Mechanisierung mathematischer Erfahrung oder: Turings Problem

by

Wilfried Sieg

February 1992

Report CMU-PHIL-24

Carnegie Mellon

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

Mechanisierung mathematischer Erfahrung
oder: Turings Problem

Wilfried Sieg
Department of Philosophy
Carnegie Mellon University
Pittsburgh

1. Der zentrale Begriff. Der Begriff der *effektiven Berechenbarkeit* ist zentral für die Logik und die Philosophie der Mathematik. Erkenntnistheoretische Überlegungen, die sich zumindest bis auf Leibniz zurückverfolgen lassen, führten in den ersten drei Jahrzenten unseres Jahrhunderts zu einer Analyse dieses Begriffs und, in den Händen von Church und Turing, 1936 zu äquivalenten mathematischen Präzisierungen. Die Theorie, die sich um diese Präzisierungen bildete, gab entscheidende praktische Impulse für die Architektur moderner digitaler Computer und stellte die begrifflichen Grundlagen für die Informatik, die Künstliche Intelligenz und die Kognitive Psychologie bereit. Hier haben wir ein neueres Beispiel aus der Geschichte der Wissenschaften, in dem Entwicklungen der Mathematik - stark philosophisch motiviert - ihren "Anwendungen" um Jahrzehnte in prästabilierter Harmonie voraus sind.

1.1. Hintergrund. Die Problematik erwuchs innerhalb zweier Traditionen der Logik und Mathematik, die nach angemessenen symbolischen Darstellungen von Problemen und deren algorithmischer Lösung suchten. Diese Traditionen trafen in Leibniz zusammen. Er sah algorithmische Lösungen mathematischer und logischer Probleme als Paradigmata der Lösung von Problemen schlechthin. Erinnern Sie sich bitte daran, daß er den Disputanten in einer Kontroverse in einem *beliebigen* Gebiet empfahl, sich an einen Tisch zu setzen, eine Feder in die Hand zu nehmen, und zu sagen *Calculemus!* Diese Empfehlung war auf seinen Hoffnungen für eine *lingua philosophica* und einen *calculus ratiocinator* begründet. Aber erst die *Begriffsschrift* Freges realisierte einen Teil der Leibnizschen Hoffnungen, indem sie die "Formalisierung" mathematischer Beweise erlaubte. Sie stellte nicht nur eine reiche Sprache mit Quantoren und Relationen zur Verfügung, sondern verlangte auch, daß alle Annahmen in Beweisen explizit aufgeführt werden und daß alle Beweisschritte innerhalb eines logischen Kalküls vollzogen werden. Frege betrachtete die letzte Forderung, korrekterweise, als eine Verschärfung der axiomatischen Methode, die auf die Euklidschen *Elemente* zurückgeht. Mit dieser Verschärfung der axiomatischen Methode verfolgte Frege das Ziel, die erkenntnistheoretische Natur von Theoremen zu erkennen, die durch eine Folge von Schlüssen mit relevanten Axiomen verkettet sind.[1]

Durch die Lückenlosigkeit der Schlussketten wird erreicht, daß jedes Axiom, jede Voraussetzung, Hypothese, oder wie man es sonst nennen will, auf denen ein Beweis beruht, ans Licht gezogen wird; und so gewinnt man eine Grundlage für die Beurtheilung der erkenntnistheoretischen Natur des bewiesenen Gesetzes.

Das konnte aber nur erfolgreich durchgeführt werden, so erkannte Frege wie Leibniz vor ihm, wenn in die Schlußschritte kein inhaltliches Wissen einfließt: sie müssen auf Grund der Form der in ihnen vorkommenden Sätze als korrekt erkannt werden. Und tatsächlich konstatierte Frege, daß in seinem logischen System "das Schließen wie eine Rechnung durchgeführt wird".[2] Gödel berief sich 1933 auf Frege und Peano als er auf "das hervorstechende Merkmal der Schlußregeln" in einem formalen mathematischen System hinwies: "sie [die Regeln] sind rein formal, d.h. sie beziehen sich nur auf die äußere Struktur der Formeln, nicht auf deren Bedeutung, so daß sie von jemandem, der nichts von der Mathematik versteht, oder von einer Maschine angewendet werden können."

1.2. Fragestellungen. Die Frage, was denn "rein formale" Regeln seien, ist direkt verknüpft mit der mathematischen Frage *Was ist ein Algorithmus?*. Die äquivalenten Antworten, die auf diese mathematische Frage 1936 gegeben wurden, haben seit den Fünfziger Jahren fundamentale Bedeutung für die noch immer philosophische Frage

---

[1] Das spezifisch Fregesche Ziel, den analytischen Charakter der Mathematik zu erweisen, soll uns hier nicht weiter beschäftigen. - Das folgende Zitat findet man in *Grundgesetze*, p.vii.

[2] [1984], p. 237. In anderen Schriften [1969], p. 39, betonte Frege, "das ganze [Denken] kann nie durch eine Maschine besorgt oder durch eine rein mechanische Tätigkeit ersetzt werden. Wohl lässt sich der Syllogismus in die Form einer Rechnung bringen, die freilich auch nicht ohne Denken vollzogen werden kann, aber doch durch die wenigen festen und anschaulichen Formen, in denen sie sich bewegt, eine grosse Sicherheit gewährt."

Sind wir (reduzierbar auf) Maschinen? erlangt. Denn einerseits benutzt die überzeugendste Antwort auf die mathematische Frage den Begriff der Turing-Maschine, und andererseits stellen Turing-Maschinen (oder äquivalente Begriffe) einen präzisen Rahmen bereit, in dem Modelle kognitiver Prozesse formuliert werden können. Einige für unser Jahrhundert als charakteristisch angesehene wissenschaftliche "Revolutionen" haben in grundlegender Weise diesen Rahmen genutzt: zum Beispiel die Linguistik (hier weise ich auf Chomsky hin) und die Kognitive Psychologie.

Ich will mich auf zwei Fragen konzentrieren. Die *erste Frage* ist die Hauptfrage: sie ist vorgängig und will wissen, warum und in welchem Sinne die Turingsche Charakterisierung der effektiven Berechenbarkeit korrekt ist.[3] In der Literatur ist es merkwürdigerweise umstritten, was Turing eigentlich charakterisieren wollte. Zum Beispiel behauptete Gödel, daß Turing den Begriff *menschlicher* Effektivität analysieren wollte, aber daß seine Analyse nur für *mechanische* Effektivität überzeugend ist. Die *zweite Frage* ist ein Spezialfall der kognitiv-psychologischen: sie will wissen, ob mathematische Erfahrung adäquat "mechanisiert", d.h. durch formale Systeme repräsentiert werden kann. Hilbert behauptete schon 1928 in seiner programmatisch-formalistischen Haltung, daß die Antwort "Ja" sei: die bestimmten Regeln, nach denen das Formelspiel der Mathematik vorgeht und "in denen die *Technik unseres Denkens* zum Ausdruck kommt", bilden ein "abgeschlossenes System, das sich auffinden und endgültig angeben läßt". Liest man Hilbert weiter, dann ist man versucht, die von ihm begründete Beweistheorie als einen Vorläufer der Kognitiven Psychologie zu betrachten: "Die Grundidee meiner Beweistheorie ist nichts anderes, als die Tätigkeit unseres Verstandes zu beschreiben, ein Protokoll über die Regeln aufzunehmen, nach denen unser Denken tatsächlich verfährt." -- Bevor ich diesen beiden Fragen nachgehe, will ich relevante historische Zusammenhänge skizzieren.

---

[3] Ich möchte bemerken, daß meine Überlegungen, die diese Frage betreffen, Diskussionen mit Tamburrini und (historischen) Arbeiten von Davis, Gandy und Kleene viel verdanken. - Für stilistische Verbesserungen meiner Arbeit danke ich Thomas Polzin herzlich.

2. Metamathematik. Die ausgefeilt formale Darstellung Freges verhinderte natürlich nicht die Herleitung eines Widerspruchs aus den Fregeschen Grundgesetzen -- durch Russell und Zermelo. Für Hilbert ergab sich deshalb die Frage, wie man die Nicht-Herleitbarkeit eines Widerspruchs aus einer präzis formulierten Theorie mathematisch überzeugend erweisen könne. Frege hatte die Möglichkeit, in seiner Begriffsschrift mechanisch zu schließen und einige Probleme algorithmisch zu lösen, nicht als einen der logisch wichtigen Aspekte betrachtet; einen Aspekt, den Hilbert betonte und radikalisierte: er wollte die formale Darstellung mathematischer Beweise dazu nutzen, den Gebrauch klassischer Theorien zum Nachweis elementarer arithmetischer Sätze zu rechtfertigen. Und das sollte auf finite Weise geschehen, ohne auf den (problematischen) Inhalt der Theorien Bezug zu nehmen. Die Rechtfertigung einer formalen Theorie im Sinne dieser instrumentalistischen Fassung des Hilbertschen Programms impliziert die Widerspruchsfreiheit der untersuchten Theorie.[4]

2.1. Entscheidbarkeit. Was immer als dubios angesehen werden kann bei diesem auf der Oberfläche radikal-formalistischen Programm, Hilbert hat den enormen Verdienst, mathematisch präzise Begriffe und Methoden zur Untersuchung formaler Theorien eingeführt und faszinierende Fragen gestellt zu haben; Fragen übrigens, die in modifizierter Form noch immer verfolgt werden. Die neue *metamathematische* Perspektive ist in Hilbert und Ackermanns 1928 erschienenem Buch *Grundzüge der theoretischen Logik* aufs klarste beschrieben:

> Die mathematische Logik leistet aber noch mehr als eine Verschärfung der Sprache durch die symbolische Darstellung der Schlußweisen. Nachdem einmal der logische Formalismus feststeht, kann man erwarten, daß eine systematische, sozusagen rechnerische Behandlung der logischen Formeln möglich ist, die etwa der Theorie der Gleichungen in der Algebra entspricht. ...

Insbesondere sollte ein zweites Problem, eng verknüpft mit dem der Widerspruchsfreiheit, einer *rechnerischen Behandlung* zugänglich

---

[4] Eine detaillierte Analyse des Hilbertschen Programms und seiner Entwicklung findet man in meiner Arbeit "Relative Consistency and Accessible Domains".

gemacht werden. Es ist das Entscheidungsproblem für die Prädikatenlogik, dessen klassische Formulierung für Allgemeingültigkeit, bzw. Erfüllbarkeit man ebenfalls in den *Grundzügen* findet:

Das Entscheidungsproblem ist gelöst, wenn man ein Verfahren kennt, das bei einem vorgelegten logischen Ausdruck durch endlich viele Operationen die Entscheidung über die Allgemeingültigkeit bzw. Erfüllbarkeit erlaubt.

Hilbert und Ackermann betonten die "grundsätzliche Wichtigkeit" einer Lösung dieses Problems und nannten es "das Hauptproblem der mathematischen Logik". Tatsächlich würde eine positive Lösung eine Methode bereitstellen, Unabhängigkeits- und Widerspruchsfreiheitsbeweise zu führen -- zumindest für endlich axiomatisierte Theorien. Aber diese letzte Einschränkung wurde nicht als wesentlich betrachtet.[5] Den Forschern in der Hilbert Schule war klar, daß eine positive Lösung für die Prädikatenlogik -- zusammen mit der Annahme der endlichen Axiomatisierbarkeit von Theorien und der quasi-empirischen Vollständigkeit der *Principia Mathematica* -- die Entscheidbarkeit jeder (im Rahmen der *Principia* formulierbaren) mathematischen Frage implizieren würde. Für einige war das ein hinreichender Grund, eine negative Lösung des Entscheidungsproblems zu erwarten; von Neumann, zum Beispiel, schrieb 1927:

Es scheint also, daß es keinen Weg gibt, um das allgemeine Entscheidungskriterium dafür, ob eine gegebene Normalform a beweisbar ist, aufzufinden. (*Nachweisen können wir freilich gegenwärtig nichts. Es ist auch gar kein Anhaltspunkt dafür vorhanden, wie ein solcher Unentscheidbarkeitsbeweis zu führen wäre.*) ... Und die Unentscheidbarkeit ist sogar die Conditio sine qua non dafür, daß es überhaupt einen Sinn habe, mit den heutigen heuristischen Methoden Mathematik zu treiben. An dem Tage, an dem die Unentscheidbarkeit aufhörte, würde auch die Mathematik im heutigen Sinne aufhören zu existieren; an ihre Stelle würde eine absolut mechanische Vorschrift treten, mit deren Hilfe jedermann von jeder gegebenen Aussage entscheiden könnte, ob diese bewiesen werden kann oder nicht. - Wir müssen uns also auf den Standpunkt stellen: Es ist allgemein unentscheidbar, ob eine gegebene Normalform beweisbar ist oder nicht.

Warum glaubte von Neumann, daß gar kein Anhaltspunkt vorhanden sei, wie ein Unentscheidbarkeitsbeweis geführt werden könne? Schließlich gab es wohlbekannte mathematische Beweise für die

---

[5] Herbrand, zum Beispiel, schrieb 1930: "And in general we can contrive so as to make all usual mathematical arguments in theories that have only a determinate finite number of hypotheses. Thus we can see the importance of this problem, whose solution would allow us to decide with certainty with regard to the truth of a proposition in a determinate theory."

Unmöglichkeit, gewisse geometrische Probleme mit eingeschränkten Konstruktionsmitteln zu lösen (z. B. die Würfelverdoppelung mit Zirkel und Lineal). Aber hier liegt ja das Problem: für eine negative Lösung braucht man eine Abgrenzung der zulässigen Konstruktionen, im Falle des Entscheidungsproblems also eine präzise Antwort auf die Frage "Was sind absolut mechanische Vorschriften?".

2.2. Rekursionen. Beispiele mechanischer Vorschriften lieferten die in der mathematischen Praxis seit langem verwendeten primitiv rekursiven Funktionen. Die Addition, die Multiplikation, die Folge der Primzahlen, und die Fibonacci Zahlen können alle durch "primitive Rekursion" definiert werden. Das allgemeine Schema dieser Art der Rekursion führt von primitiv rekursiven Funktionen g und h auf folgende Weise zu einer neuen Funktion f:

$$f(x_1, \ldots, x_n, 0) = g(x_1, \ldots, x_n)$$
$$f(x_1, \ldots, x_n, y') = h(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y))$$

Die Definitionsgleichungen für f können als Rechenvorschriften aufgefasst werden. Interessiert war man aber wirklich an Vorschriften für das Operieren mit endlichen (syntaktischen) Gebilden, wie arithmetischen und logischen Formeln. Gödel führte eine grundlegende Methode ein, um syntaktische Operationen auf arithmetische zurückzuführen: die Methode der zahlentheoretischen Kodierung oder der Arithmetisierung; heute sprechen wir deshalb auch von *Gödelisierung*. Tatsächlich wurden die primitiv rekursiven Funktionen in seiner klassischen Arbeit "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I" (1931) benutzt, um die (Operationen auf den) kodierten syntaktischen Strukturen einer Version der *Principia Mathematica* zu beschreiben.

Vom finiten Standpunkt aus war die Einschränkung auf primitiv rekursive Funktionen für die Beschreibung der Syntax formaler Systeme angemessen; von einer breiteren Perspektive gab es jedoch keinen Grund, warum andere effektive Verfahren ausgeschlossen werden sollten. Daß es effektive, aber nicht primitiv rekursive Verfahren gibt, war von Ackermann schon Mitte der Zwanziger Jahre bewiesen worden. Gödel bemühte sich deshalb in den 1934 gehaltenen Princeton Vorlesungen, seine Unvollständigkeitssätze so zu formulieren, daß sie weniger auf einen speziellen Formalismus Bezug nahmen. Dieses Ziel war bereits durch den Titel der Vorlesungen angedeutet: "On undecidable propositions of formal mathematical systems". In dem einleitenden ersten Paragraphen besprach Gödel den Begriff eines *formalen mathematischen Systems* und forderte:

> the rules of inference, and the definitions of meaningful formulas and axioms, [have to] be constructive; that is, for each rule of inference there shall be a finite procedure for determining whether a given formula B is an immediate consequence (by that rule) of given formulas $A_1, \ldots, A_n$, and there shall be a finite procedure for determining whether a given formula A is a meaningful formula or an axiom.

Jedoch benutzte Gödel wieder primitiv rekursive Funktionen und Relationen für die Beschreibung der Syntax. Er betrachtete die primitiv rekursive Definierbarkeit von Formeln und Beweisen als "präzise Bedingung, die in der Praxis ausreicht als Ersatz für die unpräzise Forderung von §1, daß die Klasse der Axiome und die Beziehung der unmittelbaren Folgerung konstruktiv seien".[6] Aber was benötigt wurde war wirklich ein Begriff, der nicht nur in der Praxis ausreicht, sondern in prinzipieller Hinsicht genügt: Gödel bezeichnete es als eine "wichtige Eigenschaft" der primitiv rekursiven Funktionen, daß ihr Wert für jedes Argument mit einem endlichen Verfahren berechnet werden kann, und fügte in Fußnote 3 die Bemerkung an:

> The converse seems to be true if, besides recursions according to the scheme (2) [der primitiven Rekursion], recursions of other forms (e.g., with respect to two variables simultaneously) are admitted. This cannot be proved, since the notion of finite computation is not defined, but it can serve as a heuristic principle.

Die Frage war natürlich, welche neuen Rekursionsgleichungen können hinzugenommen werden. Gödel glaubte, einem Vorschlag Herbrands zu folgen, als er im letzten Paragraphen seiner Vorlesungen einen allgemeinen Begriff der rekursiven Funktion angab.

> If ϕ denotes an unknown function, and $\psi_1, \ldots, \psi_k$ are known functions, and if the ψ's and ϕ are substituted in one another in the most general fashions and certain pairs of resulting expressions are equated, then, if the resulting set of functional equations has one and only one solution for ϕ, ϕ is a recursive function.

---

[6] Gödel, Collected Works I, p. 361.

Gödel schränkte "Herbrands Definition" in zwei Hinsichten ein. Er forderte zunächst, daß die linken Seiten der Funktionalgleichungen in einer Standardform mit ϕ als äußerstem Symbol gegeben sind. Ferner sollte gelten: für jede Menge natürlicher Zahlen $k_1, \ldots, k_l$ gibt es genau ein m, so daß $\phi(k_1, \ldots, k_l)=m$ eine abgeleitete Gleichung ist. Die Regeln, die in Ableitungen benutzt werden durften, waren von einfacher Art: Variablen dürfen durch Ziffern ersetzt werden und, wenn eine Gleichung der Form $\phi(k_1, \ldots, k_l)=m$ abgeleitet ist, darf man Vorkommen von $\phi(k_1, \ldots, k_l)$ auf der rechten Seite einer anderen abgeleiteten Gleichung durch m ersetzen.

Der wesentliche Punkt, in dem er über Herbrand hinauszugehen meinte, war für Gödel die präzise Spezifikation der mechanischen Regeln für die Ableitung von Gleichungen oder, um es angemessener auszudrücken, für die Ausführung einer Rechnung. In einem Brief an van Heijenoort vom 14. August 1964 schrieb er: "*Gerade durch die Spezifikation der Rechenregeln wurde ein mathematisch brauchbarer und fruchtbarer Begriff erhalten*".[7] Und er war sich sicher, daß Herbrand seinen Vorschlag (in einem Brief) "genau wie auf Seite 26 meiner Vorlesungsnotizen, d.h. ohne bezug auf Berechenbarkeit" formuliert hatte.[8] Gödels Erinnerung war nicht korrekt, wie wir nach der Entdeckung von Herbrands Brief durch Dawson wissen. Herbrand hatte sogar gefordert, daß die eindeutige Berechenbarkeit des Funktionswertes für jede Folge von Argumenten "intuitionistisch" bewiesen werden muß. Damit will ich nicht sagen, daß Gödels Einschätzung falsch war, sondern vielmehr betonen, daß Gödel einen noch wichtigeren Schritt genommen hatte; er hatte den Begriff der rekursiven Funktion von einem epistemologisch eingeschränkten Beweisbegriff gelöst.

3. Church 1936. Was war mit dem neuen Begriff erreicht? - Eine weite Klasse von berechenbaren Funktionen war charakterisiert, die

---

7 [van Heijenoort 1985], p. 115.

8 In einem Brief an van Heijenoort vom 23. 4. 1963, zitiert in [Herbrand 1971], p. 283. Gödel hatte Herbrands Brief nicht finden können, als er van Heijenoort schrieb; aber der Brief wurde von John W. Dawson vor einigen Jahren in Gödels Nachlass entdeckt. (Vgl. [Sieg 1992].)

alle bekannten effektiv berechenbaren Funktionen umfasste. Mit der von mir gerade zitierten Fußnote 3 in der Princeton Vorlesung schien Gödel sogar eine Art "Churchsche These" formuliert zu haben. Das war jedoch nicht der Fall; in einem teilweise in [Davis 1982] veröffentlichten Brief vom 15. Februar 1965 schrieb Gödel:

> ... it is not true that footnote 3 is a statement of Church's Thesis. The conjecture stated there only refers to the equivalence of "finite (computation) procedure" and "recursive procedure". However, I was, at the time of these lectures, not at all convinced that my concept of recursion comprises all possible recursions; ... .

3.1. Die These. Zu jener Zeit, d.h. im Jahre 1934, war Gödel auch nicht von Churchs Vorschlag überzeugt, Berechenbarkeit durch λ-Definierbarkeit zu explizieren. Er nannte diesen Vorschlag in einer Unterhaltung mit Church zu Beginn des Jahres 1934 "völlig unbefriedigend".[9] Trotz dieser nicht gerade ermutigenden Reaktion Gödels formulierte Church seine "These" -- für Gödels rekursive Funktionen -- in einem der wundervoll kurzen, gerade zehnminütigen Beiträge zum Treffen der Amerikanischen Mathematischen Gesellschaft in New York am 19. April 1935. Die Zusammenfassung seines Vortrages [Church 1935] beginnt so:

> Following a suggestion of Herbrand, but modifying it in an important respect, Gödel has proposed (in a set of lectures at Princeton, N.J., 1934) a definition of the term *recursive function*, in a very general sense. In this paper a definition of *recursive function of positive integers* which is essentially Gödel's is adopted. <u>And it is maintained that the notion of an effectively calculable function of positive integers should be identified with that of a recursive function, since other plausible definitions of effective calculability turn out to yield notions that are either equivalent to or weaker than recursiveness.</u>

In der berühmten Arbeit von 1936, "An unsolvable problem of elementary number theory", wiederholte Church seinen Vorschlag, den vagen Begriff der effektiv berechenbaren Funktion mit dem mathematischen der rekursiven Funktion zu identifizieren. Neben dem schon erwähnten quasi-empirischen Faktum, daß alle als berechenbar bekannten Funktionen rekursiv waren, betonte Church die Äquivalenz von λ-Definierbarkeit und Rekursivität. Diese Äquivalenz stärkte nach Church

[9] Church in einem Brief an Kleene vom 29. November 1935, zitiert in [Davis 1982], p. 9.

... the reasons adduced below for believing that they [diese präzisen Begriffe] constitute as general a characterization of this notion [i.e. effektive Berechenbarkeit] as is consistent with the usual intuitive understanding of it.[10]

Die tiefergehende Analyse, auf die in diesem Zitat angespielt wird, gab Church in Paragraph 7 seiner Arbeit.

Church führt dort zwei Methoden vor, die sich - wie er sagt - in natürlicher Weise anbieten, um den Berechenbarkeitsbegriff für zahlentheoretische Funktionen *informal* zu charakterisieren: die erste benutzt den Begriff des Algorithmus; die zweite den der Berechenbarkeit in einer Logik (nach festen Regeln). Dann argumentiert er, daß beide nicht zu einer allgemeineren Definiton führen. Da die Struktur der Argumente in beiden Fällen parallel ist, betrachte ich nur die zweite Methode. Church führt ein logisches System **L** ein, dessen Sprache das Gleichheitszeichen = enthält, ein Symbol { }( ) für die Anwendung einer einstelligen Funktion auf ihr Argument und Ziffern für die positiven ganzen Zahlen. Er gibt für einstellige Funktionen F dann diese Definition:

F wird *effektiv berechenbar* genannt genau dann wenn es einen Ausdruck f in **L** gibt, so daß gilt: $\{f\}(\mu)=\nu$ ist ein Theorem von **L** genau-dann-wenn $F(m)=n$; wo μ und ν Ausdrücke sind, die für die positiven ganzen Zahlen m and n stehen.

Dann behauptet Church, F ist rekursiv, wenn **L** einer Bedingung genügt, die im Wesentlichen ausdrückt, daß das Theoremprädikat von **L** rekursiv aufzählbar ist. Natürlich folgt die Behauptung (für uns) mit einer einfachen Anwendung des μ-Operators!

3.2. Kern des Arguments. Aber wie argumentiert Church für die rekursive Aufzählbarkeit des Theoremprädikats? Er beginnt damit, Bedingungen zu formulieren, denen jedes logische System genügen muß, "wenn es überhaupt den Zwecken dienen soll, für die ein System der symbolischen Logik gewöhnlich intendiert ist". Diese Bedingungen verallgemeinern, wie er explizit bemerkt, die von Gödel für ein formales mathematisches System eingeführten Forderungen (vergleiche Abschnitt 2.2): (i) jede Regel muß eine effektiv bere-

[10] [Church 1936], Fußnote 3, p. 90 in [Davis].

chenbare Operation sein, (ii) die Menge der Regeln und Axiome (wenn unendlich) muß effektiv aufzählbar sein und (iii) die Beziehung zwischen einer positiven ganzen Zahl und dem Ausdruck, der sie bezeichnet, muß effektiv bestimmbar sein.

Diese Bedingungen werden dann von Church, "mit Bezugnahme auf ein System von Gödelnummern für die Ausdrücke der Logik", so *interpretiert*: (i') jede Regel muß eine rekursive Operation sein, (ii') die Menge der Regeln und Axiome (wenn unendlich) muß rekursiv aufzählbar sein, und (iii') die Beziehung zwischen einer positiven ganzen Zahl und dem Ausdruck, der sie bezeichnet, muß rekursiv sein. Unter diesen Umständen ist das Theoremprädikat von **L** in der Tat rekursiv aufzählbar. Aber da für die *Interpretation* der Bedingungen nicht argumentiert wird, hängt der letzte Schritt von der These ab, die gerade etabliert werden soll! Diese Anwendung der These ist nicht direkt zirkelhaft wie es auf den ersten Blick scheinen mag: der Begriff der effektiven Berechenbarkeit wird ja auf den der Ableitbarkeit in einer Logik zurückgeführt, und dann wird der Gedanke, daß in diesem logischen Formalismus mit effektiven Regeln operiert wird, durch die Bedingungen (i') bis (iii') verschärft. Jedoch ist die Verschärfung sicherlich ein problematischer Stolperstein für die Churchsche Analyse.

4. Absoluter Stolperstein? Der Begriff, den Church in seinem Argument betrachtete, ist ein natürlicher und fruchtbarer. Er entspricht dem Begriff der Entscheidungsdefinitheit für Relationen und Klassen, der in [Gödel 1931] eingeführt worden war. In der Abhandlung "Über die Länge von Beweisen" definierte Gödel 1936 die Berechenbarkeit einer Funktion in einem formalen System S (wie Church) und betonte in einer bei der Drucklegung hinzugefügten Bemerkung die Absolutheit des so definierten Begriffs:

Es läßt sich übrigens zeigen, daß eine Funktion, die in einem der Systeme S_i oder auch in einem System transfiniter Stufe berechenbar ist, auch schon in S_1 berechenbar ist, so daß also der Begriff "berechenbar" in gewissem Sinn "absolut" ist, während alle sonst bekannten metamathematischen Begriffe (z.B. beweisbar, definierbar etc.) sehr wesentlich vom zu Grunde gelegten System abhängen.

Obwohl hier nur Absolutheit relativ zu den *formalen* Systemen S_i behauptet wird, war diese Beobachtung doch für Gödel von großer Bedeutung. In einem Brief an Kreisel schrieb er 1968: "Daß meine [Unvollständigkeits] Resultate für alle möglichen formalen Systeme gültig waren, begann mir erst plausibel zu werden ... wegen der *Bemerkung* auf Seite 83 in 'The Undecidable' ... Aber ich wurde erst durch Turings Arbeit völlig überzeugt".[11] In seinem Beitrag zur Princeton Bicentennial Conference betonte Gödel 1946 noch einmal diese Absolutheit des Berechenbarkeitsbegriffs:

> Tarski has stressed in his lecture (and I think justly) the great importance of the concept of general recursiveness (or Turing's computability). It seems to me that this importance is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e. one not depending on the formalism chosen.[12]

1965 fügte Gödel dieser Feststellung eine Fußnote hinzu und erklärte Absolutheit oder Unabhängigkeit von einem bestimmten Formalismus etwas allgemeiner als in der früheren Abhandlung: "Um präziser zu sein: eine Funktion auf den ganzen Zahlen ist berechenbar in jedem formalen System, das die Arithmetik enthält, genau dann wenn sie in der Arithmetik berechenbar ist; und eine Funktion f heißt berechenbar in S, wenn es in S einen berechenbaren Term gibt, der f repräsentiert." Wie in der früheren Arbeit ist Absolutheit also relativ zu Formalismen (für Erweiterungen) der Arithmetik erklärt, und das angeführte Faktum ist kaum hilfreich, den *allgemeinen* Begriff eines formalen Systems zu fassen!

Die Idee, daß Rechnungen in erweiterten Formalismen der Arithmetik durchgeführt werden, daß aber der spezifische Charakter der zugrunde liegenden Kalküle keine wesentliche Rolle spielt, haben Hilbert und Bernays 1939 im zweiten Band ihrer *Grundlagen der Mathematik* auf die befriedigendste Weise analysiert. In Supplement II, "Eine Präzisierung des Begriffs der berechenbaren Funktion und der Satz von CHURCH über das Entscheidungsproblem", definieren sie

---

[11] In [Odifreddi 1990], p. 65. "Bemerkung" in 'The Undecidable' [Davis 1965] bezieht sich auf die gerade zitierte Bemerkung über Absolutheit.

[12] In [Davis 1965], p. 84.

eine Funktion f als *regelrecht auswertbar*, wenn es *irgendeinen deduktiven Formalismus* S gibt, so daß f in S berechenbar ist. Die deduktiven Formalismen, die in Betracht gezogen werden können, müssen drei "Rekursivitätsbedingungen" genügen. Die wesentliche Bedingung fordert, daß die Theoreme von S durch eine primitiv rekursive Funktion aufgezählt werden können oder, was äquivalent ist, daß das Beweisprädikat von S primitiv rekursiv ist. Dann wird bewiesen, (i) daß die regelrecht auswertbaren Funktionen alle in einem engen zahlentheoretischen Formalismus (Z^0) berechnet werden können, und (ii) daß die in (Z^0) berechenbaren Funktionen genau die rekursiven sind. Diese meines Erachtens viel zu wenig beachtete Analyse ist ein wirklicher Abschluß der kurzen Entwicklung von Entscheidungsdefinitheit zu "absoluter" Berechenbarkeit; aber *sie räumt den Stolperstein, den wir in der Untersuchung von Church und den Bemerkungen Gödels bemerkt haben, nicht aus dem Wege, sondern legt ihn nur klar zu Tage.* -- Turing nimmt die den Rechnungen in einem Kalkül unterliegenden Prozesse als Ausgangspunkt seiner Analyse; ihm gelingt es, den Stolperstein zu überspringen. Wie, das werden wir jetzt nachvollziehen.

5. Turing 1936. Der Begriff der Maschinen-Berechenbarkeit, den Turing einführte, wurde von Gödel als Basis für "eine präzise und unfraglich adäquate Definition des allgemeinen Begriffs eines formalen Systems" betrachtet; Church stellte von diesem Begriff fest, er habe - im Vergleich zu dem der Rekursivität oder der λ-Definierbarkeit - "den Vorteil, daß er die Identifizierung mit der Effektivität im gewöhnlichen (nicht explizit definierten) Sinne *unmittelbar evident* macht...".[13] Was unterscheidet den Turingschen Vorschlag so dramatisch von dem Churchschen? Eine Antwort darauf muß gefunden werden, denn die naive Betrachtung von Turing-Maschinen bestätigt wohl kaum die Bemerkungen, die Gödel und Church formulierten.

---

[13] Gödels Bemerkung findet man im Postscriptum zu seiner Princeton Vorlesung [Davis 1965], p. 71. - Der Vergleich wurde von Church in einer bemerkenswerten Besprechung der Turingschen Arbeit gemacht, die im J. Symbolic Logic 1937, 2(1), 42-43 veröffentlicht wurde.

5.1. Axiomatische Bedingungen. Turings Arbeit "On computable numbers, with an application to the Entscheidungsproblem" beginnt mit einer kurzen Beschreibung der in ihrem Titel angedeuteten Gegenstände, d.h. der "berechenbaren Zahlen" oder, genauer, "der reellen Zahlen, deren Dezimalentwicklung mit endlichen Mitteln berechenbar ist". Turing weist jedoch sofort darauf hin, daß die fundamentalen Probleme die gleichen sind, wenn man zum Beispiel berechenbare Funktionen einer ganzzahligen Variablen betrachtet. Was bedeutet es also, wenn eine reele Zahl die Eigenschaft hat, "mit endlichen Mitteln berechenbar zu sein"? Turing gesteht:

> This requires rather more explicit definition. No real attempt will be made to justify the definitions given until we reach §9. For the present I shall only say that the justification lies in the fact that the human memory is necessarily limited.

In Paragraph 9 argumentiert Turing, daß die Operationen seiner Maschinen alle jene einschließen, "die bei der Berechnung einer Zahl benutzt werden." Er geht dieses Problem nicht direkt an, sondern versucht die "wirkliche Frage" zu beantworten, nämlich: "*Was sind die möglichen Prozesse, die bei der Berechnung einer Zahl ausgeführt werden können?*"

Wir müssen uns ganz klar vor Augen führen, welche Situation analysiert wird. Turing stellt sich einen menschlichen Rechner, einen "Calculator" vor, der Symbole auf Papier schreibt und vorgegebenen Rechenregeln folgt.[14] Das Papier ist in Quadrate unterteilt, wie die Seiten eines Rechenheftes für Kinder. Der zwei-dimensionale Charakter dieses Rechenraumes wird nicht als wesentlich für die Rechnung betrachtet, und Turing geht direkt über zu einem eindimensionalen Band, das in Quadrate unterteilt ist. Was bestimmt die Schritte des Calculators, und welche elementaren Operationen kann er ausführen? Nach Turing wird das Verhalten des Calculators in jedem Augenblick *eindeutig* durch zwei Faktoren bestimmt: (1) die Symbolkonfiguration, die er beobachtet, und (2) seinen "state of mind", seinen "internen Zustand". Diese Forderung nenne ich die

---

14 Die im Folgenden isolierten Bedingungen für einen Calculator sind aus Turings Überlegungen extrahiert und vereinfachen seine Analyse. - Man kann auch Turings Überlegungen *direkt* für einen Calculator durchführen, der in einem zwei-dimensionalen Rechenraum operiert.

Determiniertheitsbedingung **(D)**; sie garantiert, daß Rechnungen deterministisch sind. Die für die Fixierung der "Umstände" relevanten symbolischen Strukturen müssen für den Calculator *unmittelbar* (auf einen Blick) *erkennbar* sein. Daß das Verhalten von früheren Beobachtungen abhängig sein kann, wird durch die internen Zustände ermöglicht: sie reflektieren die "Erfahrung" des Calculators.[15] Turing fordert nun, daß der Calculator zwei *Endlichkeitsbedingungen* genügen muß:

(1.1) es gibt eine feste endliche Zahl von symbolischen Strukturen, die der Calculator unmittelbar erkennen kann;
(1.2) es gibt eine feste endliche Anzahl von internen Zuständen, die in Betracht gezogen werden können.

Folglich gibt es nur endlich viele verschiedene Kombinationen von symbolischen Strukturen und internen Zuständen. Aber solche Kombinationen bestimmen nach **(D)** eindeutig das Verhalten des Calculators; d.h. <u>es kann durch eine endliche Liste von Befehlen bestimmt</u> werden. -- Turing (p. 136) will Operationen des Calculators isolieren, die so elementar sind, "daß es nicht einfach ist, sich eine weitere Unterteilung vorzustellen". Die möglichen Operationen sind auf folgende Weise eingeschränkt:

(2.1) nur Symbole auf beobachteten Quadraten können verändert werden, und die neue (durch die Veränderung erhaltene) symbolische Struktur kann unmittelbar erkannt werden;
(2.2) die Verteilung der beobachteten Quadrate kann geändert werden, und die neue symbolische Struktur kann unmittelbar erkannt werden; darüberhinaus muß jedes der neu beobachteten Quadrate innerhalb einer festen Entfernung L von einem unmittelbar vorher beobachteten Quadrat liegen.

---

[15] Turing verknüpft für seine Maschinen "state of mind" mit "memory" in §1, l.c., p. 117: "By altering its m-configuration [d.h. "state of mind"] the machine can effectively remember some of the symbols which it has 'seen' (scanned) previously." Dieser Punkt wird auch in [Kleene 1988], p. 22, betont: "A person computing is not constrained to working from just what he sees on the square he is momentarily observing. He can remember information he previously read from other squares. This memory consists in a state of mind, his mind being in a different state at a given moment of time depending on what he remembers from before."

Die beiden letzten Bedingungen implizieren, daß ein Calculator nur endlich viele elementare Operationen ausführen kann; da die Operationen mit einer Änderung des internen Zustandes einhergehen können, schließt Turing (p. 137):

> The most general single operation must therefore be taken to be one of the following: (A) A possible change (a) of symbol [wie in (2.1)] together with a possible change of state of mind. (B) A possible change (b) of observed squares [wie in (2.2)] together with a possible change of state of mind.

5.2. Theorem und These. Mit dieser Analyse der elementaren Operationen ist es nicht schwierig, die Behauptung zu beweisen, daß jede durch einen Calculator durchgeführte Rechnung von einer Turing-Maschine ausgeführt werden kann. Tatsächlich konstruiert Turing Maschinen, die die Arbeit eines Calculators direkt simulieren, und behauptet (p. 137):

> The machines just described do not differ very essentially from computing machines as defined in § 2, and corresponding to any machine of this type a computing machine can be constructed to compute the same sequence, that is to say the sequence computed by the computer [dem Calculator, in meiner Terminologie].

Damit haben wir *Turings Theorem*: Wenn eine Zahlenfolge durch einen Calculator berechnet werden kann, der den Bedingungen **(D)** und **(1.1)** - **(2.2)** genügt, so kann sie durch eine Turing-Maschine berechnet werden.

Turings Analyse untermauert die *These*, daß mechanische Verfahren genau solche sind, die von einem Calculator ausgeführt werden können, der den einschränkenden Bedingungen genügt. Aber warum soll man diese Analyse als korrekt akzeptieren? -- Offensichtlich kam Turing zu den einschränkenden Bedingungen durch die Betrachtung der Schranken unseres sinnlichen und geistigen Apparats *wenn wir mechanische Rechnungen durchführen.* Er sah den wesentlichen Punkt, wie Sie sich vielleicht erinnern, in den Schranken des Gedächtnisses; das war die entscheidende Rechtfertigung der Bedingungen **(1.1)** - **(2.2)**. Das scheint mir -- beim gegenwärtigen Stand unseres Wissens -- höchst plausibel zu sein, weil jede dieser Bedingungen durch die Beschränkung unseres sinnlichen

Apparates motiviert ist; und diese Beschränkungen können scheinbar auf solche unseres Gedächtnisses reduziert werden.[16]

Gödel behauptete in einer kurzen Notiz mit dem Titel "A philosophical error in Turing's work" (veröffentlicht im zweiten Band der *Collected Works*), daß Turing in seiner Arbeit aus dem Jahre 1936 zeigen wollte, "geistige Verfahren können nicht über mechanische Verfahren hinausgehen". Er konstatierte korrekt, daß Turings Analyse *diese* These[17] nicht untermauert. Obwohl sich Gödels Interpretation explizit auf p. 136 in [Davis 1965] bezieht, ist sie weder werkbezogen, noch historisch-systematisch schlüssig. Tatsächlich kann sie durch Turing selbst explizit widerlegt werden, der in seiner Arbeit "Systems of logic based on ordinals" schreibt:

> A function is said to be "effectively calculable" if its values can be found by some purely mechanical process. ... We may take this statement literally, understanding by a purely mechanical process one that can be carried out by a machine. It is possible to give a mathematical description, in a certain normal form, of the structures of these machines. The development of these ideas leads to the author's definition of a computable function, and to an identification of computability [mit Hilfe von Turing Maschinen] with effective calculability.

Der 1939 veröffentlichte Artikel repräsentiert Untersuchungen, die Turing für seine Dissertation unter Church während der beiden vorhergehenden Jahre in Princeton durchgeführt hatte.

6. Effektive Erkenntnis. Was den "Fehler" in Turings Arbeit angeht, so sollten wir von einem Interpretationsfehler Gödels und nicht von einem philosophischen Irrtum Turings sprechen. Aber von diesem Fehler unberührt ist die Unterscheidung, die Gödel in dieser Notiz (und auch in anderen Bemerkungen) zwischen menschlich- und maschinen-effektiver Berechenbarkeit machte; sie ist wichtig für die Interpretation der Unvollständigkeitssätze. Jene Sätze waren

---

[16] In [Sieg 1992] findet man ein ausführlicheres Argument für diese Reduzierbarkeit.

[17] Diese *These* wird auch von anderen Autoren als Turingsche These bezeichnet; zum Beispiel von Judson Webb, der in seinem Buch [1980], p. 9, schreibt: "This [i.e., Turing's thesis] is indeed a very strong thesis, for it says that *any* effective procedure whatever, using whatever 'higher cognitive processes' you like, is after all mechanizable." In Kapitel IV, p. 230, formuliert Webb (und verteidigt als Turingsche These) die Version: "Every humanly effective computation procedure can be simulated by a Turing machine."

für Gödel nicht nur als metamathematische Resultate von Interesse, sondern hatten "Implikationen für die traditionellen philosophischen Probleme über das Wesen der Mathematik". Für solche Implikationen argumentierte Gödel in seiner Gibbs Vorlesung (1951), und seine Argumente führen uns direkt zu prinzipiellen Fragen, die die Kognitiven Wissenschaften betreffen.

6.1. Gödels Disjunktion. Dank der Turingschen Analyse können die Unvollständigkeitssätze in voller Allgemeinheit für *alle* widerspruchsfreien *formalen Systeme* S formuliert werden, die ein Minimum an elementarer Arithmetik enthalten. Die Sätze drücken für Gödel aus, daß die Mathematik, wenn sie als Bestand wahrer Aussagen aufgefaßt wird, nicht durch eine *mechanische Aufzählung* ihrer Theoreme ausschöpfbar ist; d.h. keine Turing-Maschine kann alle ihre Sätze erzeugen. Das Erste Unvollständigkeitstheorem gibt ja schon einen arithmetischen Satz an, der von S unabhängig ist. Gödel betont jedoch, daß das Zweite Theorem dieses Phänomen der Unausschöpfbarkeit besonders evident macht.

For, it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertions about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.

Behauptet jemand dies, so widerspricht er sich: wenn er alle Axiome und Regeln als korrekt erkennt, so erkennt er auch die Konsistenz des Systems. Damit hat er jedoch eine *mathematische* Einsicht gewonnen, die nicht aus den Axiomen folgt.

Um die Bedeutung dieser Sachlage zu erläutern, unterscheidet Gödel zwischen "objektiver" und "subjektiver" Mathematik. Die Mathematik im objektiven Sinn besteht aus *allen* wahren mathematischen Sätzen; die im subjektiven Sinn aufgefaßte Mathematik besteht nur aus der Teilklasse der *(menschlich) beweisbaren* Sätze. Für die objektive Mathematik kann es natürlich kein umfassendes formales System geben; aber ist es nicht ausgeschlossen, daß es für die evidenten Axiome der subjektiven Mathematik ein mechanisches

Aufzählungsverfahren gibt.[18] Gäbe es ein solches Verfahren, dann wäre -- zumindest was die Mathematik anbelangt -- der menschliche Geist einer Turing-Maschine äquivalent, und es gäbe mathematische Probleme, die durch keine dem menschlichen Geist einsichtigen Beweise entschieden werden könnten. Nennen wir mit Gödel solche Probleme *absolut unentscheidbar*, so haben wir den folgenden *Hauptsatz* bewiesen: entweder sind die evidenten Axiome der Mathematik nicht mechanisch aufzählbar, oder es gibt arithmetische Probleme, die absolut unentscheidbar sind.

Die philosophischen Implikationen, die Gödel nach einem Zwischenschritt zu ziehen sucht, betreffen seinen aus veröffentlichten Schriften bekannten, aber doch obskuren Platonismus. Für meine Überlegungen ist der Zwischenschritt wichtiger; in ihm betont Gödel das philosophische Interesse des Hauptsatzes und erläutert, ohne dafür zu argumentieren, die erste Alternative durch ein "d.h.": "d.h. der menschliche Geist übertrifft (sogar innerhalb des Reiches der Reinen Mathematik) unendlich die Fähigkeiten jeder endlichen Maschine". Gödel betrachtet die zweite Alternative als falsch, da sie impliziert, daß der menschliche Geist Fragen aufwirft, die er nicht lösen kann; und das ist für ihn ein Widerspruch![19] Akzeptierte man dieses reductio ad absurdum Argument und Gödels Erläuterung der ersten Alternative, so wäre die Unmöglickeit einer Mechanisierung mathematischer Erfahrung erwiesen. -- Nach Gödel "übertrifft" der menschliche Geist jede endliche Maschine durch das Erkennen neuer evidenter Axiome. Das könnte, so spekuliert Gödel, zum Beispiel in der Mengenlehre auf systematische Weise geschehen; und deshalb sei es denkbar, daß es ein menschlich-effektives Verfahren zur Gewinnung neuer Axiome gibt, das nicht mechanisch durchführbar ist. Doch bevor wir uns dieser höchst spekulativen Vermutung (in 6.3.) noch einmal kurz zuwenden, will ich beschreiben, wie zu Beginn der Entwicklung der kognitiven Wissenschaft *für* die Mechanisierbarkeit (von Aspekten) intellektueller Erfahrung argu-

---

18 In diesem Fall können wir uns nicht sicher sein, daß alle aufgezählten Sätze korrekt sind.

19 In Hao Wangs "From Mathematics to Philosophy", London 1974, ist diese Gödelsche Überzeugung auf den Seiten 324 bis 326 detaillierter beschrieben.

mentiert worden ist; dabei beschränke ich mich, paradigmatisch, auf Arbeiten von Newell und Simon.

6.2. Simulation kognitiver Prozesse. Die Argumente für Mechanisierbarkeit waren auf der Oberfläche nicht von allgemeiner Natur: vielmehr sollte, wie Newell und Simon 1963 betonten, durch konkrete Programme demonstriert werden, daß "Mechanismen einer gewissen Art von Verhalten fähig sind, das gewöhnlich als charakteristisch für Organismen betrachtet wird" (p. 401). Von besonderem Interesse war für sie das "Verhalten beim Lösen von Problemen" (problem solving behavior), und sie stellten fest:

> The processes of problem solving, it is turning out, are the familiar processes of noticing, searching, modifying the search direction on the basis of clues, and so on. ... It looks more and more as if problem solving is accomplished through complex structures of familiar simple elements. The *growing proof* [meine Betonung] is that we can simulate problem solving in a number of situations using no more than these simple elements as the building blocks of our programs.

In Artikeln (teilweise mit Shaw) aus den Jahren 1957 bis 1963 betrachteten sie als ein Paradebeispiel die Entdeckung von Beweisen logischer Theoreme durch ein Computer Programm. Ihrer Logic Theory Machine (oder kurz LT) gelang es, einige der aussagenlogischen Sätze aus den *Principia Mathematica* zu beweisen. Daß LT daran scheiterte, andere Sätze zu beweisen, wurde nicht als Versagen von LT betrachtet; denn wesentlich war nicht logische Vollständigkeit, sondern korrekte Simulation kognitiver Prozesse, die der menschlichen Beweissuche angeblich unterliegen. Ich lasse die komplexe Fragen nach Korrektheitskriterien und Suchprinzipien beiseite, denn sie setzen voraus, daß man Antworten auf zwei fundamentalere Fragen gegeben hat: In welchem Sinne können kognitive Prozesse durch Maschinen simuliert werden, und in welchem Rahmen wird nach Lösungen eines Problems gesucht?

Um eine schlüssige Antwort auf die erste Frage geben zu können, machten Newell und Simon (1963, p. 425) die "implizite psychologische Annahme":

that the central nervous system is a symbol-processing system, that its memory is organized in terms of lists and list structures of associated symbols, and that it is capable of executing sequences of behaviors organized as hierarchical list structures.

Kognitive Prozesse werden also als symbolische Prozesse verstanden und können "deshalb" durch Programme für geeignete digitale Computer spezifiziert werden. Bevor das "deshalb" begründet wird, möchte ich betonen, daß der Begriff "symbol-processing system" oder "physical symbol system", wie er von Newell und Simon synonym verwendet wurde, wirklich zentral ist. Über diesen Begriff schrieb Newell rückblickend im Jahre 1980 (p. 136):

In my estimation, the most fundamental contribution so far of artificial intelligence and computer science to the joint enterprise of cognitive science has been the notion of *physical symbol system*. ... the hypothesis is that humans are instances of physical symbol systems, and, by virtue of this, mind enters into the physical universe.
In my own view this hypothesis sets the terms on which we search for a scientific theory of mind. What we all seek are the further specifications of physical symbol systems that constitute the human mind or that constitute systems of powerful and efficient intelligence.

Die *physical symbol systems*, kurz PSS, die kognitive Prozesse simulieren können sollen, sind digitale Computer mit einer besonderen Architektur, die in dem obigen Zitat von Newell und Simon angedeutet ist. Programme für solche Maschinen sollen es gestatten, *ganz beliebige* symbolische Prozesse zu spezifizieren; vgl. [Newell und Simon 1963], p. 366. Wenn dies korrekt ist und man die obige psychologische Annahme macht, dann können die kognitiven Prozesse von Menschen natürlich als PSS-Prozesse beschrieben werden. In einem bemerkenswerten Rückgriff auf die Überlegungen der Logiker aus den Dreißiger Jahren wird auf die These von Church Bezug genommen, um die Klasse "beliebiger symbolischer Prozesse" präzis charakterisieren und die Universalität der PSS beweisen zu können.[20] Wie schlüssig diese Argumentation für die Simulierbarkeit kognitiver Prozesse durch die PSS ist, möge der Leser selbst beurteilen; ich möchte nur betonen, daß ein - wegen der Turingschen Analyse notwendigerweise - enger Begriff kognitiver Prozesse ins Auge gefaßt wird.

---

20 In [Newell und Simon 1963], Fußnote 13 auf p. 422, und in [Newell 1980], p. 150.

Wie eng der Begriff ist, erhellt sich auch aus der Antwort auf die Frage, in welchem Rahmen kognitive Prozesse nach Lösungen von Problemen suchen. Im Falle der Beweissuche oder des Schachspielens, um auf ein anderes häufig betrachtetes Beispiel hinzuweisen, liegt ein nach mechanischen Regeln aufgebauter "Suchraum" vor, in dem Lösungen gefunden werden müssen. (Solche Suchräume werden gewöhnlich als endlich verzweigte Bäume mathematisch dargestellt.) Die Suche nach Lösungen besteht in einem heuristisch gesteuerten Durchschreiten (von einem Teil) dieses *fest vorgegebenen* Problemraums. In [Newell und Simon 1963] wird auf Seite 403 für den aufmerksamen Leser betont, daß ein wesentlicher Unterschied zwischen dem Lösen von Problemen in einem solchen Suchraum und dem Erfinden (oder Erweitern) eines Problemraumes besteht, aber auch zwischen dem Verbessern spezifischer und allgemeiner Techniken.

They [Menschen] cannot solve algebra problems without knowing the rules of algebra. To do so would be *inventing* algebra, not solving algebra problems, *two very different tasks indeed* [meine Hervorhebung]. But humans bring to a new task environment a whole kit of tools, which, when combined with knowledge of the specific area, enables them to solve problems in it. A theory that purports to be an explanation of human problem solving has to have the same property -- it has to contain some general problem-solving means, some way of acquiring knowledge of specific task environments and of improving both its general and specific techniques (learning capabilities).

Betrachtet man einen "Meta-Prozess" der Erweiterung des Suchraumes oder der Einführung neuer (allgemeiner) Techniken, so stellt sich natürlich sofort die Frage, ob ein solcher Prozess ebenfalls mechanisierbar ist. Turing machte in seiner Arbeit von 1936 (p. 139) eine relevante Vermutung über *nicht-uniforme Verfahren* zur Berechnung reeller Zahlen δ:

It is (so far as we know at present) possible that any assigned number of figures of δ can be calculated, but not by a uniform process. When sufficiently many figures of δ have been calculated, an essentially new method is necessary in order to obtain more figures.

In seiner Arbeit über "Ordinal Logics", die während der Fünfziger und Sechziger Jahre von Feferman systematisch erweitert und abgeschlossen wurde, betrachtete Turing *uniforme und effektive Erweiterungen* von (unvollständigen) Systemen der Arithmetik um relevante Gödel-Sätze. Transfinite Iteration dieses Erweiterungsprozesses entlang allen konstruktiven Ordinalzahlen liefert, so zeigte Feferman, ein vollständiges (aber nicht mehr formales) System der Arithmetik. Solch ein mechanisch-iteratives Verfahren ist für eine heuristisch motivierte, intelligible Suche nach Beweisen mathematischer Sätze nur von höchst theoretischem Interesse. Turing sah das klar und hatte schon 1947, als er mit dem Entwurf wirklicher Rechner beschäftigt war, andere "Verfahren" ins Auge gefasst: Maschinen sollten so programmiert sein, daß sie "gelegentlich, wenn ein guter Grund vorliegt", ihr eigenes Programm in unvorhergesehener Weise ändern.

What we want is a machine that can learn from experience. The possibility of letting the machine alter its own instructions provides the mechanism for this, but this of course does not get us very far.[21]

6.3. Reflektieren: Induktion & Analogie. Das bringt uns im Falle der Mathematik weiter, so argumentierte Turing [1947], p. 124, wenn wir nicht darauf bestehen, daß Maschinen "unfehlbar" sind, und wenn wir ihnen eine hinreichend gute mathematische "Ausbildung" geben:

... if a machine is expected to be infallible, it cannot also be intelligent. There are several mathematical theorems which say almost exactly that. But these theorems say nothing about how much intelligence may be displayed if a machine makes no pretence at infallibility. ... A human mathematician has always undergone an extensive training. This training may be regarded as not unlike putting instruction tables into a machine. One must therefore not expect a machine to do a very great deal of building up of instruction tables on its own. No man adds very much to the body of knowledge, why should we expect more of a machine?

In der Arbeit "Intelligent Machinery", die Turing einige Monate später schrieb, betonte er (auf Seite 23) diese Art 'kultureller Suche' und stellte sie der 'intellektuellen Suche' in einem festen Rahmen, zum Beispiel dem der *Principia Mathematica*, gegenüber:

As I have mentioned, the isolated man does not develop any intellectual power. It is necessary for him to be immersed in an environment of other men, whose techniques he absorbs during the first twenty years of his life. He may then perhaps do a little research of his own and make a very few discoveries which are passed on to other men. From this point of view the search for new techniques must be regarded as carried out by the human community as a whole, rather than by individuals.

---

21 [Turing 1947], p. 123.

Diese Turingschen Auffassungen stehen nicht im Konflikt mit denen von Newell und Simon, aber auch nicht mit denen Gödels, wenn er auf Turings 'philosophischen Fehler' und ein möglicherweise nicht-mechanisches, effektives Verfahren hinweist:

What Turing disregrads completely is the fact that *mind, in its use, is not static, but constantly developing*, i.e., that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding. There may exist systematic methods of actualizing this development, which could form part of the procedure. Therefore, although at each stage the number and the precision of the abstract terms at our disposal may be *finite*, both (and, therefore, also Turing's number of *distinguishable states of mind*) may *converge toward infinity* in the course of the application of the procedure.

Natürlich ist es nicht nur der Geist eines Individuums, der immer mehr abstrakte Begriffe schärfer erfasst, sondern es ist die gesamte menschliche Gemeinschaft, die durch neue Begriffe Suchräume erweitert und versucht, mit neuen Methoden Beweise zu erbringen. Der wirkliche Gegensatz zwischen Gödel und Turing liegt in ihren spekulativ-ontologischen Erwartungen: Gödel meinte, daß sich die Behauptung, es gebe keinen Geist unabhängig von der Materie, als Vorurteil unserer Zeit erweisen werde; Turing, denke ich, teilte diese Auffassung nicht. Aber unabhängig von diesen philosophischen Erwartungen, ist es ganz wesentlich herauszuarbeiten, in welcher Weise der menschliche Geist über den beschränkten Rahmen eines festen Suchraumes und die fixen Methoden einer bestimmten Turing-Maschine hinausgeht. Darum beschreibe ich jetzt sehr schematisch zwei Aspekte mathematischer Erfahrung, die in diesem Kontext wichtig sein könnten.

Bereiche, deren Elemente durch (deterministische) induktive Definitionen erzeugt werden, sind Ihnen allen bekannt: aus der Mathematik die natürlichen Zahlen; aus der Metamathematik die Formeln und Beweise bestimmter logischer Systeme. Deterministisch bedeutet hier nur, daß die Objekte auf genau eine Art erzeugt werden. Diese induktiven Definitionen und auch die nicht-elementaren, schon durch Brouwer eingeführten induktiven Definitionen von konstruktiven Ordinalzahlen lassen uns den Aufbau mathematischer Objekte *begreifen*. Und damit ist, so scheint mir, eine fundamentale und objektive Quelle unserer Erkenntnis mathematischer Prinzipien für die Bereiche gegeben, die von jenen Objekten konstituiert werden; denn die passenden Definitions- und Beweisprinzipien folgen dem begriffenen Aufbau direkt. Diese grundlegende Idee kann auf allgemeinere Bereiche, deren Elemente keine traditionell-konstruktiven Züge haben, ausgedehnt werden. Das trifft zu für Segmente der sogenannten kumulativen Hierarchie, die ja das intendierte Modell der Zermelo-Fraenkelschen Mengenlehre ist. Mit diesem Fall hat sich Gödel auseinandergesetzt und den Prozess der Formulierung von immer stärkeren Unendlichkeitsaxiomen[22] als ein Beispiel betrachtet, wie -- möglicherweise -- die Einführung von mehr und mehr abstrakten Begriffen Teil einer systematischen Methode ist, die nicht durch ein mechanisches Verfahren simuliert werden kann.

Diesem schematisch beschriebenen *quasi-konstruktiven Aspekt* mathematischer Erfahrung mit besonderen, uniform erzeugten Objekten steht ein *konzeptioneller Aspekt* gegenüber, der in der modernen strukturellen Auffassung der Mathematik seinen Ausdruck gefunden hat: seine Entwicklung reicht von Dedekind über Emmy Noether zu Bourbaki und der Theorie der Kategorien. Hier handelt es sich um abstrakte Begriffe, die aus der mathematischen Praxis herausdistilliert wurden, um Analogien zu präzisieren, d.h. komplexe Zusammenhänge zwischen mehreren, scheinbar getrennten Theorien zu entdecken und axiomatisch zu fassen. Auf diese Weise lehrt uns die axiomatische Methode -- wie Bourbaki es einmal ausdrückte -- "nach den tiefliegenden *gemeinsamen Gründen* einer solchen Entdeckung zu suchen, die *gemeinsamen Ideen* dieser scheinbar sehr verschiedenen Theorien zu finden .... und sie ins richtige Licht zu setzen." Axiomatisch charakterisierte Begriffe *ohne intendierte, quasi-konstruktive Modelle* wie der einer Gruppe, eines topologischen Raumes oder einer differenzierbaren Mannigfaltigkeit fallen in diese Kategorie, aber auch der eines vollständigen, geordneten

---

22 Analoge Fragen können für die finitistischen und intuitionistischen Begründungsversuche formuliert werden, d.h. für primitiv rekursive Funktionale und konstruktive Ordinalzahlen der zweiten Zahlenklasse.

Körpers. Mir scheint, daß der Begriff der mechanischen Berechenbarkeit wegen der Analyse in 5.1. auch hierher gehört.

7. Aufgaben. Die Verschärfung von axiomatischen zu formalen Theorien wurde in der Mathematik durch erkenntnistheoretische Überlegungen motiviert. Eine wichtige Überlegung forderte, daß Beweise einer Theorie auf radikal intersubjektive Weise überprüft werden sollen: eine Überprüfung darf nur Operationen (des Geistes) involvieren, die den bei arithmetischen Rechnungen benutzten ähnlich sind. Turing postulierte 1936 allgemeine, einschränkende Bedingungen für einen passenden Begriff der Berechenbarkeit. In seiner Arbeit "Intelligent Machinery" aus dem Jahre 1947 formulierte er auf Seite 21, was ich als *Kernproblem* der Kognitiven Psychologie bezeichnen möchte:

> If the untrained infant's mind is to become an intelligent one, it must acquire both discipline and initiative. So far we have been considering only discipline. ... But discipline is certainly not enough in itself to produce intelligence. That which is required in addition we call initiative. This statement will have to serve as a definition. *Our task is to discover the nature of this residue as it occurs in man, and to try and copy it in machines.*

Die Aufgabe des Kopierens oder, besser, Simulierens ist besonders schwierig, einige würden behaupten *unmöglich*, im Fall des mathematischen Denkens. So werden wir in naheliegender Weise zu den Fragen geführt: Was ist die Natur dieses Rests[23], von dem Turing spricht; d.h. was sind wesentliche Aspekte mathematischer Erfahrung? Sind sie mechanisierbar? - Ich habe versucht eine vorläufige Antwort auf die erste Frage zu geben. Was die zweite Frage angeht, so habe ich noch nicht einmal eine Vermutung, wie sie beantwortet werden wird. Könnte es sein, daß ein mathematisch hinreichend erfahrener Computer ein neues Unendlichkeitsaxiom vorschlägt oder einen überzeugenden Begriff der Berechenbarkeit für parallel arbeitende Maschinen einführt?

Aspekte mathematischer Erfahrung stellen eine äußerst wichtige Komponente des Turingschen Problems dar, und wir sollten

---

[23] Die amerikanische Künstlerin Agnes Martin sagte in anderem Kontext: "... the absolute trick in life is to find the rest."

diese Aspekte mit großem Nachdruck verfolgen: durch historische Fallstudien, theoretische Analysen, psychologisches Experimentieren und auch -- ganz im Sinne Turings -- durch Maschinensimulation. Für das letzte Projekt, Maschinensimulation, müssen wir über die Konstruktion von "Rahmen" für interaktives, Computer-unterstütztes Beweisen hinauskommen und zur Entwicklung von Programmen gelangen, die direkt und erfogreich nach menschlich einsehbaren Beweisen in komplexen mathematischen Theorien suchen: *wir haben erst sehr kleine Schritte in diese Richtung unternommen.*

Auf die Frage, was wir von solchen Untersuchungen erwarten können, möchte ich mit einem Zitat aus Brechts "Galileo", das auf die Physik Bezug nimmt, antworten:

> Eine Hauptursache der Armut der Wissenschaften ist meist eingebildeter Reichtum. Es ist nicht ihr Ziel, der unendlichen Weisheit eine Tür zu öffnen, sondern eine Grenze zu setzen dem unendlichen Irrtum.

In der Tat erwarte ich nicht, daß die heutige Kognitive Wissenschaft der unendlichen Weisheit die Tür öffnet; aber ich hoffe, daß sie eine angemessenere begriffliche Fassung unserer intellektuellen Erfahrung vorbereitet, so wie die Mechanik Galileos die Mechanik Newtons vorbereitete. Und das ist offensichtlich nicht nur von theoretischem Interesse.

BIBLIOGRAPHIE

Church, Alonzo
1935 An unsolvable problem of elementary number theory. Preliminary report (abstract); Bull. Amer. Math. Soc. 41, 332-333.
1936 An unsolvable problem of elementary number theory; Amer. J. Math. 58, 345-363; auch in [Davis 1965], 89-107.

Davis, Martin
1965 (Hrg.) *The Undecidable*; New York
1982 Why Gödel didn't have Church's Thesis; Information and Control 54, 3-24.

Feferman, Solomon
1988 Turing in the land of O(z); in [Herken], 113-147.

Frege, Gottlob
1893 *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*; Jena.
1969 *Nachgelassene Schriften*; Hermes, Kambartel, Kaulbach (Hrg.), Hamburg.
1984 *Collected papers on mathematics, logic, and philosophy*; B. McGuinness (Hrg.), Oxford.

Gandy, Robin
1980 Church's Thesis and principles for mechanisms; in: Barwise, Keisler, and Kunen (Hrg.), *The Kleene Symposium*, Amsterdam; 123-148.
1988 The confluence of ideas in 1936; in: [Herken], 55-111.

Gödel, Kurt
1933 The present situation in the foundations of mathematics; wird erscheinen in Gödels *Collected Works*, volume III.
1951 Some basic theorems on the foundations of mathematics and their implications; Gibbs Lecture, wird erscheinen in Gödels *Collected Works*, volume III.
1986 *Collected Works*, volume I, Oxford.
1990 *Collected Works*, volume II, Oxford.

Herbrand, Jacques
1971 *Logical Writings*; Warren Goldfarb (Hrg.), Cambridge (Mass.).

Herken, Rolf
1988 (Hrg.) *The Universal Turing Machine - a half-century survey*; Oxford.

Hilbert David
1928 Die Grundlagen der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 6, 65-85.

Hilbert, David und Bernays, Paul
1935 *Grundlagen der Mathematik*; vol. I, Berlin.
1939 *Grundlagen der Mathematik*; vol. II, Berlin.

Kleene, Stephen C.
1981 Origins of recursive function theory; Annals History of Computing 3, 52-66.
1988 Turing's analysis of computability, and major applications of it; in: [Herken], 17-54.

Newell, Allen
1980 Physical symbol systems; Cognitive Science 4, 135-183.

Newell, Allen, Shaw, J. Cliff und Simon, Herbert A.
1957 Empirical explorations of the logic theory machine; Proc. of the Western Joint Computer Conference, New York, 218-239.

Newell, Allen und Simon, Herbert A.
1963 Computers in psychology; in: *Handbook of Mathematical Psychology*, vol. I, R.D. Luce, R.R. Bush, E. Galanter (Hrg.), New York, 362-428.

Odifreddi, Piergiorgio
1990 *About logic and logicians* - A palimpsest of essays by Georg Kreisel, volume II: Mathematics; manuscript.

Sieg, Wilfried

1990 Relative consistency and accessible domains; Synthese 84, 259-297.

1992 Mechanical procedures and mathematical experience; *Mathematics and Mind*, A. George (Hrg.), Oxford University Press.

Tamburrini, Guglielmo

1988 *Reflections on Mechanism*; Ph.D. Thesis, Columbia University, New York.

Turing, Alan

1936 On computable numbers, with an application to the Entscheidungsproblem; Proc. London Math. Soc. 42, 230-265; auch in [Davis 1965], 116-151.

1939 Systems of logic based on ordinals; Proc. London Math. Soc. 45, 161-228; auch in [Davis 1965], 155-222.

1947 Lecture to the London Mathematical Society on 20 February 1947; in: *A.M. Turing's ACE Report of 1946 and Other Papers*, B.E. Carpenter und R.W. Doran (Hrg.), MIT-Press, 1986, 106-124.

1947a Intelligent Machinery; geschrieben im September 1947 und erschienen in: *Machine Intelligence 5*, Edingburgh, 3-23.

van Heijenoort, Jean

1967 (Hrg.) *From Frege to Gödel*; Cambridge (Mass.).

1985 Jacques Herbrand's work in logic and its historical context; in: van Heijenoort *Selected Essays*; Naples, 99-122.

von Neumann, Johan

1927 Zur Hilbertschen Beweistheorie; Mathematische Zeitschrift 26, 1-46.

Webb, Judson C.

1980 *Mechanism, Mentalism, and Metamathematics*; Dordrecht.

1990 Introduction to Remark 3 of [Gödel 1972a] in *Collected Works II*, 292-304.