edward f. moore

Neue, Forum, Mai 1973, Wien
S. 47–49

was kann ein automat?

(1973)

übersetzt,ausgewählt und kommentiert von peter weibel

WEIBEL:

die abendländische generallinie ist die dialektik von abstraktion und sinnlichkeit, von geist und körper, von bewußtsein und welt. sie ist begleitet von den begriffen erfahrung, erklärung, experiment und getrieben von der sehnsucht, mechanische verfahren und vorrichtungen aufzufinden. was die erklärung betrifft, sollen sie die denkschritte und -weisen vereinfachen; was das experiment angeht, die effektivität und operationale reichweite steigern; und was die erfahrung betrifft, ihre veränderung bewirken.

jene **modelle,** die wir automaten nennen, stehen als abstrakte wie empirische objekte (d. h. als mathematische theorie wie als computer), als wissenschaft vom künstlichen zum studium der natürlichen welt wie als artefakt unseres organischen gehirns zum studium abstrakter systeme, in dem oben skizzierten zusammenhang, allerdings auf einem niveau, das die zäsur einer epistemologischen differenz von evolutionärem rang durchgegangen ist.

ramon LULL (geb. 1232) war einer der ersten, der ein mechanisches verfahren als hilfe der beweisführung konstruieren wollte. doch er fand nur die kombinatorische permutation. mit ihrer hilfe konnte er z. b. über 4000 fragen von der art stellen wie: kann der allmächtige und allwissende gott dem gläubigen die freie wahl zwischen himmel und hölle lassen? wie bewegen sich engel in einem augenblick von einem platz zum andern? wie sprechen engel zueinander? etc.

lord STANHOPE erfand den „demonstrator", die erste mechanische vorrichtung für die lösung von problemen der formalen logik (1879).

die logische maschine von JEVON (geb. 1835) verwendete wahrscheinlich erstmals auf der grundlage von BOOLEs algebraischer logik die wahrheitstafeln für die problemlösung.

SHANNON (1938) benützte die korrespondenz von elektrischen netzwerken und dem aussagenkalkül zur konstruktion und vereinfachung von schaltungen. mit ihm beginnt vielleicht am sichtbarsten jene differenz, die die automatentheorie von früheren versuchen trennt. nicht nur, daß durch die einführung der tabularischen wahrheitswerte eine deduktive kraft und keine verbalen hirngespinste die symbolmanipulationen bestimmen, also die unsinnigkeit der vorher erwähnten fragen klar ist. wesentlich ist, daß die **automaten nur scheinbar bloß mechanische** vorrichtungen sind, in wahrheit jedoch physikalische implementierungen geistiger strukturen. als von mathematischen theorien erzeugte physikalische objekte sind sie weder rein subjektive noch objektive sinnesempfindungen (abstrakte objekte, die existieren oder nicht), sondern mögen einen aspekt der objektiven realität darstellen, der allerdings zwischen uns und der realität eine andere beziehung als die der sinnesempfindungen errichtet.

in jenem bereich von philosophie und mathematik, wo die logik als quelle der wahrheit denunziert, anderseits die mathematische theorie so verfeinert wird, daß ihre inneren untersuchungen erhellende und befreiende konsequenzen auch für nichtmathematische bereiche des denkens haben, stellt sich die frage nach der objektiven existenz der mathematischen objekte und ihrer beschaffenheit als „genaue replica der frage nach der objektiven existenz der äußeren welt" und ihrer beschaffenheit (GÖDEL, 1947). automatentheorie verändert unsere auffassung von der wirklichkeit in dem maße, wie einst die mikrophysik und quantenphysik, die z. b. verschiedene erklärungen des universums als zugleich richtig erlaubte.

die automatentheorie zeigt die wirklichkeit als produkt unserer modelle. sie ist von jenem geist, mit dem gödel selbst sein unvollständigkeitstheorem interpretierte, das allen finiten bestrebungen der mathematik ein ende setzte, „entweder gibt es mathematische objekte außerhalb von uns oder sie sind unsere eigenen konstruktionen und der geist ist nicht mechanistisch" (vortrag von gödel, brown university, ende 1940) und sich dann für die „nicht-mechanistische tätigkeit des geistes" entschied.

als allgemeine theorie der informationsverarbeitung klärt die automatentheorie das verhältnis von epistemologie und ontologie (was man wissen kann und was da ist). z. b. kann die frage des a priori wissens, des synthetischen wissens etc. präzis als relation der inputinformation und outputinformation untersucht werden. MOORE stellt zum beispiel die grenzen der modellbildung fest. werden durch experimente modelle in tatsachen überführt? nein. was verifiziert man mit einem experiment? das modell, die interpretation. dadurch sinkt die erklärung der natur zur beschreibung ab.

als selbständige theoretische richtung formierte sich die automatentheorie seit etwa 1954 und für eine breitere öffentlichkeit seit 1956 mit dem erscheinen von „automata studies" („Studien zur Theorie der Automaten", hrg. Shannon/McCarthy. Verlag Rogner & Bernhard, München 1973, übersetzt von Kaltenbeck/Weibel). sie ist nicht der kybernetik gleichzusetzen, obwohl das nicht heißt, daß es nicht eine beide umfassende theorie geben könnte. die kybernetik beruht auf statistik und wahrscheinlichkeit, die automatentheorie hingegen auf logik und diskreter mathematik.

gemäß den drei in der praxis vorherrschenden typen gibt es maschinen, die information (folgen von zeichen, worten etc.) annehmen, also erkennen, maschinen, die information annehmen, umwandeln und wieder ausgeben, und maschinen die information nur ausgeben und erzeugen. mit diesen drei grundtypen von automaten, auch akzeptoren, transduktoren und generatoren genannt, sind der reihe nach fragen der entscheidbarkeit, berechenbarkeit und aufzählbarkeit verknüpft.

der erste automat in diesem sinn entstand durch eine präzisierung des begriffs der berechenbarkeit: die TURINGmaschinen (1936) mit einem unendlichen band. zur unterscheidung von den turing-maschinen nennt man die anderen automaten endliche automaten, mitunter auch sequentielle maschinen, endliche zustands-automaten etc. viele weitere unterscheidungen von automatenklassen sind möglich.

die elementarste klasse von automaten ist der endliche automat (der ein transduktor ist). sein mathematisches modell kann man als quintupel definieren: $A = (S, X, Y, f, g)$. S ist die endliche menge der zustände, X die endliche menge der inputs, Y die endliche menge des outputs. g ist die zustandsfunktion, f die outputfunktion.

das verhalten von automaten betrifft die beobachtbaren phänomene, wenn der automat in einer black box eingeschlossen ist, und man nur input und output wahrnehmen kann. das innere verhalten wird schematisiert als zustand. die struktur eines automaten verkörpert die prinzipien der konstruktion eines automaten aus seinen teilen.

nachstehend wird nun eine der folgenreichsten und erfinderischsten abhandlungen zur theorie der automaten (abgedruckt in „automata studies", 1956) in auszügen wiedergegeben. MOORE hat darin mit wenig formalismen und großer gedankenkraft den begriff des endlichen automaten begründet, der eine theoretische vielfalt sondergleichen entstehen ließ.

MOORE:
diese arbeit betrachtet endliche automaten vom experimentellen gesichtspunkt aus. das heißt nicht, daß sie von resultaten irgendwelcher experimente an wirklichen physikalischen objekten berichtet, sondern sie befaßt sich vielmehr mit der frage, welche art von schlüssen aus äußeren experimenten über die inneren zustände einer endlichen maschine gezogen werden können. um die begriffliche natur dieser experimente zu betonen, wurde vom physiker das wort „gedanken-experimente" für den titel ausgeliehen.

zwei arten von experimenten werden in dieser arbeit betrachtet. die erste art, die der sogenannten einfachen experimente, wird in abbildung 1 beschrieben.

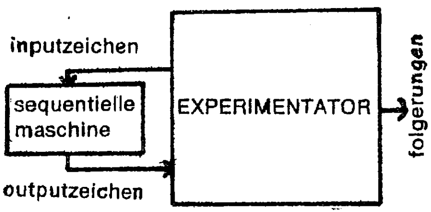

ABBILDUNG 1. schematische darstellung eines einfachen experiments

ein exemplar der sequentiellen maschine, das experimentell beobachtet wird, wird nacheinander gewisse inputzeichen vom experimentator erhalten.

der experimentator wird bestimmen, welche endliche inputzeichenfolge er in die maschine gibt, ob entweder eine feste folge oder eine, in der jedes zeichen von den vorigen outputzeichen abhängt. diese inputzeichenfolge in kombination mit der outputzeichenfolge werden wir das ergebnis des experiments nennen.

die zweite in dieser arbeit betrachtete art von experimenten ist die des in abbildung 2 gezeigten mehrfachen experiments.

in diesem fall hat der experimentator zugang zu mehreren exemplaren der gleichen maschine, von denen jedes anfangs im gleichen zustand ist. der experimentator kann jedem dieser K exemplare eine andere inputfolge senden und von jedem die entsprechende outputfolge empfangen.

in jedem dieser zwei arten von experimenten kann man sich den experimentator als einen menschen oder eine andere maschine vorstellen, der versucht, antworten auf fragen über die beschaffenheit der maschine oder ihres ursprünglichen zustands zu erhalten.

es mag lehrreich sind, einige situationen zu betrachten, in denen diese art von theorie als mathematisches modell dienen könnte.

wir erlegen der tätigkeit des experimentators eine etwas künstliche einschränkung auf. es ist ihm nicht erlaubt, die maschine aufzumachen und deren teile zu betrachten, um nachzusehen, welcher art sie sind und wie sie miteinander verbunden sind. die betrachteten maschinen sind immer nur sogenannte „schwarze kästen" (black boxes), die nur in ausdrücken ihrer inputs und outputs beschrieben werden, doch aus denen keine information bezüglich der inneren konstruktion gewonnen werden kann.

wiederum eine andere situation, wofür diese theorie ein mathematisches modell ist, tritt in dem fall ein, wenn ein psychiater an einem patienten versuche anstellt. er gibt dem patienten (hauptsächlich verbale) inputs und notiert die (ebenso hauptsächlich verbalen) outputs, die er dazu verwendet, herauszufinden, was mit dem patienten los ist. die black box-einschränkung entspricht ungefähr der unterscheidung zwischen dem psychiater und dem gehirnchirurgen.

schließlich gibt es noch eine situation, deren mathematisches modell dies möglicherweise ist: wenn ein wissenschaftler irgendeiner art ein experiment durchführt. in der physik, chemie, oder in beinahe jeder anderen wissenschaft entsprechen die inputs, die ein experimentator in seinen versuch eingibt, und die outputs, die daraus resultieren, nicht genau den dingen, die der experimentator durch die ausführung dieses experiments zu erfahren wünscht. wegen der durch die unbeugsamen naturgesetze auferlegten einschränkungen wird der experimentator häufig gezwungen, seine fragen in indirekter form zu stellen. diese einschränkungen sind in ihrer wirkung auf die organisation des experiments einigermaßen der black box-einschränkung vergleichbar.

der zustand, in dem die maschine zu einer gegebenen zeit ist, hängt nur von seinem zustand zur vorhergehenden zeit und vom vorhergehenden inputzeichen ab. das outputzeichen zu einer gegebenen zeit hängt nur vom laufenden zustand der maschine ab. eine tabelle, die diese zustandsübergänge und outputs angibt, wird als die definition einer maschine verwendet. um diese konventionen zu veranschaulichen, betrachten wir das folgende beispiel einer maschine:

maschine a

gegenwärtiger zustand

voriger zustand	voriger input 0	voriger input 1
q_1	q_4	q_3
q_2	q_1	q_3
q_3	q_4	q_4
q_4	q_2	q_2

gegenwärtiger zustand	gegenwärtiger output
q_1	0
q_2	0
q_3	0
q_4	1

diese zwei tabellen geben die vollständige definition einer maschine. in der oberen tabelle wird der gegenwärtige zustand der maschine als eine funktion des vorherigen zustands und des vorherigen inputs angegeben. in der unteren tabelle wird der gegenwärtige output der maschine als eine funktion des gegenwärtigen zustands angegeben.

ein experiment kann an dieser maschine vollzogen werden, indem ihr eine spezielle inputfolge eingegeben wird. zum beispiel die folge 000100010.

wenn die maschine anfangs im zustand q_1 ist, dann würde das ergebnis dieses experiments sein:

0	0	0	1	0	0	0	1	0
q_1	q_4	q_2	q_1	q_3	q_4	q_2	q_1	q_3
0	1	0	0	0	1	0	0	0

die erste zeile ist die folge von inputs, die zweite zeile die folge von zuständen, und die dritte die folge von outputs. die letzten zwei zeilen können aus der ersten durch den gebrauch der tabularischen definition der maschine a oder ihres zustandsdiagramms erhalten werden. es soll betont werden, daß nur die unterste der drei zeilen durch den experimentator beobachtbar ist, und daß die zustandsfolge versteckt ist, und nur gebraucht wird, um zu den beobachtbaren resultaten des experiments zu kommen oder diese zu erkären.

man nehme an, daß dieselbe oben erwähnte inputfolge einer maschine a eingegeben wird, die einen anderen anfangszustand hat. das ergebnis dieses experiments wäre, je nachdem ob der anfangszustand q_2, q_3 oder q_4 ist, eines der folgenden:

0	0	0	1	0	0	0	1	0
q_2	q_1	q_4	q_2	q_3	q_4	q_2	q_1	q_3
0	0	1	0	0	1	0	0	0

0	0	0	1	0	0	0	1	0
q_3	q_4	q_1	q_3	q_4	q_2	q_1	q_3	
0	1	0	0	0	1	0	0	0

0	0	0	1	0	0	0	1	0
q_4	q_2	q_1	q_4	q_2	q_1	q_4	q_2	q_3
1	0	0	1	0	0	1	0	0

obwohl in diesem beispiel für ein experiment der maschine eine vorher bestimmte inputzeichenfolge eingegeben wurde, sollte nicht angenommen werden, daß dies die einzige zulässige art von experiment ist. im allgemeinen können die inputs in die maschine von ihren vorigen outputs abhängen, wodurch der verlauf des experiments sich verzweigen kann.

obwohl ein verzweigungsexperiment der allgemeinste typ eines deterministischen experiments ist, können die meisten experimente, die für die beweise dieser arbeit erforderlich sind, einfache folgen sein. zum beispiel ist das kürzeste einfache experiment, das verwendet werden kann, um zwischen zwei zuständen (derselben oder einer anderen maschine) zu unterscheiden, einfach eine folge.

ein zustand q_i einer maschine s heißt ununterscheidbar von einem zustand q_j von s_j genau dann, wenn jedes experiment, das an s im zustand q_i beginnend durchgeführt wird, dasselbe ergebnis produziert, wie wenn es im zustand q_j begänne.

ein paar von zuständen gilt als unterscheidbar, wenn sie nicht ununterscheidbar sind. daher ist q_i unterscheidbar von q_j genau dann, wenn es ein experiment gibt, dessen ergebnis davon abhängt, in welchem dieser zwei zustände s zu beginn des experiments war.

die maschinen s und t werden ununterscheidbar genannt, genau dann, wenn sie nicht unterscheidbar sind, d. h. mit anderen worten, wenn die beiden folgenden bedingungen gelten:

1. Zu jedem Zustand q_i von s und zu jedem experiment gibt es einen zustand q_j von t derart, daß das experiment beginnend mit s im zustand q_i das gleiche resultat liefert, wie wenn es mit t im zustand q_j begonnen wird.

2. zu jedem zustand q_j von t und zu jedem experiment gibt es einen zustand q_i von s derart, daß das experiment beginnend mit t im zustand q_j das gleiche ergebnis liefert, wie wenn es mit s im zustand q_i begonnen wird.

wenn s von t ununterscheidbar ist, dann sind die zwei maschinen in ihrem verhalten gleich (obgleich sie in ihrer struktur differieren können), und können als austauschbar angesehen werden. in der praktischen anwendung wirklicher maschinen kann der hersteller aus dieser äquivalenz vorteile ziehen und diejenige von den zwei maschinen produzieren, die billiger zu bauen ist, leichter zu reparieren oder irgendeine andere wünschenswerte innere eigenschaft hat.

das erste theorem, das wir beweisen, betrifft eine interessante qualitative eigenschaft der maschinen.

THEOREM 1: es gibt eine maschine, so daß je zwei ihrer zustände unterscheidbar sind, aber es gibt kein einfaches experiment, das bestimmen kann, in welchem zustand die maschine zu beginn des experiments war.

die maschine a, die wir bereits früher beschrieben haben, erfüllt die bedingungen dieses theorems. das vorher beschriebene experiment wird zwischen jedem zustandspaar außer dem paar (q_1, q_3) unterscheiden. das heißt: sei irgendein anderes zustandspaar gegeben, so wird, wenn man weiß, daß die maschine zu beginn des experiments in einem zustand dieses paars war, die anwendung dieses experiments einen output ergeben, der von dem zustand abhängt, in dem die maschine war. um zwischen q_1 und q_3 unterscheiden zu können, sollte das experiment darin bestehen, die folge 11 anzuwenden. das ergebnis ist:

1	1	1	1
q_1	q_3	q_3	q_4
0	0	0	1

Daher gibt es ein einfaches experiment, das zwischen jedem zustandspaar unterscheiden kann. weiters kann das mehrfache experiment, das zwei exemplare der maschine gebraucht, wobei es eine der zwei vorher erwähnten folgen zu jeder schickt, genug information erhalten, um vollständig zu bestimmen, in welchem zustand die maschine zu beginn des experiments war.

um den beweis zu vervollständigen, muß nur gezeigt werden, daß, wenn nur ein exemplar der maschine a gegeben ist, es kein experiment gibt, das bestimmen kann, ob sie zu beginn des experiments im zustand q_1 war.

es ist klar, daß jedes experiment zwischen q_1 und q_4 unterscheiden wird, da das erste outputzeichen verschieden ist. doch jedes einfache experiment, das zwischen q_1 und q_2 unterscheidet, kann q_1 nicht von q_3 unterscheiden. um dies zu sehen, halte man fest, daß jedes mit dem input 1 beginnende experiment es nicht ermöglicht, q_1 von q_2 zu unterscheiden (da in beiden fällen der erste output 0 und der zweite zustand q_3 ist, so daß keine künftigen inputs verschiedene outputs erzeugen können). ähnlich ermöglicht es kein experiment, das mit dem input 0 beginnt, q_1 von q_3 zu unterscheiden.

dieses resultat kann man als ein diskretwertiges analogon zum unbestimmtheitsprinzip von heisenberg ansehen. um die parallele zu verdeutlichen, werden sowohl das unbestimmtheitsprinzip als auch dieses theorem in ähnlicher sprache neu formuliert.

der zustand eines elektrons e wird als bestimmt angenommen, wenn sowohl seine geschwindigkeit als auch seine position bekannt sind. es können experimente ausgeführt werden, die folgende zwei fragen beantworten:

1. welche position hatte e zu beginn des experiments?

2. welche geschwindigkeit hatte e zu beginn des experiments?

im fall der maschine a können experimente ausgeführt werden, die folgende zwei fragen beantworten:

1. war a zu beginn des experiments im zustand q_2?

2. war a zu beginn des experiments im zustand q_3?

in beiden fällen ändert die durchführung des experiments, das frage 1 beantwortet, den zustand des systems, so daß die antwort auf frage 2 nicht erhalten werden kann. mit anderen worten: es ist nur möglich, eine teilinformation über die frühere geschichte des systems zu gewinnen, da die ausführung von experimenten bewirkt, daß das system seine vergangenheit „vergißt".

können wir aufgrund der analogie zum unschärfeprinzip auch behaupten, daß der künftige zustand der maschine a aus vergangenen experimentellen resultaten nicht vorausgesagt werden kann? hier endet die analogie. obwohl wir experimentell nicht feststellen können, in welchem zustand sich die maschine a zu beginn des experiments befand, können wir dennoch feststellen, in welchem zustand sie am ende des experiments ist. in der tat, am ende des zuerst beschriebenen experiments wird die maschine a in einem besonders vorbestimmten zustand (unabhängig von ihrem anfangszustand) sein, nämlich im zustand q_3.

trotz der unvollständigkeit der analogie scheint es interessant, daß es in diesem diskreten, deterministischen system eine entsprechung zur unschärfe-relation gibt. irgendwelche anwendungen dieses beispiels auf kausalität, freien willen oder andere metaphysische probleme werden dem leser überlassen.

THEOREM 2: es sei eine maschine s gegeben und ein an s ausgeführtes mehrfaches experiment. dann gibt es andere von s experimentell unterscheidbare maschinen, für die das ursprüngliche experiment dasselbe ergebnis gehabt haben würde.

dieses resultat bedeutet, daß es niemals möglich sein wird, experimente an einer vollständig unbekannten maschine durchzuführen, die zu ihrer identifizierung innerhalb der klasse aller sequentiellen maschinen genügten. wenn wir jedoch die klasse auf eine kleinere einschränken, mag dies möglich sein.

die bis jetzt bewiesenen theoreme waren hauptsächlich mit qualitativen fragen befaßt, z.b.: ob es möglich oder unmöglich ist, experimente auszuführen, die fragen über den angeblichen zustand einer maschine oder ihre innere struktur beantworten. die restlichen theoreme werden sich mit der frage befassen, wieviele schritte diese experimente benötigen, und ihre beweise werden methoden für den entwurf solcher experimente miteinschließen (z. b. eine methode zur auffindung des kürzesten experiments, das zwischen zwei zuständen unterscheidet).

es gibt viele probleme bezüglich der umgekehrten frage der zerlegung einer maschine in bestandteile. es sei eine maschine mit n zuständen gegeben. unter welchen bedingungen kann sie als eine kombination von zwei maschinen dargestellt werden, die n_1 und n_2 zustände haben, so daß $n_1 n_2 = n$? unter welchen bedingungen ist die zerlegung eindeutig?

eine art beschreibung der tätigkeit der ingenieure, wenn sie wirkliche automaten konstruieren, ist die, daß sie mit einer umfassenden beschreibung einer maschine beginnen und sie nacheinander in kleinere und kleinere maschinen zerlegen, bis schließlich die einzelnen relais oder vakuumröhren erreicht sind. die wirksamkeit einer solchen methode mag durch eine theoretische untersuchung über solche zerlegungen bestimmt werden. dies mag auch ein licht werfen auf die berechtigung, mit der psychiater hoffen können, den geist in ich, überich, es etc. zu unterteilen.

Pumpwerk mit Wasserradantrieb. London Bridge 1749.