verursacht wird.

Die Ansicht, Maschinen könnten nicht Anlaß zu Überraschungen geben, ist meiner Meinung nach die Folge eines Trugschlusses, dem Philosophen und Mathematiker besonders unterliegen. Er besteht in der Annahme, daß sich, sobald eine Tatsache dem Verstand präsentiert wird, auch gleichzeitig sämtliche Folgen dieser Tatsache dem Verstand offenbaren. Bei vielen Gelegenheiten ist das eine sehr nützliche Annahme, man vergißt jedoch zu leicht, daß sie falsch ist. Eine natürliche Folge davon ist, daß man schließlich annimmt, daß die schlichte Herleitung von Folgerungen aus Daten und allgemeinen Prinzipien kein Verdienst sei.

(7) *Das Argument der Stetigkeit innerhalb des Nervensystems.* Das Nervensystem ist mit Sicherheit keine diskrete Maschine. Ein kleiner Informationsfehler bezüglich der Größe eines Nervenimpulses, der eine Nervenzelle trifft, kann auf die Größe des austretenden Impulses einen großen Einfluß haben. Es könnte argumentiert werden, daß, weil dem so ist, man nicht erwarten kann, in der Lage zu sein, das Verhalten des Nervensystems mit einem diskreten System nachzuahmen.

Es ist wahr, daß sich eine diskrete Maschine von einer stetigen Maschine unterscheiden muß. Wenn wir aber an den Bedingungen

⁶ D. R. Hartree, *Calculating Instruments and Machines*, New York 1949, S. 70.

des Imitationsspiels festhalten, wird der Fragesteller nicht in der Lage sein, aus diesem Unterschied irgendeinen Vorteil zu ziehen. Die Situation läßt sich verdeutlichen, wenn wir eine einfachere stetige Maschine betrachten. Ein Differentialanalysator eignet sich hierfür gut. (Ein Differentialanalysator ist eine bestimmte Art Maschine, nicht vom diskreten Typ, die bei bestimmten Berechnungen verwendet wird.) Einige von ihnen geben ihre Antworten in ausgedruckter Form und sind also geeignet, am Spiel teilzunehmen. Für einen Digitalrechner wäre es nicht möglich, exakt vorherzusagen, welche Antworten der Differentialanalysator auf ein Problem geben würde, aber er wäre sehr wohl in der Lage, die richtige Art von Antwort zu geben. Wenn man z. B. nach dem Wert von π fragt (ungefähr 3,1416), wäre es vernünftig, zwischen den Zahlen 3,12; 3,13; 3,14; 3,15; 3,16 mit den Wahrscheinlichkeiten 0,05; 0,15; 0,55; 0,19; 0,06 eine Zufallswahl zu treffen. Unter diesen Umständen wäre es für den Fragesteller sehr schwierig, den Differentialanalysator vom Digitalrechner zu unterscheiden.

(8) *Das Argument aus dem informellen Charakter des Verhaltens.* Es ist unmöglich, eine Menge von Regeln aufzustellen, die festlegen, was ein Mensch in jeder denkbaren Situation tun sollte. Es mag z. B. eine Regel geben, daß man bei Rotlicht im Straßenverkehr anhalten muß und bei Grün freie Fahrt hat; was aber tun, wenn Rot und Grün infolge einer Störung gleichzeitig aufleuchten? Man könnte vielleicht entscheiden, daß Anhalten das Sicherste ist. Doch eine neue Schwierigkeit könnte sich später aus dieser Entscheidung ergeben. Der Versuch, Verhaltensregeln für alle Eventualitäten festzulegen, selbst solche, die bei Verkehrsampeln auftauchen, scheint unmöglich zu sein. Mit alledem stimme ich überein.

Von daher wird argumentiert, daß wir keine Maschinen sein können. Ich werde versuchen, das Argument wiederzugeben, fürchte jedoch, daß ich ihm kaum Gerechtigkeit widerfahren lassen werde. Es lautet etwa so: »Besäße jeder Mensch eine festumrissene Menge von Verhaltensregeln, nach denen er sein Leben regelt, so wäre er nicht besser als eine Maschine. Aber es gibt keine solchen Regeln, also können Menschen nicht Maschinen sein.« Ganz offensichtlich handelt es sich hierbei um einen falsch angewandten Mittelbegriff. Ich glaube nicht, daß das Argument jemals so formuliert wird, aber ich glaube, das ist trotzdem das verwendete Argument. Jedoch könnte es

eine Verwechslung zwischen »Verhaltensregeln« und »Verhaltensgesetzen« geben, die die Sache verschleiern. Unter »Verhaltensregeln« verstehe ich Vorschriften, z. B. »Halte bei Rotlicht an«, nach denen man sich richten und deren man sich bewußt sein kann. Unter »Verhaltensgesetzen« verstehe ich Naturgesetze, die den menschlichen Körper betreffen, z. B. »wenn du ihn kneifst, schreit er.« Ersetzen wir im erwähnten Argument »Verhaltensregeln, nach denen er sein Leben regelt« durch »Verhaltensgesetze, die sein Leben regeln«, ist das Hindernis des falsch angewandten Mittelbegriffs nicht mehr unüberwindlich. Denn wir halten es nicht allein für wahr, daß von Verhaltensgesetzen gelenkt sein impliziert, eine Art Maschine zu sein (wenn auch nicht notwendig eine diskrete Maschine), sondern daß auch umgekehrt eine solche Maschine sein impliziert, von solchen Gesetzen gelenkt sein. Es gelingt uns jedoch nicht so leicht, uns vom Nichtvorhandensein vollständiger Verhaltensgesetze zu überzeugen wie vom Fehlen vollständiger Verhaltensregeln. Die einzige uns bekannte Möglichkeit, solche Gesetze herauszufinden, besteht in wissenschaftlicher Beobachtung, und wir kennen sicher keinen Fall, in dem wir behaupten könnten: »Wir haben genügend nachgeforscht. Es gibt keine solchen Gesetze.«

Wir können noch zwingender zeigen, daß jede solche Feststellung ungerechtfertigt wäre. Denn angenommen, wir dürften sicher sein, solche Gesetzmäßigkeiten aufzudecken, falls sie existieren. Dann sollte es bei einer diskreten Maschine gewiß möglich sein, durch Beobachtung genügend über sie zu erfahren, um ihr zukünftiges Verhalten vorherzusagen, und zwar innerhalb einer vernünftigen Zeitspanne, sagen wir, innerhalb von tausend Jahren. Dies scheint jedoch nicht der Fall zu sein. Dem Manchester-Rechner habe ich ein kurzes Programm eingegeben, welches nur 1000 Speichereinheiten benützt; die Maschine antwortet nach Eingabe einer sechzehnstelligen Zahl innerhalb von zwei Sekunden mit einer anderen. Ich traue es niemandem zu, aus diesen Antworten das Programm so hinreichend kennenzulernen, daß er in der Lage ist, irgendwelche künftige Antworten auf noch nicht probierte Zahlenwerte vorherzusagen.

(9) *Das Argument von der außersinnlichen Wahrnehmung.* Ich nehme an, daß der Leser mit der Idee der außersinnlichen Wahrnehmung vertraut ist sowie mit der Bedeutung ihrer vier Bereiche, nämlich Telepathie, Hellsehen, Prophetie und Psychokinese. Diese irritieren-

den Phänomene scheinen allen unseren üblichen wissenschaftlichen Vorstellungen zu widersprechen. Wie gerne würden wir sie leugnen! Unglücklicherweise sind die statistischen Beweise, zumindest für Telepathie, überwältigend. Es ist sehr schwierig, unsere Vorstellungen so umzumodeln, daß diese neuen Tatsachen hineinpassen. Hat man sie erst einmal akzeptiert, so scheint der Schritt zum Glauben an Geister und Kobolde nicht groß. Die Vorstellung, daß unsere Körper sich nach bekannten physikalischen und anderen, bis jetzt noch unentdeckten, doch irgendwie ähnlichen Gesetzen bewegen, wäre eine der ersten, die man aufgeben müßte.

Ich halte dieses Argument für sehr gewichtig. Als Erwiderung kann man sagen, daß viele wissenschaftliche Theorien in der Praxis ausführbar erscheinen, obwohl sie zu der außersinnlichen Wahrnehmung im Widerspruch stehen; daß man in der Tat sehr hübsch vorankommt, wenn man sie außer acht läßt. Das ist ein ziemlich schlechter Trost, und der Verdacht liegt nahe, daß gerade Denken die Art von Phänomen ist, bei der außersinnliche Wahrnehmung besonders relevant sein könnte.

Ein spezifischeres Argument, welches sich auf außersinnliche Wahrnehmung gründet, könnte folgendermaßen lauten⁷: »Man spiele das Imitationsspiel mit einem Menschen, der als telepathischer Empfänger geeignet ist, und einem Digitalrechner als Zeugen. Der Fragesteller kann Fragen stellen wie: »Zu welcher Farbe gehört die Karte in meiner rechten Hand?« Der Mensch gibt dank Telepathie oder Hellseherei bei 400 Karten 130mal die richtige Antwort. Die Maschine kann nur rein zufällig raten und gibt vielleicht 104 richtige Antworten, so daß der Fragesteller die beiden Teilnehmer richtig identifiziert.« Hier eröffnet sich eine interessante Möglichkeit. Angenommen, der Digitalrechner verfügt über einen Zufallsgenerator. Dann wird es natürlich sein, diesen für die Wahl der jeweiligen Antwort zu benutzen. Dieser Zufallsgenerator wird nun aber den psychokinetischen Kräften des Fragestellers unterworfen sein. Vielleicht könnte diese Psychokinese die Maschine öfter richtig raten lassen, als man aufgrund einer Wahrscheinlichkeitsrechnung erwartet, so daß der Fragesteller dadurch nicht in der Lage wäre, richtig zu identifizie-

ren. Andererseits könnte er, ohne überhaupt zu fragen, doch in der Lage sein, richtig zu raten, und zwar durch Hellsehen. Bei außersinnlicher Wahrnehmung ist alles möglich.

Läßt man Telepathie zu, so wird es nötig sein, unseren Test enger zu fassen. Man könnte die Situation analog zu einer solchen betrachten, die aufträte, wenn der Fragesteller mit sich selbst spräche und einer der Teilnehmer an der Wand horchte. Alle Forderungen wären erfüllt, wenn man die Beteiligten jeweils in einen »telepathie-sicheren Raum« stellte.

7. LERNENDE MASCHINEN

Der Leser wird geahnt haben, daß ich über keine sehr überzeugenden Argumente positiver Art verfüge, um meine Ansichten zu stützen. Besäße ich sie, hätte ich nicht solche Mühe auf mich genommen, Trugschlüsse in anderen Auffassungen aufzuzeigen. Die Beweise, die ich habe, werde ich nun anführen.

Wir wollen kurz auf Lady Lovelaces Einwand zurückkommen, welcher besagte, daß die Maschine nur das tun kann, was wir ihr zu tun befehlen. Man könnte sagen, daß ein Mensch der Maschine eine Idee »injizieren« kann, daß die Maschine bis zu einem gewissen Grad antwortet und dann in Schweigen verfällt, wie eine von einem Hammer angeschlagene Klaviersaite. Ein weiterer Vergleich wäre ein Atomreaktor unterhalb der kritischen Größe: die injizierte Idee soll hier einem Neutron entsprechen, welches von außen in den Reaktor eindringt. Jedes solche Neutron wird eine gewisse Störung verursachen, die schließlich abklingt. Ist der Reaktor jedoch hinreichend groß, wird die durch das eingedrungene Neutron verursachte Störung höchstwahrscheinlich eine Kettenreaktion auslösen, die schließlich zur Zerstörung des Reaktors führt. Gibt es eine entsprechende Erscheinung für den menschlichen Verstand bzw. für Maschinen? Für den menschlichen Verstand scheint dies der Fall zu sein. Der Verstand der meisten Menschen scheint »unkritisch« zu sein, d. h. er entspricht bei dieser Analogie den Reaktoren unterkritischer Größe. Eine einem solchen Verstand präsentierte Idee ruft als Erwiderung im Durchschnitt weniger als eine Idee hervor. Nur ein winziger Teil ist überkritisch: Eine einem solchen Verstand mitgeteilte Idee ruft eine

⁷ Vgl. dazu den Bericht über die Experimente mit *Nimrod*, in diesem Band S. 118. – *Ann. d. Hrsg.*

ganze »Theorie« hervor, bestehend aus sekundären, tertiären und noch fernerliegenden Ideen. Bei Tieren ist der Verstand eindeutig unterkritisch. In Anlehnung an diese Analogie fragen wir: »Läßt sich eine superkritische Maschine konstruieren?«

Auch der Vergleich mit den Schalen einer Zwiebel ist dienlich. Wenn wir die Funktionen des Verstandes oder Gehirns betrachten, stoßen wir auf bestimmte Vorgänge, die wir in rein mechanischen Terms erklären können. Dies entspricht unseres Erachtens nicht dem wirklichen Verstand: sie sind eine Art Schale, die wir entfernen müssen, wenn wir den wirklichen Verstand finden wollen. Um das herum, was übrigbleibt, finden wir aber wieder eine Schale, die es zu entfernen gilt, und so weiter. Wenn wir auf diese Weise fortschreiten, stoßen wir dann jemals auf den »wirklichen« Verstand oder kommen wir schließlich zu der Schale, die nichts enthält? Im letzteren Fall ist der ganze Verstand mechanisch. (Er wäre jedoch keine diskrete Maschine; wir haben dies diskutiert.)

Diese letzten beiden Abschnitte erheben keinen Anspruch darauf, überzeugende Argumente zu sein. Sie sollten eher als »Darlegungen, die Glauben wecken möchten« beschrieben werden.

Der einzig wirklich befriedigenden Untermauerung der am Anfang von Abschnitt 6 ausgesprochenen Ansicht besteht darin, das Ende des Jahrhunderts abzuwarten und dann das beschriebene Experiment durchzuführen. Was aber können wir in der Zwischenzeit sagen? Welche Schritte sollten wir jetzt unternehmen, wenn das Experiment erfolgreich sein soll?

Wie dargelegt, ist dies hauptsächlich eine Frage der Programmierung. Fortschritte in der Ingenieurwissenschaft müssen ebenfalls erzielt werden, aber es scheint unwahrscheinlich, daß diese den Erfordernissen nicht angemessen sein werden. Schätzungen hinsichtlich der Speicherkapazität des Gehirns schwanken zwischen 10^{10} und 10^{15} bits. Ich tendiere zu den niedrigeren Werten und glaube, daß bei höheren Denkvorgängen nur ein sehr kleiner Bruchteil benützt wird. Der größte Teil wird wahrscheinlich zur Speicherung visueller Eindrücke verwendet. Es sollte mich überraschen, wenn mehr als 10^9 erforderlich wären, um das Imitationsspiel befriedigend zu spielen, zumindest gegen einen Blinden. (Beachte: Die Kapazität der *Encyclopaedia Britannica*, 11. Auflage, beträgt 2×10^9 .) Eine Speicherkapazität von 10^7 wäre, sogar bei dem gegenwärtigen Stand der Technik,

eine sehr praktikable Möglichkeit. Es ist wahrscheinlich nicht erforderlich, die Rechengeschwindigkeit der Maschinen überhaupt zu vergrößern. Teile moderner Maschinen, die man als Analoga zu Nervenzellen betrachten kann, arbeiten ca. tausendmal schneller als die letzteren. Dies sollte ein »Sicherheitsspielraum« sein, der Geschwindigkeitsverluste verschiedenster Art ausgleichen könnte. Unser Problem besteht nun darin, wie die Maschinen für das Spiel zu programmieren sind. Bei meinem gegenwärtigen Arbeitstempo stelle ich pro Tag ein etwa tausendziffriges Programm her, so daß etwa sechzig Arbeiter, die fünfzig Jahre hindurch ununterbrochen arbeiten, die Aufgabe bewältigen könnten, sofern nichts im Papierkorb landet. Eine etwas schnellere Methode scheint wünschenswert.

Bei dem Versuch, den Verstand eines erwachsenen Menschen nachzuahmen, müssen wir uns über den Vorgang klar werden, der zu seinem gegenwärtigen Zustand geführt hat. Es lassen sich drei Komponenten feststellen:

- (a) der Anfangszustand des Verstandes, sagen wir bei Geburt,
- (b) die Erziehung, der er unterworfen wurde,
- (c) andere Erfahrungen, denen er unterworfen war und die nicht als Erziehung zu beschreiben sind.

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. Vermutlich ist das kindliche Gehirn mit einem Notizbuch vergleichbar, das man beim Schreibwarenhändler kauft. Wenig Mechanismus und viele leere Blätter. (Mechanismus und Geschriebenes sind bei unserer Betrachtungsweise nahezu synonym.) Wir hoffen, daß das kindliche Gehirn so wenig Mechanismus hat, daß es leicht programmiert werden kann. Wir können, als erste Näherung, annehmen, daß der Aufwand an Erziehungsarbeit im wesentlichen der gleiche ist wie bei einem menschlichen Kind.

Wir haben somit unser Problem zweigeteilt: in das Kind-Programm und den Erziehungsprozeß. Beides bleibt sehr eng miteinander verbunden. Wir können nicht erwarten, auf Anhieb eine gute Kind-Maschine zu finden. Man muß mit einer solchen Maschine Leherexperimente durchführen, um festzustellen, wie gut sie lernt.

Man kann es dann mit einer anderen versuchen und sehen, ob sie besser oder schlechter lernt. Es besteht ein offensichtlicher Zusammenhang zwischen diesem Prozeß und der Evolution, und zwar durch folgende Gleichsetzungen:

Struktur der Kind-Maschine	=	Erbgut
Veränderungen der Kind-Maschine	=	Mutationen
Natürliche Auswahl	=	Werturteil des Experimentators.

Man darf jedoch hoffen, daß dieser Prozeß schneller vor sich geht als die Evolution. Das Überleben des Stärkeren ist eine langsame Methode zum Messen von Vorteilen. Der Experimentator sollte, durch Anwendung von Intelligenz, in der Lage sein, den Prozeß zu beschleunigen. Ebenso wichtig ist die Tatsache, daß er nicht auf zufällige Mutationen beschränkt ist. Wenn er den Grund für irgendeine Schwäche feststellen kann, so kann er sich wahrscheinlich auch die Mutationsart vorstellen, die ihr abhilft.

Es wird nicht möglich sein, die Maschine dem gleichen Unterrichtsprozeß zu unterziehen wie ein normales Kind. Sie wird z. B. keine Beine haben, so daß man sie nicht auffordern könnte, hinauszugehen und den Kohleneimer zu füllen. Möglicherweise hat sie keine Augen. Aber wie gut diese Mängel auch immer durch geschickte Maschinenbaukunst ausgeglichen sein mögen, man kann die Kreatur nicht zur Schule schicken, ohne daß sich die anderen Kinder maßlos über sie lustig machten. Man muß ihr einen besonderen Unterricht angedeihen lassen. Die Füße, Augen, usw. brauchen uns nicht zu sehr zu bekümmern. Das Beispiel der *Helen Keller*⁸ zeigt, daß Erziehung möglich ist, vorausgesetzt, daß wechselseitige Kommunikation zwischen Lehrer und Schüler durch irgendwelche Mittel erfolgen kann.

Normalerweise verbinden wir den Lehrprozeß mit Bestrafungen und Belohnungen. Auf diesem Prinzip lassen sich einige einfache Kind-Maschinen konstruieren oder programmieren. Die Maschine muß so konstruiert werden, daß Ereignisse, die einem Bestrafungssignal unmittelbar vorausgingen, sich nur mit sehr geringer Wahrscheinlichkeit wiederholen, während ein Belohnungssignal die Wahrscheinlichkeit für eine Wiederholung derjenigen Ereignisse, die

es auslösten, erhöhen würde. Diese Definitionen setzen seitens der Maschine keinerlei Gefühle voraus. Ich habe mit einer derartigen Kind-Maschine einige Experimente durchgeführt und sie erfolgreich ein paar Dinge gelehrt, die Lehrmethode war jedoch zu unorthodox, um das Experiment als wirklich erfolgreich betrachten zu können.

Die Verwendung von Bestrafungen und Belohnungen kann bestenfalls ein Teil des Lehrprozesses sein. Grob ausgedrückt: verfügt der Lehrer gegenüber dem Schüler über keine anderen Kommunikationsmittel, so reicht die Informationsmenge, die ihm übermittelt werden kann, nicht über die Gesamtzahl angewandter Belohnungen und Bestrafungen hinaus. Bis ein Kind gelernt hat, das Wort »Casabianca« aufzusagen, wäre es wahrscheinlich sehr unglücklich, wenn das Wort nur mit Hilfe einer »Zwanzig Fragen«-Methode gefunden werden könnte, wobei jedes »Nein« mit Schlägen verbunden wäre. Es ist deshalb nötig, daß man über gewisse andere, »nicht-emotionale« Kommunikationskanäle verfügt. Sind diese vorhanden, so ist es möglich, einer Maschine durch Bestrafungen und Belohnungen beizubringen, daß sie den in irgendeiner Sprache, z. B. einer symbolischen Sprache erteilten Befehlen gehorcht. Diese Befehle müssen über die »nicht-emotionalen« Kanäle übermittelt werden. Die Verwendung jener Sprache wird die Anzahl der erforderlichen Bestrafungen und Belohnungen stark vermindern.

Über die für die Kind-Maschine angemessene Komplexität mögen die Meinungen auseinandergehen. Man könnte versuchen, die Maschine im Einklang mit den allgemeinen Prinzipien so einfach wie möglich zu gestalten. Zum andern könnte man ein vollständiges System logischer Schlußweisen »einbauen«⁹. Im letzteren Fall wäre der Speicher größtenteils mit Definitionen und Sätzen ausgefüllt. Die Sätze könnten von ganz unterschiedlichem Status sein, z. B. wohlbegründete Tatsachen, Vermutungen, mathematisch bewiesene Sätze, autoritative Feststellungen, Ausdrucksweisen, die die logische Form von Sätzen, aber keinen Glaubwürdigkeitswert haben. Gewisse Sätze lassen sich als »Befehle« beschreiben. Die Maschine sollte so konstruiert sein, daß jeder Befehl, der als »wohlbegründet« klassifiziert ist, automatisch die entsprechende Aktion nach sich zieht. Zur Illu-

⁸ Helen Keller wurde als Kleinkind blind und taub. Sie berichtet von ihrer Entwicklung in ihren beiden Büchern: *Die Geschichte meines Lebens*, Stuttgart 1904; sowie *Meine Welt*, Stuttgart 1909. – *Anm. d. Hrsg.*

⁹ Oder besser »hineinprogrammieren«, denn unsere Kind-Maschine wird in einen Digitalrechner programmiert. Das logische System muß jedoch nicht gelernt werden.

stration wollen wir annehmen, daß der Lehrer sagt: »Mache jetzt deine Hausaufgaben.« Das mag bewirken, daß »Der Lehrer sagt: »Mache jetzt Deine Hausaufgaben«« unter die wohlbegründeten Tatsachen eingereiht wird. Eine weitere solche Tatsache könnte lauten: »Alles, was der Lehrer sagt, ist wahr.« Beides zusammengekommen könnte schließlich dazu führen, daß der Befehl »mache jetzt deine Hausaufgaben« unter die wohlbegründeten Tatsachen eingereiht wird, was aufgrund der Konstruktion der Maschine bedeutet, daß mit den Hausaufgaben wirklich begonnen wird; dieser Effekt ist sehr zufriedenstellend. Die Schlußweisen der Maschine müssen nicht derart sein, daß sie die exaktesten Logiker zufriedenstellen. Zum Beispiel braucht es keine Hierarchie der Typen zu geben. Dies muß nicht heißen, daß Typenfehler (type fallacies) auftreten werden, genausowenig wie wir an Felsen abstürzen müssen, nur weil zufällig kein Schutzzaun vorhanden ist. Geeignete Befehle (*innerhalb* der logischen Systeme ausgedrückt und nicht Bestandteile der Regeln des Systems), z. B. »Verwende keine Klasse, außer sie ist Unterklasse einer vom Lehrer ausdrücklich erwähnten«, können einen ähnlichen Effekt haben wie »Gehe nicht zu nahe an den Rand«.

Die Befehle, die eine Maschine, die keine Gliedmaßen hat, befolgen kann, müssen weitgehend von intellektuellem Charakter sein, wie bei dem Beispiel der Hausaufgaben. Unter solchen Befehlen sind diejenigen wichtig, die die Reihenfolge festlegen, in der die Regeln des betreffenden logischen Systems anzuwenden sind. Denn bei der Anwendung eines logischen Systems existiert auf jeder Stufe eine Vielfalt von Alternativschritten, von denen jeder angewendet werden darf, soweit die Regeln des logischen Systems dabei nicht verletzt werden. Dieses Auswählen macht den Unterschied zwischen einem brillanten und einem törichten Denker, nicht den Unterschied zwischen einem folgerichtigen und einem trugschlüssigen. Richtlinien, die zu derartigen Befehlen führen, könnten lauten: »Verwende den Modus Barbara, sobald Sokrates erwähnt wird«, oder: »Erweist sich eine Methode als schneller denn eine andere, so benütze nicht die langsamere.« Einige davon könnten »autoritär gegeben«, andere hingegen von der Maschine selbst, z. B. durch wissenschaftliche Induktion, aufgestellt werden.

Die Vorstellung einer lernenden Maschine mag einigen Lesern paradox vorkommen. Wie können sich die Operationsregeln der Ma-

schine ändern? Sie sollen, unabhängig von der Vergangenheit der Maschine und unabhängig von den Änderungen, die mit ihr geschehen, vollständig beschreiben, wie die Maschine reagieren wird. Insofern sind die Regeln absolut zeit-invariant. Das ist durchaus richtig. Die Erklärung dieses Paradoxons besteht darin, daß die Regeln, die beim Lernprozeß geändert werden, nur von ziemlich wenig anmaßender Art sind, indem sie nur vorübergehend Gültigkeit beanspruchen. Der Leser möge dies mit der Verfassung der Vereinigten Staaten vergleichen.

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. Dies dürfte in stärkstem Maße auf die spätere Erziehung einer Maschine zutreffen, die aus einer Kind-Maschine bewährter Konstruktion (Programm) hervorgegangen ist. Das steht in offensichtlichem Gegensatz zu dem normalen Vorgang, bei dem man eine Maschine zum Rechnen benützt: man hat dann das Ziel, beim Rechenvorgang immer ein klares Bild vom augenblicklichen Zustand der Maschine zu haben. Dieses Ziel läßt sich nur mit Mühe erreichen. Die Ansicht, daß »die Maschine nur das tun kann, was wir ihr zu befehlen wissen«¹⁰, erscheint in dieser Hinsicht seltsam. Die meisten Programme, die wir in die Maschine eingeben können, werden zur Folge haben, daß die Maschine etwas tut, was wir uns überhaupt nicht erklären können oder als völlig zufälliges Verhalten auffassen. Intelligentes Verhalten besteht vermutlich in einem Abweichen vom völlig disziplinierten Verhalten bei Rechenvorgängen, aber einem ziemlich geringfügigen, das nicht zu zufälligem Verhalten oder sinnlosen, sich wiederholenden Schleifen führt. Ein weiteres wichtiges Ergebnis für die Vorbereitung unserer Maschine auf ihre Rolle im Imitationsspiel mit Hilfe eines Lehr- und Lernprozesses ist es, daß »menschliche Fehlbarkeit« wahrscheinlich auf ziemlich natürliche Weise unter den Tisch fällt, d. h. ohne besondere »Nachhilfe«. (Der Leser sollte das mit der Ansicht auf S. 167f. in Einklang bringen.) Prozesse, die gelernt sind, zeitigen nicht eine hundertprozentige Sicherheit des Ergebnisses; wäre dies der Fall, könnten sie nicht verlernt werden.

10 Vergleiche Lady Lovelaces Feststellung, die nicht das Wort »nur« enthält.

Es ist wahrscheinlich klug, in eine lernende Maschine ein Zufallselement mit einzubeziehen (s.S.155). Ein zufälliges Element erweist sich bei der Suche nach der Lösung eines Problems als sehr nützlich. Angenommen, wir wollten eine Zahl zwischen 50 und 200 finden, die gleich dem Quadrat der Summe ihrer Ziffern ist, so könnten wir mit 51 anfangen, es dann mit 52 versuchen, bis wir auf eine Zahl mit dieser Eigenschaft stoßen. Eine andere Möglichkeit bestünde darin, Zahlen zufällig auszuwählen, bis wir eine passende finden. Diese Methode hat den Vorzug, daß es unnötig ist, sich die schon untersuchten Zahlen zu merken, jedoch den Nachteil, daß man möglicherweise dieselbe zweimal ausprobiert, aber das ist nicht sehr wesentlich, wenn es mehrere Lösungen gibt. Die systematische Methode hat den Nachteil, daß es in dem Bereich, der zuerst untersucht werden muß, einen beträchtlichen Block ohne eine einzige Lösung geben kann. Der Lernprozeß läßt sich jetzt als das Suchen nach einer Verhaltensform betrachten, die den Lehrer (oder ein anderes Kriterium) zufriedenstellt. Da es wahrscheinlich eine große Anzahl an befriedigenden Lösungen gibt, erscheint die Zufallsmethode besser als die systematische. Man beachte auch, daß sie bei dem analogen Prozeß der Evolution verwendet wird. Allerdings ist dort die systematische Methode nicht möglich. Wie könnte man sich die verschiedenen genetischen Kombinationen, die ausprobiert wurden, merken, um Wiederholungen zu vermeiden?

Wir dürfen hoffen, daß Maschinen schließlich auf allen rein intellektuellen Gebieten mit dem Menschen konkurrieren. Aber mit welchen sollte man am besten beginnen? Auch dies ist eine schwierige Entscheidung. Viele glauben, daß eine sehr abstrakte Tätigkeit, beispielsweise das Schachspielen, am geeignetsten wäre. Ebenso kann man behaupten, daß es das beste wäre, die Maschine mit den besten Sinnesorganen auszustatten, die überhaupt für Geld zu haben sind, und sie dann zu lehren, englisch zu verstehen und zu sprechen. Dieser Prozeß könnte sich wie das normale Unterrichten eines Kindes vollziehen. Dinge würden erklärt und benannt werden, usw. Wiederum weiß ich nicht, welches die richtige Antwort ist, aber ich meine, daß man beide Ansätze erproben sollte.

Wir können nicht sehr weit vorausschauen, aber wir sehen vieles, was getan werden muß.

THE STATE OF THE ART

VORTRAG VOR DER LONDON MATHEMATICAL SOCIETY
AM 20. FEBRUAR 1947