

# WAS IST DIE BERECHNUNGSHYPOTHESE?

Georg Schwarz<sup>1</sup>

## Vorbemerkung zur Übertragung ins Deutsche.

Im allgemeinen halte ich es für Unfug, englische Ausdrücke selbst dann ins Deutsche zu übernehmen, wenn sie ohne weiteres übersetzbar wären. Ein Paradebeispiel dafür ist der im deutschsprachigen Raum weitverbreitete Gebrauch von "Artificial Intelligence", der sich hartnäckig gegenüber dem naheliegenden "Künstliche Intelligenz" behauptet. Ich gewinne dabei manchmal den Eindruck, daß die Verwendung des englischen Namens für eine Forschungsrichtung bereits als ein Beitrag zu deren Legitimierung verstanden wird.

Manchmal gibt es aber, wie etwa für einen zentralen Begriff einer im englischsprachigen Raum entwickelten Forschungsrichtung, kein offensichtliches deutschsprachiges Gegenstück. Dies trifft gerade auch für die Grundlagendiskussion in der kognitiven Wissenschaft ("cognitive science") zu. Im Falle des für diese Diskussion zentralen Begriffs "computationalism" stand ich vor eben diesem Problem. Da mir kein deutschsprachiges Gegenstück bekannt ist (falls eines vorhanden und in Gebrauch ist, wäre ich für eine diesbezügliche Verständigung dankbar), standen zwei Möglichkeiten offen: entweder "computationalism" beizubehalten, was aber meinem (deutschen) Sprachgefühl nicht zuzumuten war; oder eine Übertragung zu finden, die mit eben diesem Sprachgefühl vereinbar ist. Im folgenden Text schlage ich daher "Berechnungshypothese" als deutschsprachiges Gegenstück vor. Das ebenfalls häufig verwendete Adjektiv "computational" wird dementsprechend entweder als "berechnende(r)/(s)" übersetzt, oder dem zugehörigen Hauptwort als "Berechnungs-" vorangestellt.

Weitaus größere Schwierigkeiten bereiteten Begriffe wie "mind", "cognition", "cognitive" oder "cognize". Letzteres Verb ist ein Kunstwort, dessen Entstehungsgeschichte hier wiederzugeben zu weit führen würde. Wo es im englischen Original vorkommt, habe ich eine durch den jeweiligen Kontext bestimmte sinngemäße deutsche Nachdichtung versucht.

Zum Begriff "cognition" sei bloß gesagt, daß er auch im Englischen nicht immer eindeutig verwendet wird. In seinem weitesten Sinne umfaßt er neben Denk- und Schlußfolgerungs- auch Wahrnehmungs- und Handlungsprozesse, also alles, was in einem nichttrivialen Sinn mit Informationsverarbeitung zu tun hat. Als solcher wird er meist mit "kognitive Prozesse" oder ähnlichen Adjektiv-Substantiv-Paaren übersetzt, da ich wieder einmal einem "Kognition" im Deutschen nichts abgewinnen kann (obwohl es bereits Verwendung findet).

Die Übersetzung anderer problematischer Ausdrücke erfolgte aus dem jeweiligen inhaltlichen Zusammenhang.

Georg Schwarz

## I: Einführung; Fabel.

Die Arbeitshypothese des Großteils der gegenwärtigen kognitiven Wissenschaft besagt, daß ein System dann kognitiv ist, wenn es geeignete Funktionen berechnet. Um diese Behauptung besser bewerten zu können, ist es notwendig, jene erklärende Rolle zu verstehen, die die Berechnung im Gesamtzusammenhang der kognitiven Wissenschaft spielt. Ich beginne daher meine Ausführungen damit, kurz zwischen einer Reihe verschiedener Berechnungsbegriffe zu unterscheiden, die manchmal in Diskussionen über dieses Thema durcheinandergebracht werden.

Ich stelle dann anhand eines einfachen, aber unproblematischen Beispiels die grundsätzlichen Bestandteile einer berechnenden Erklärung vor. Da nun aber laut der Berechnungshypothese Systeme deswegen kognitiv sind, weil sie bestimmte Funktionen berechnen, muß eine vollständige Erklärung solcher Systeme auch bestimmen, was diese Funktionen sind. Obwohl dieses Problem zwar von der ursprünglichen Frage, was eine berechnende Erklärung ist, unabhängig ist, stelle ich trotzdem kurz dar, was der herrschenden Meinung zufolge eine Funktion zur kognitiven macht.

Was aber, wenn sich einige oder sogar die meisten kognitiven Phänomene jener Art unabhängiger Bestimmungen entziehen, wie sie für den Erfolg der herrschenden Meinung notwendig ist? Da diese Möglichkeit zunehmend wahrscheinlicher wird, skizziere ich abschließend einige der Folgerungen, die sich aus der "Widerspenstigkeit" des Bestimmungsproblems ergeben.

## II. Drei Berechnungsbegriffe

Es ist mittlerweile allgemein bekannt, daß die Berechnung eine wichtige Rolle in der kognitiven Wissenschaft spielt. Worin nun diese Rolle genau besteht, ist aber oft viel weniger richtig eingeschätzt. Das wird vor allem dann klar, wenn man sich die Diskussion rund um die Behauptung, daß "der menschliche Geist eine Rechenmaschine sei", zu Gemüte führt. Die zumeist von Gegnern dieser Behauptung vorgebrachten Argumente beziehen ein Sammelsurium verschiedenster Themen ein, die von bestimmten meta-mathematischen Resultaten (z.B. daß jedes formale System mehr als eine Interpretation hat) bis hin zu den technischen Einzelheiten bestehender Computerarchitekturen (z.B.: Wie kann eine sequentiell arbeitende Rechenmaschine überhaupt dazu beitragen, die vielen parallel ablaufenden Prozesse zu erklären, die im menschlichen Gehirn, das dem Verstandesvermögen zugrundeliegt, statthaben?) reichen.

All dies deutet darauf hin, daß eine Reihe verschiedener Berechnungsbegriffe, die eigentlich auseinandergehalten werden sollten, durch und durch vermischt wurden. Dazu kommt noch, daß nur einer dieser Begriffe jenen Anforderungen Rechnung trägt, die von einer naturwissenschaftlichen Untersuchung kognitiver Prozesse verlangt werden, Es mag daher von Nutzen sein, diese verschiedenen Begriffe kurz zu behandeln, was gleichzeitig auch als Einleitung zum eigentlichen Thema dienen wird.

**Berechenbarkeit.** Die Theorie der berechenbaren Funktionen beschäftigt sich damit zu bestimmen, welche (Klassen von) Funktionen berechnet bzw. nicht berechnet werden können. Diesem Unternehmen liegt die Church-Turing-These zugrunde, die in ihrer einfachsten Form besagt, daß jede Funktion, die überhaupt berechenbar ist, von einer Turingmaschine berechnet werden kann.

Eine Turingmaschine wiederum ist eine mathematische Konstruktion, die aus folgenden Bestandteilen besteht: einer endlichen Anzahl von eindeutig bestimmten Regeln, die der Manipulation von Symbolen aus einer ebenfalls festgelegten endlichen Menge dienen; sowie einem (möglicherweise unendlich großen) Speicher, der dem Lesen, der Verarbeitung und dem Schreiben dieser Zeichen dient. Eine Funktion ist genau dann berechenbar, wenn es eine Turingmaschine gibt, die, wenn sie von jenen Zeichen ausgeht, die die Argumente der Funktion repräsentieren, nach einer endlichen (aber möglicherweise gigantischen) Anzahl von Regelanwendungen die entsprechenden Zeichen für die Werte der Funktion produziert. Eine derartige Turingmaschine realisiert einen Algorithmus für die Berechnung dieser Funktion.

Die Berechenbarkeit einer Funktion ist also definiert mittels der Existenz (im mathematischen Sinne) von geeigneten Turingmaschinen. Es sei dabei betont, daß niemals auf die tatsächlichen Funktionsweisen physikalischer Systeme Bezug genommen wurde. Ganz

allgemein gesprochen lassen sich die Resultate der Theorie der Berechenbarkeit auf die oben erwähnte Debatte nur indirekt anwenden, in der Form negativer Einschränkungen. Wenn eine bestimmte Funktion nicht berechenbar ist, wenn sich aber kognitive Prozesse nichtsdestoweniger als berechnende herausstellen, dann ist für den Vertreter der Berechnungshypothese klar, daß diese Funktion nicht Teil des kognitiven Apparats ist.

Auf der anderen Seite gibt es aber Gegner der Berechnungshypothese, die behaupten, daß die bloße Existenz nichtberechenbarer Funktionen bereits zeige, daß "der menschliche Verstand kein Computer sein könne". Aber das allein ist natürlich nicht genug. Selbst wenn gezeigt werden könnte ' daß Menschen tatsächlich nichtberechenbare Funktionen erfüllen (was eine viel detailliertere, aber nicht selbst berechenbare Bestimmung dieser Funktionen notwendig machen würde), und das ist bisher niemandem gelungen, würde das doch nur zeigen, daß ein Aspekt des kognitiven Apparats kein Rechenprozeß ist.

All dies sollte eigentlich kaum überraschen. Die Berechnungshypothese nimmt an, daß kognitive Systeme Funktionen berechnen; die Existenz nichtberechenbarer Funktionen würde nur dann eine Widerlegung darstellen, wenn gezeigt werden könnte, daß diese Funktionen in der kognitiven Ökologie auch wirklich eine wesentliche Rolle spielen. Das ist aber bisher nicht gezeigt worden.

Damit soll natürlich nicht gesagt sein, daß die Berechnungshypothese überhaupt nicht widerlegt werden kann. Ich will hier einzig und allein betonen, daß Resultate aus der Berechenbarkeitstheorie, für sich allein genommen, nur wenig zur Debatte beitragen. Die wirkliche Auseinandersetzung findet in einem anderen Zusammenhang statt, wie wir später sehen werden.

Bevor ich das Thema der theoretischen Berechenbarkeit endgültig zur Seite lege, möchte ich ein Forschungsgebiet erwähnen, das sich mit der Frage der praktischen Berechenbarkeit beschäftigt. Die Komplexitätstheorie untersucht unter anderem, wie "teuer" es ist, bestimmte (Klassen von) Funktionen zu berechnen. Eine Methode für die Bestimmung dieser "Kosten" untersucht, wie die Anzahl der für die Berechnung des Funktionswertes erforderlichen Rechenschritte mit zunehmender "Länge" des Funktionsarguments zunimmt.

Ohne weiter ins Detail zu gehen, sei hier bloß gesagt, daß die Berechnung bestimmter Funktionen so unerschwinglich kostspielig ist (die Anzahl der benötigten Rechenschritte steigt exponentiell mit der Länge des Inputs), daß sie, obwohl zwar theoretisch berechenbar, für alle praktischen Zwecke rechnerisch nicht bewältigbar ist. Die hier erzielten Resultate dienen ebenfalls dazu, den Untersuchungsbereich der Berechnungshypothese einzuschränken. Man kann nämlich davon ausgehen, daß derartige Funktionen nicht in natürlichen Systemen vorkommen. Schließlich würde der Zeit- und Energieaufwand, der für das Berechnen dieser inflationären Funktionen erforderlich wäre, wahrscheinlich bei weitem jeden Vorteil überwiegen, den diese Funktionen für den betreffenden biologischen Organismus bieten könnten.

Rechner. Ein anderer Begriff der Berechnung, der in der Literatur gefunden werden kann, steht in engem Zusammenhang mit gegenwärtig vorhandenen Rechnerarchitekturen. Insbesondere wird oft darauf hingewiesen, daß Computer sequentielle Maschinen sind, während es doch mittlerweile bekannt ist, daß Hirne mehrere Aufgaben gleichzeitig auf höchst parallele Weise erledigen. Daher, so die Behauptung, könne das Gehirn/der Geist kein Computer sein.

Es ist zweifelsohne der Fall, daß die meisten gegenwärtigen Rechner (immer noch) sogenannte Von-Neumann-Maschinen sind, was heißt, daß sie aus einem Eingabe/Ausgabe-Modul, einem Speicher und nur einem Verarbeitungselement bestehen. Und sie eignen sich

daher besonders für die Implementierung serieller Algorithmen. Aber das eben skizzierte Argument wäre nur dann gültig, wenn alle Rechenprozesse notwendigerweise sequentiell wären. Die rapide Entwicklung von parallelen Algorithmen sollte es aber klar machen, daß der durch alle Rechenprozesse aufgespannte Raum nicht mit den Designprinzipien jener Maschinen verwechselt werden darf, die gegenwärtig für seine Untersuchung zur Verfügung stehen. In jedem Fall werden sich etwa bestehen gebliebene Verwirrungen spätestens dann in nichts auflösen, wenn parallele Architekturen weite Verbreitung gefunden haben werden.

Berechnung. Dem aufmerksamen Leser wird es nicht entgangen sein, daß meine bisherigen Absichten hauptsächlich negativ waren. Ich behandelte zwei verschiedene Berechnungsbegriffe und wies sie beide als entweder irrelevant oder ernstlich verworren zurück. Was noch zu tun bleibt, ist, die richtige Rolle der Berechnung bereitzustellen.

### **III: Die erklärende Rolle der Berechnung**

Die Arbeitshypothese des Großteils der gegenwärtigen kognitiven Wissenschaft besagt, daß Systeme kognitiv sind aufgrund der Berechnung geeigneter Funktionen. Wie sollen wir diese Behauptung verstehen? Ich beginne mit einer Veranschaulichung der Berechnungshypothese, da sie, wie so viele Dinge, leichter anhand eines Musterbeispiels verstanden wird als durch Bezugnahme auf eine abstrakte Charakterisierung.

Rechnen. Während einer archäologischen Expedition graben Sie ein seltsam aussehendes Gerät aus dem Boden. Eine genauere Untersuchung ergibt, daß dieses Gerät einen Deckel hat. Unter diesem wiederum verbergen sich zwölf Knöpfe, von denen jeder ein seltsames Zeichen in seine Oberfläche geritzt hat, und ein kleines Fenster. Die Neugierde verleitet Sie zum Experiment, und Sie stellen fest, daß nach bestimmten Knopfdruckfolgen bestimmte Zeichen in dem Fenster erscheinen, wobei diese Anzeigen nur aus jenen Zeichen bestehen, die auch auf den Knöpfen aufscheinen. Überdies scheinen die Anzeigen mit den ihnen jeweils vorangegangenen Knopfdruckfolgen in einem nicht ganz zufälligen Verhältnis zu stehen.

Da Sie von dem immer interessanter werdenden Verhalten dieses Geräts fasziniert sind, beschließen Sie, ein wenig Kryptographie zu versuchen.

Die Anzeige verwendet immer nur zehn der insgesamt zwölf seltsamen Zeichen, also dauert es nicht lange, bis Sie diese auf die zehn Ziffern unseres Dezimalsystems abbilden. Unter dieser Interpretation vollzieht das Gerät Operationen über Zahlen, daher nehmen Sie zusätzlich an, daß einer der beiden übrigen Druckknöpfe zur Trennung der Argumente dient, während der andere eine Art Eingabetaste ist, wie etwa "=" bei einem Taschenrechner. Der Rest ist einfach, denn nach ein paar Versuchen zeigt es sich, daß die Ausgabe des Geräts (sein Anzeigezustand) durchwegs als das Produkt jener Zahlen interpretiert werden kann, die die jeweils vorangegangene Tastendruckfolge interpretieren. Wie sich soeben herausgestellt hat, ist Ihr altertümliches Gerät eine Multipliziermaschine. Was ist passiert?

Erfüllen. Die Multiplikation ist die bekannte Relation zwischen einem Zahlenpaar und einer dritten Zahl, dem Produkt der ersten beiden.

Rechenmaschinen hingegen sind mechanische oder elektronische Geräte, deren Verhalten durch die Gesetze der Physik bestimmt ist. Wie ist es überhaupt möglich, und auf Grund welcher Umstände, daß Rechner die Produktfunktion erfüllen?

Was unser Gerät zur Multipliziermaschine macht, ist die Tatsache, daß sich bestimmte Knopfdruckfolgen und Anzeigezustände systematisch als Zahlen und deren Produkte interpretieren ließen. Ein Gerät wird also dadurch zur Multipliziermaschine, daß eine Interpretationsfunktion existiert (im mathematischen Sinn von Existenz), die systematisch

physikalische Zustandsübergänge auf die Argumente und Werte der Produktfunktion abbildet. Und es ist genau diese Interpretationsfunktion, die physikalische Zustände des Systems als Repräsentationen der Argumente und Werte der erfüllten Funktion erkennen läßt, in unserem Fall als Ziffern.

Jedes physikalische System mit den erforderlichen repräsentierenden Zuständen gilt daher als Multipliziermaschine.

Eine Multipliziermaschine ist also ein Gerät, bei dem das Hervorrufen von den Repräsentationen zweier Zahlen die Repräsentation ihres Produkts verursacht. Um daher zu erklären, wie ein Gerät multipliziert, muß man (zumindest) erklären, wie es möglich ist, daß Repräsentationen von Multiplikator und Multiplikand Repräsentationen von (den richtigen) Produkten verursachen.

Eine mögliche Lösung für dieses Problem besteht darin, zu zeigen, daß das Gerät die Repräsentationen der Produkte aus den Repräsentationen von Multiplikator und Multiplikand berechnet. Die Arbeitshypothese des Vertreters der Berechnungshypothese besagt, daß Systeme wie unsere Multipliziermaschine die Funktionen, die sie erfüllen, berechnen, und daß aufgrund dieser Berechnung die einschlägigen semantischen Regelmäßigkeiten gewährleistet sind (daß eben die Repräsentationen zweier Zahlen die Repräsentation ihres Produkts verursachen).

Berechnen. Eine Möglichkeit, eine Funktion zu erfüllen, besteht darin, sie zu berechnen. Rechenmaschinen erfüllen die Produktfunktion, indem sie diese berechnen. Ein frei fallender Apfel hingegen erfüllt zwar die Funktion  $D=(at^2)/2$ , berechnet sie aber nicht. Worin liegt der Unterschied?

Wir selbst berechnen Produkte meist, indem wir zuerst die Teilprodukte berechnen und diese dann addieren. Dieser Algorithmus ist, wie alle symbolischen Algorithmen, für ein bestimmtes Zeichensystem definiert. Er funktioniert zum Beispiel nicht für das römische Ziffernsystem. Im Fall des Teilproduktalgorithmus sind die Gegenstände der Berechnung — die Dinge, die der Algorithmus manipuliert — die Ziffern unseres Dezimalsystems. Während die Exekution eines Algorithmus immer zur Folge hat, daß Regeln zur Manipulation von Gegenständen gefolgt wird, müssen diese aber nicht unbedingt Symbole sein. Diese Gegenstände müssen nicht einmal Repräsentationen welcher Art auch immer sein. Die auf einem Rechner ausgeführten Algorithmen manipulieren zwar zu meist Symbole, aber viele alltägliche und nützliche Algorithmen tun das nicht. Augenfällige Beispiele für letztere sind Kochrezepte und Gebrauchsanweisungen.

Der Berechnungshypothese zufolge besteht das Berechnen einer Funktion  $f$  im Exekutieren eines Algorithmus, der genau dann die Ausgabe  $o$  bei Eingabe  $i$  liefert, wenn  $f(i)=o$ . Das Problem, die Berechnung einer Funktion zu erklären, reduziert sich daher auf das Problem zu erklären, was es heißt, einen Algorithmus für diese Funktion zu exekutieren. Die offensichtliche Strategie hier besteht im Ausschachten der Idee, daß die Exekution eines Algorithmus mit Schritten verbunden ist. Während einige dieser Schritte ihrerseits die Berechnung anderer Funktionen erforderlich machen können, wie etwa das Addieren der Teilprodukte im obigen Beispiel, werden die Basisschritte als Funktionen verstanden, die das exekutierende System (oder einer seiner Bestandteile) einfach erfüllt. Einen Algorithmus zu exekutieren heißt, seine Grundschritte zu erfüllen, und zwar, wie wir sagen, "in der richtigen Reihenfolge". Kurz gesagt: eine Funktion zu berechnen, heißt, einen Algorithmus für sie zu exekutieren, und die Exekution eines Algorithmus besteht in der disziplinierten Erfüllung der Basisschritte.

Es gilt noch zu erklären, unter welchen Bedingungen das Erfüllen von Schritten diszipliniert ist. Im Prinzip ist diese Frage einfach zu beantworten. Die grundlegende Idee hier ist, daß ein System  $d$  deswegen die Schritte in der richtigen Reihenfolge erfüllt, weil sich diese Schritte gegenseitig kausal beeinflussen: jede Erfüllung eines Schrittes ist ein Ereignis in  $d$ , und Ereignisse in  $d$  haben Wirkungen in  $d$ ; und unter diesen Wirkungen befinden sich Ereignisse, die die Erfüllung anderer Schritte ausmachen, und so weiter. Die Kausalstruktur des exekutierenden Systems übernimmt die Haftung für die Disziplin, nach der wir gesucht haben.

Erklären. Ein physikalisches System berechnet also eine Funktion (anstatt sie bloß zu erfüllen), wenn es einen Algorithmus für diese Funktion exekutiert. Im Gegensatz dazu hat der fallende Apfel nicht seine Fallkurve auf Newtons Kopf hin berechnet. Es ist wichtig, sich darüber im klaren zu sein, daß Berechnung, derart verstanden, ein abstrakter Kausalprozeß ist — das heißt, ein Kausalprozeß, der abstrakt als Algorithmus bestimmt ist.

Genau das ist ja, was vonnöten ist, wenn Berechnung überhaupt bei der Erklärung des Verhaltens physikalischer Systeme eine Rolle spielen soll.

Es gibt viele Kausalprozesse, und nur manche davon sind Fälle von Funktionsberechnungen. Es sind aber genau die letzteren, für die eine berechnende Erklärung zuständig ist. Überdies gibt es viele Möglichkeiten, Funktionen zu erfüllen, auch berechenbare (d.h. Funktionen für die Algorithmen existieren), die nicht die Exekution eines Algorithmus voraussetzen.

Eine berechnende Erklärung für die Multiplizierfähigkeit unseres Geräts wird daher versuchen, zuerst repräsentierende Zustände und einen über sie definierten Algorithmus zu identifizieren, wobei dieser die "richtige" (abstrakte) Kausalstruktur des Geräts bestimmen muß, d.h. jene Kausalstruktur, die die Übergänge von einem repräsentierenden Zustand zu einem anderen garantiert. Dabei mag unser Wissen um existierende Algorithmen (wie Teilprodukte, oder aufeinanderfolgende Addition) helfen, ebenso wie unsere Vertrautheit mit verschiedenen formalen Systemen für die Repräsentation von Zahlen (binär, dezimal, etc.). Andererseits könnte sich unser System als gigantische Nachschlagtabelle herausstellen, vielleicht in der Form eines assoziativen Netzwerks. Und es mag sich dabei herausstellen, daß die Grundsätze für den Entwurf leistungsfähiger künstlicher Systeme von wenig Nutzen sind, wenn es um die Erklärung der Vorgänge in natürlichen Systemen geht.

Bisher haben wir uns ausschließlich mit der Darstellung der berechnenden Erklärung beschäftigt. Was wir noch nicht behandelt haben, ist die Frage, was im Fall von kognitiven Systemen dementsprechend erklärt werden soll.

Kognitiv. Wenn kognitive Systeme als berechnende Systeme erklärt werden sollen, müssen sie bestimmte Funktionen erfüllen. Einer (durch und durch rationalistischen) Auffassung zufolge ist das Verhalten eines Systems einem bestimmten Bereich gegenüber kognitiv (und ist nicht eine bloße Reaktion auf diesen), wenn dieses Verhalten die diesem Bereich eigenen erkenntnistheoretischen Beschränkungen beachtet. Das heißt, daß sein Verhalten relativ zu seiner Eingabe und seinem inneren Zustand schlüssig, oder gerechtfertigt, oder vernünftig ist. Diese erkenntnistheoretischen Beschränkungen können wir uns so vorstellen, daß sie eine Funktion bestimmen, was dann zur Folge hat, daß ein System, das diese Beschränkungen beachtet, die entsprechende Funktion erfüllt. Die Berechnungshypothese nimmt nun an, daß Systeme kognitiv sind, indem sie einschlägige kognitive Funktionen berechnen. Und zwar erledigen sie das auf dieselbe Art und Weise wie Multipliziermaschinen die Produktfunktion berechnen, d. h. indem sie einen Algorithmus exekutieren, welcher, ausgehend von Repräsentationen der Argumente der Funktion, die Repräsentationen der entsprechenden Werte liefert.

Repräsentation. Die Verpflichtung zu Repräsentationen für sich genommen ist übrigens nicht ein hervorstechendes Merkmal der Berechnungshypothese.

Beinahe jeder in der kognitiven Wissenschaft glaubt mittlerweile an die Existenz von Repräsentationen. Und alle glauben, daß Repräsentationen kausal verknüpft sind. Was aber die Berechnungshypothese doch unterscheidet, ist, daß es nicht bloß kausale, sondern berechnende Prozesse sind, die das Auftreten von Repräsentationen disziplinieren. Die Objekte der semantischen Interpretation — die Repräsentationen — sind ident mit den Objekten der Berechnung — die vom Algorithmus manipulierten Gegenstände. Es ist diese zusätzliche Annahme, die die Berechnungshypothese von all jenen unterscheidet, die nicht davon ausgehen, daß Repräsentationen buchstäblich eine bestimmte Gestalt haben, eine Gestalt, die ihre berechnende Manipulation ermöglicht.

Ein anderer Punkt, den ich in diesem Zusammenhang für erwähnenswert halte, betrifft die vor kurzem aufgeflammete Diskussion um den Unterschied zwischen sogenannten "klassischen" Modellen und dem Neuen Konnektionismus.

Es ist wichtig zu verstehen, daß die Berechnungshypothese als solche weder zu irgendeinem Repräsentationssystem verpflichtet noch auf ein solches beschränkt ist. Ganz im Gegenteil, die meisten konnektionistischen Modelle haben selbst mit den extremsten logizistischen Modellen etwas gemein, nämlich ein starkes Bekenntnis zu Repräsentationen. Worin sie sich, und zwar radikal, unterscheiden, ist, welches Repräsentationssystem (und folglich welche Klassen von Algorithmen) sie verwenden, nicht aber, ob sie überhaupt von Repräsentationen Gebrauch machen.

#### **IV. Das Bestimmungsproblem**

Von besonderer Bedeutung für die Berechnungshypothese ist die Annahme, daß es unabhängig bestimmbare kognitive Funktionen gibt, d.h. daß kognitive Fähigkeiten bestimmbar sind ohne Bezugnahme auf die Art und Weise, wie sie irgendein System erfüllt oder berechnet. Eine kognitive Funktion, wie wir uns erinnern, ist eine Funktion, die eine Reihe erkenntnistheoretischer Beschränkungen ausdrückt, deren Berücksichtigung als Erfüllung der Funktion selbst gilt. Die Berechnungshypothese beinhaltet daher, daß sich diese Beschränkungen immer als berechenbare Funktion ausdrücken lassen.

Doch gibt es mittlerweile Grund zur Annahme, daß diese Beschränkungen, zumindest für manche Bereiche, nicht einmal bestimmbar sind, ganz zu schweigen von einer Bestimmung in Form einer berechenbaren Funktion.

Eine Analogie mit der Naturwissenschaft läßt dies ziemlich glaubhaft erscheinen. Eine Einzelwissenschaft kodifiziert, was es in ihrem Zuständigkeitsbereich zu wissen gibt. Aber eine Einzelwissenschaft ist nicht für jeden x-beliebigen Bereich möglich. Zum Beispiel ist eine wissenschaftliche Untersuchung der Temperatur als solche nur dann möglich, wenn Wärmephänomene einen unabhängigen Untersuchungsbereich bilden. Das heißt, daß die Thermodynamik nur bis zu jenem Ausmaß möglich ist, in dem die Wärmephänomene von Gesetzen geregelt sind, die diese Phänomene in einer "ihnen eigenen" Weise, und damit unabhängig, beschreiben. Im Gegensatz dazu ist eine unabhängige Einzelwissenschaft für Bekleidung kaum möglich, da es keine besonderen Kleidungs Gesetze gibt. Kleidungsstücke fallen natürlich unter alle möglichen Gesetze, z.B. die Gesetze der Physik, der Chemie, der Ökonomie; es steht also außer Frage, daß eine wissenschaftliche Untersuchung von Kleidungsstücken möglich ist. Es gibt außerdem zweifelsohne "Regeln" für den Kleidungsbereich: was man für welche Gelegenheit anzieht, welche Gewebe mit welchen Temperaturen gewaschen werden sollen usw. Aber es ist fraglich, ob sich diese Regeln in einem Expertensystem zusammenfassen lassen, und folglich fraglich, ob eine Ansammlung

von Regeln gefunden werden kann, die jene Beschränkungen ausdrückt, die Menschen ständig erfüllen, wie auch immer sie das tun.

Möglicherweise kann die menschliche Leistung in diesen Bereichen deshalb nicht im Befolgen von Regeln bestehen weil sie nicht in Form von Regeln beschreibbar ist; und vielleicht ist sie deshalb nicht in Form von Regeln beschreibbar, weil sie letztendlich anhand von Beispielen gesteuert ist.

Was wenn? Die mögliche Unlösbarkeit des Bestimmungsproblems würde auch den spektakulären Mangel an Erfolgen erklären, der die Versuche, kognitives Verhalten zu synthetisieren, begleitet hat. Wie ein Gutteil der Geschichte der Künstlichen Intelligenz zeigt, steht das Ausmaß der "Wohldefiniiertheit" einer Funktion oft in genau umgekehrtem Verhältnis dazu, wie gut Menschen sie verarbeiten: je "knackiger" die Funktion, desto unkomplizierter ihre Ausführung auf einem Rechner, wie z.B. im Fall des Beweisen von Theoremen. Es zeigt sich aber nur zu oft, daß Menschen bei solchen Funktionen besonders schwache Leistungen erbringen, wie z.B. im Fall des Beweisen von Theoremen. Auf der anderen Seite wiederum verfolgen all die "unsauberen Sachen", die Menschen so problemlos beherrschen, wie z.B. das Erkennen verschiedener Handschriften, die Künstliche Intelligenz seit ihrem Entstehen (von dem berüchtigten und anscheinend immerwährend entweichenden "gesundem Hausverstand" einmal ganz abgesehen).

Sollte sich also das Bestimmungsproblem als unlösbar herausstellen, was anscheinend immer wahrscheinlicher wird, dann steht die kognitive Wissenschaft ohne autonomen Untersuchungsbereich da. Das muß aber nicht so verheerend sein, wie es klingen mag. Denn während sich die Annahme einer "Kognition" an sich als illusorisch herausgestellt haben wird, werden wir ja weiterhin ständig von Beispielen dessen umgeben sein, was wir erklären wollen. Kognitive Fähigkeiten werden dann ostentativ als das identifiziert werden, was Menschen tun, wenn sie ... ihre Großmutter erkennen, bestimmte Probleme lösen, den Weg nach Hause finden usw. Um diese Fähigkeiten zu erklären wird man nun aber viel mehr Augenmerk auf jene Möglichkeiten und Beschränkungen legen müssen, die durch die biologischen und/oder berechnenden Prozesse im menschlichen Zentralnervensystem bestimmt sind.

Dem vor kurzem explodierten Interesse am Konnektionismus mag bereits dieser Einstellungswechsel zugrundeliegen.

1 Da die folgende Arbeit erst vergleichsweise spät in Auftrag gegeben wurde, habe ich bereits vorhandenes Material verwendet. Das trifft vor allem auf Teil III zu, der zum Großteil beinahe wortwörtlich einem Aufsatz entnommen ist, den ich gemeinsam mit Robert Cummins geschrieben habe. Dieser Aufsatz, "Connectionism, computation, and cognition", wurde für den von T. Horgan und J. Tienson herausgegebenen Sammelband *Connectionism and the Philosophy of Mind* geschrieben, der Anfang nächsten Jahres bei Kluwer erscheinen wird. Für eine gründliche Untersuchung der Berechnungshypothese verweise ich auf: R. Cummins (1989). *Meaning and Mental Representation*. Boston, MA: The MIT Press/A Bradford Bock.

Da die folgende Arbeit außerdem vor allem dafür gedacht ist, einem interessierten, aber nicht unbedingt spezialisierten Publikum die wesentlichsten Merkmale der Berechnungshypothese vorzustellen, habe ich, um die Lektüre etwas zu erleichtern, auf die sonst üblichen bibliographischen Hinweise verzichtet.