

## Ist eine Sprache des Geistes möglich?\*

(In: A. Burri (ed.) *Sprache und Denken*.  
Berlin/New York: Walter de Gruyter 1997, 75-92)

Ansgar Beckermann

### 1.

Kognitionswissenschaften – in einem weiten Sinn – sind einfach alle die Wissenschaften, die sich mit der Analyse und Erklärung kognitiver Leistungen und Fähigkeiten befassen. Wenn man jedoch von *der* Kognitionswissenschaft im Singular spricht, dann ist in der Regel mehr gemeint. Für die Kognitionswissenschaft ist nicht nur ein bestimmter Forschungsgegenstand charakteristisch, sondern auch ein bestimmter Erklärungsansatz: der Informationsverarbeitungsansatz. Stillings et al. z.B. schreiben gleich auf der ersten Seite ihres 1987 erschienenen Buches *Cognitive Science – An Introduction*: „Cognitive scientists view the human mind as a complex system that receives, stores, retrieves, transforms, and transmits information.“ (Stillings et al. 1987: 1) Der Informationsverarbeitungsansatz führt jedoch sofort weiter zum Symbolverarbeitungsansatz. Denn offenbar kann ein System nur dann Informationen empfangen, speichern und verarbeiten, wenn es über ein System von internen Repräsentationen oder Symbolen verfügt, über eine interne Sprache, in der diese Informationen codiert sind. Zumindest ist das eine naheliegende Überlegung, die Peter Hacker so formuliert hat: „... if information is received, encoded, decoded, interpreted and provides grounds for making plans, then there must be a language or system of representation in which this is all done.“ (Hacker 1987: 486f.) In der Tat ist die Annahme, daß es in kognitiven Systemen so etwas wie ein System interner Repräsentationen bzw. eine Sprache des Geistes<sup>1</sup> gibt, die zentrale Grundannahme vieler neuerer Arbeiten in den Bereichen der Kognitionspsychologie und der kognitiven Neurobiologie. Für diese Wissenschaften hat diese Annahme den Status einer empi-

---

\* Bei diesem Aufsatz handelt es sich um die leicht überarbeitete Fassung eines Aufsatzes, der zuerst erschienen ist in G. Harras (Hg.) *Die Ordnung der Wörter*. Berlin/New York: de Gruyter 1995, 120-137.

<sup>1</sup> Der Ausdruck „Sprache des Geistes“ („lingua mentis“, „language of thought“), der in diesem Zusammenhang möglicherweise zum ersten Mal von G. Harman (1973) verwendet wurde, ist der Sache nach außerordentlich irreführend. Denn die Ausdrücke der Sprache des Geistes sind auch Fodor zufolge interne *physische* Systemzustände – also z.B. bestimmte neuronale Feuermuster oder Bitmuster im Speicher eines Computers. Treffender wären deshalb Ausdrücke wie „Sprache des Gehirns“ oder „Sprache des Computers“.

rischen Hypothese, d.h. für sie sind interne Repräsentationen oder Symbole theoretische Konstrukte, die deshalb postuliert werden, weil sie gut bestätigte und systematisch besonders befriedigende Erklärungen kognitiver Leistungen ermöglichen. In der Philosophie gibt es jedoch auch Ansätze, die Annahme einer Sprache des Geistes durch sehr grundsätzliche Überlegungen zur Natur mentaler Zustände zu stützen.

Der Hauptmatador in diesem Feld ist sicher Jerry Fodor, der seine *Repräsentationale Theorie des Geistes* (RTG) über viele Jahre hinweg entwickelte,<sup>2</sup> bis er ihr in dem Buch *Psychosemantics* ihre sozusagen kanonische Form gegeben hat. In dieser Form umfaßt die RTG zunächst zwei Teilthesen:

- (1) Für jeden Organismus  $O$  und jeden Typ  $A$  intentionaler Zustände gibt es eine (funktionale/computationale) Relation  $R$ , so daß gilt:  
 $O$  ist genau dann in einem intentionalen Zustand des Typs  $A$  mit dem Inhalt  $p$ , wenn sich  $O$  in der Relation  $R$  zu einer mentalen Repräsentation  $m$  befindet und  $m$  die Bedeutung  $p$  hat.
- (2) Mentale Prozesse sind kausale Abfolgen einzelner mentaler Repräsentationen. („Mental processes are causal sequences of tokenings of mental representations“). (Fodor 1987: 17)

Während die zweite dieser beiden Thesen relativ klar ist, ist die erste vielleicht nicht ohne weiteres verständlich. Was also ist mit dieser These gemeint?

Wichtig ist zunächst, daß es Fodor bei dieser These um eine Antwort auf die Frage geht, wie bestimmte mentale Zustände – nämlich *intentionale* Zustände – *physisch* realisiert sein können. Intentionale Zustände – wie Wünsche, Überzeugungen, Hoffnungen und Befürchtungen – sind dadurch gekennzeichnet, daß sie *auf etwas gerichtet* sind, daß sie einen *Inhalt* haben. Man glaubt, *daß etwas der Fall ist*, man wünscht sich *einen bestimmten Gegenstand*, man hofft oder befürchtet, *daß ein bestimmtes Ereignis eintreten wird*, usw. Bei allen intentionalen Zuständen kann man also zwei Aspekte unterscheiden: die Art des Zustandes und seinen Inhalt. Mein Wunsch, ein neues Fahrrad zu erwerben, und mein Wunsch, einen alten Freund wiederzutreffen, sind intentionale Zustände derselben Art: beides sind Wünsche – allerdings Wünsche mit verschiedenen Inhalten. Meine Befürchtung, daß es heute regnen wird, und meine Überzeugung, daß es heute regnen wird, sind dagegen intentionale Zustände verschiedener Art. Aber auch sie haben etwas gemeinsam; sie haben denselben Inhalt: sie richten sich beide auf die Proposition, daß es heute regnen wird.

---

<sup>2</sup> Vgl. bes. Fodor (1975; 1978; 1981b; 1987).

Fodors These ist nun, daß diesen beiden Aspekten intentionaler Zustände auch verschiedene Aspekte ihrer physischen Realisierungen entsprechen: dem Inhalt eine mentale Repräsentation und dem Zustandstyp eine bestimmte computationale bzw. funktionale Relation. Mentale Repräsentationen sind dabei als *innere physische Strukturen* zu verstehen, die *etwas repräsentieren* und die insofern ähnlich wie die Sätze einer Sprache oder Landkarten oder die Kerben auf dem Griff des Revolvers eines Westernhelden *eine Bedeutung haben*. Konkret könnte man dabei z.B. an Listen-Strukturen denken wie sie etwa in der Programmiersprache LISP gebräuchlich sind. Daß ein System zu einer mentalen Repräsentation in einer bestimmten funktionalen oder computationalen Relation steht, bedeutet andererseits, daß diese Repräsentation in dem System (z.B. bei der Hervorbringung des Verhaltens des Systems) eine bestimmte funktionale oder computationale Rolle spielt. Hier kann man sich z.B. vorstellen, daß sich mentale Repräsentationen in verschiedenen Boxen (verschiedenen Speicherbereichen) – etwa einer *belief*- oder einer *desire*-Box – befinden und daß das System Repräsentationen, die sich in verschiedenen Boxen befinden, unterschiedlich verarbeitet. Auf jeden Fall gilt Fodor zufolge, daß ein System dann und nur dann z.B. die Überzeugung hat, daß es heute regnen wird, wenn es in dem System eine mentale Repräsentation, d.h. eine physische Struktur gibt, die die Bedeutung hat, daß es heute regnen wird, und wenn diese Repräsentation in dem System die für Überzeugungen charakteristische funktionale bzw. computationale Rolle spielt.

Mit den Thesen (1) und (2) ist Fodors RTG jedoch noch nicht vollständig charakterisiert. Denn eine zusätzliche – und wahrscheinlich sogar entscheidende – Annahme ist für ihn, daß mentale Repräsentationen eine satzartige innere Struktur besitzen. Zu den Thesen (1) und (2) kommt daher noch die eigentliche *Language of Thought*-These:

- (3) a. Mentale Repräsentationen sind *strukturiert*; sie haben typischerweise eine Konstituentenstruktur.
- b. Die Teile dieser Strukturen sind „transportierbar“; dieselben Teile können in verschiedenen Repräsentationen auftreten.
- c. Mentale Repräsentationen haben eine *kombinatorische Semantik*: Ihre Bedeutung hängt in regelhafter Weise von der Bedeutung ihrer Teile ab.

Ich kann an dieser Stelle nicht auf alle Gründe eingehen, die nach Fodor für die Richtigkeit der drei Thesen der RTG sprechen; aber ich will ein Hauptargument we-

nigstens andeuten.<sup>3</sup> Es ist schon deutlich geworden, daß Fodor diese Thesen im Zusammenhang einer Analyse intentionaler Zustände entwickelt hat. Genauer kann man sagen, daß Fodors Argumentation für diese Thesen von zwei Prämissen ausgeht:

- (4) Es gibt intentionale Zustände, d.h. Zustände, die zugleich kausal wirksam und semantisch evaluierbar sind.
- (5) Um kausal wirksam sein zu können, müssen intentionale Zustände physisch realisiert sein.

Wenn man diese Prämissen akzeptiert, steht man jedoch sofort vor einem Problem: Intentionale Zustände haben eine ganze Reihe von Merkmalen, die einer physischen Realisierbarkeit auf den ersten Blick im Wege zu stehen scheinen. Das einzige dieser Merkmale, auf das ich hier eingehen will, besteht darin, daß Kausalbeziehungen zwischen intentionalen Zuständen häufig Rationalitätsprinzipien bzw. semantischen Relationen zwischen ihren Inhalten entsprechen. Wenn jemand  $p$  glaubt, dann wird er in der Regel auch alle offensichtlichen Folgerungen aus  $p$  glauben. Und wenn jemand  $q$  will und glaubt, daß  $p$  eine notwendige Voraussetzung zur Erreichung von  $q$  ist, dann wird er in der Regel auch  $p$  wollen.

Damit stellt sich jedoch die Frage, wie denn physische *Mechanismen* aussehen können, die solchen Kausalbeziehungen zugrundeliegen, d.h. wie physische Mechanismen aussehen können, denen ihrerseits Rationalitätsprinzipien entsprechen. Und diese Frage führt für Fodor direkt zum Symbolverarbeitungsansatz. Denn auf der einen Seite hat die Beweistheorie, auf die sich Fodor immer wieder bezieht, gezeigt, daß der Begriff der logischen Folgerung formalisiert werden kann, d.h. daß man diesen Begriff durch Bezugnahme auf geeignete Kalküle auch rein syntaktisch charakterisieren kann. Und auf der anderen Seite haben die Computerwissenschaften gezeigt, daß solche syntaktischen Umformungsprozesse mit Hilfe von Symbolverarbeitungsprozessen physisch realisiert werden können. Mit anderen Worten: Physische Mechanismen, die Rationalitätsanforderungen respektieren, lassen sich mit Hilfe von Symbolverarbeitungsprozessen realisieren – allerdings nur unter der Voraussetzung, daß die zugrundeliegenden Symbole oder Repräsentationen strukturiert sind. Denn Symbolverarbeitungsprozesse bestehen gerade darin, daß sie Repräsentationen nach Regeln verändern, die auf die Struktur dieser Repräsentationen Bezug nehmen.

Fodors Hauptargument für die Thesen (1) – (3) kann man daher so zusammenfassen: Kausale Mechanismen, die Rationalitätsanforderungen respektieren, können nur

---

<sup>3</sup> Ausführlicher werden die Fodorschen Argumente in Beckermann (1991) erläutert.

auf Symbolverarbeitungsprozessen beruhen und Symbolverarbeitungsprozesse setzen strukturierte Repräsentationen voraus.

## 2.

Die RTG Fodors, die ich im letzten Abschnitt kurz dargestellt und erläutert habe, ist inzwischen von verschiedenen Autoren in sehr unterschiedlicher Weise kritisiert worden. Besonders in Oxford sind in den letzten Jahren aber vom späten Wittgenstein ausgehende kritische Überlegungen laut geworden, die nicht nur diese Theorie, sondern den gesamten Symbolverarbeitungsansatz radikal in Frage stellen.<sup>4</sup> Peter Hacker z.B. stellt die rhetorische Frage: „Is this [scl. the idea that there is a language in the brain] just a picturesque metaphor or helpful analogy? Or is it a symptom of widespread confusion in the presentation, description and explanation of experimental data ...?“ (op.cit.: 487) Und seine Antwort lautet in der Tat, daß die Idee eines Symbolsystems im Gehirn bzw. die Idee einer Sprache des Geistes auf einer grundlegenden Begriffsverwirrung beruhe und daher wortwörtlich *sinnlos* sei. Was sind Hackers Gründe für diese niederschmetternde Diagnose?

Zunächst charakterisiert er noch einmal die Idee, die er dann attackieren will:

The general conception at work involves the supposition that the brain has a *language* of its own, which consists of *symbols* that *represent* things. It uses the *vocabulary* of this language to *encode information* and it produces *descriptions* of what is seen ... (op.cit.: 488)

A ‘symbolic description’ is presumably an array of symbols which are so combined as to yield a true (or false) characterization of a certain aspect of the world. It must be cast in a certain language which has a vocabulary and grammar. (ibid.)

Aber was kann es bedeuten, daß das Gehirn über eine Sprache mit eigenem Wortschatz und eigener Grammtik verfügt? Was heißt es überhaupt, daß jemand über eine Sprache verfügt? „Someone who *has* a language has mastered a technique, acquired or possesses a skill of using symbols in accord with rules for their correct use, or – if you prefer – in accord with their meaning.“ (op.cit.: 491f.)

Wer eine Sprache beherrscht, verfügt also über bestimmte Fähigkeiten. Er versteht in dieser Sprache gemachte Äußerungen; er kennt die Bedeutung der Wörter dieser Sprache und kann diese Wörter verwenden, um selbst die unterschiedlichsten Sprechhandlungen auszuführen: Er kann ein Taxi rufen, nach dem Weg zum Bahn-

---

<sup>4</sup> Siehe besonders Hacker (1987); aber auch den neuen Sammelband Hyman (1991).

hof fragen, Geschichten und Witze erzählen, Wein zum Essen bestellen, einen Freund vorstellen, eine Landschaft beschreiben, und und und. Außerdem kann er, falls er einmal nicht verstanden wird, erklären, was die von ihm verwendeten Wörter bedeuten und was er mit ihnen hat sagen wollen. Alles in allem: „If [someone] understands a language he can respond in various ways to others’ uses of words and sentences, as well as correcting others’ errors, querying their unclarities and equivocations.“ (op.cit.: 492) Allein aus der Tatsache, daß Sprachbeherrschung alle diese Fähigkeiten impliziert, folgt Hacker zufolge schon zwingend, daß es im Wortsinne keinen Sinn hat, zu sagen, daß Gehirn verfüge über eine Sprache.

Only of a creature that can perform acts of speech does it make sense to say that it has, understands, uses, a language. But it is literally unintelligible to suggest that a brain, let alone *a part of a brain*, might ask a question, have or express an intention, make a decision, describe a sunset, undertake an obligation, explain what it means, insist, assert, instruct, demand, opine, classify, and so forth. (ibid.)

Um überhaupt über eine Sprache verfügen zu können, muß man fähig sein, bestimmte Handlungen auszuführen – Handlungen, die auf einer ganz anderen Ebene liegen als die, von denen man sinnvollerweise sagen kann, daß sie von einem Gehirn oder gar von Teilen eines Gehirn ausgeführt werden könnten. Gehirne oder Gehirnteile sind daher schon aus begrifflichen Gründen keine möglichen Sprachverwender.

Aber es gibt noch mehr Gründe, die Hacker zufolge zeigen, daß die Idee einer Sprache des Gehirns immer absurder wird, je mehr wir uns über ihre Implikationen klar werden. Die Ausdrücke einer Sprache, so Hacker weiter, haben eine durch Konventionen geregelte Verwendung, und jemand, der über eine Sprache verfügt, muß die korrekte Verwendung dieser Ausdrücke kennen, d.h. er muß korrekte von unkorrekten Verwendungen unterscheiden können. Ein geregelter Sprachgebrauch in diesem Sinne, ein Sprachgebrauch, der sich an Standards der Korrektheit orientiert, kann aber nur in einer sozialen Praxis fundiert sein.

For only where there is a practice of employing a sign can there also be an activity of matching the application of the sign against a standard of correctness. Since signs have a meaning, a use, only insofar as there is a convention, a standard of correctness for their application, there must be a *possibility* of correcting misuses by reference to the standard of correctness for the use of the expression which is embodied in an explanation of meaning. The use of language is essentially a normative activity. (op.cit.: 496)

Auch aus diesem Grund ist es Hacker zufolge völlig unmöglich, daß Gehirne oder gar Gehirnzellen über eine Sprache verfügen. Denn weder von Gehirnen noch von

Gehirnzellen kann man sinnvoll sagen, daß sie Konventionen folgen. Denn Konventionen können nur da befolgt werden, wo überhaupt Konventionen existieren. Und Konventionen kann es nur da geben, wo sie in einer sozialen Gemeinschaft beim Lehren und Lernen, beim Korrigieren von Fehlern und beim Erklären und Rechtfertigen von Handlungen verwendet werden.

Only of a creature who has the *ability* to make a mistake, who can *recognize* his mistake by reference to a standard, who can *correct* his action for the *reason* that it was erroneous, only of such a creature can one say that it follows and uses conventions. (ibid.)

Aus demselben Grunde ist nach Hacker auch die Rede von cerebralen Karten völlig sinnlos. Denn auch Karten sind nur Karten *von etwas*, wenn es entsprechende Konventionen gibt. Eine bestimmte Gegend kann nur dann mit Hilfe einer Karte repräsentiert werden, wenn bei der Erstellung der Karte spezifische kartographische Konventionen befolgt wurden einschließlich bestimmter konventionell geregelter Projektionsmethoden wie etwa der Mercator-Projektion.

So there are no representing maps without conventions of representation. There are no conventions of representation without a *use*, by intelligent, symbol-employing creatures, of the representation. And to *use* a representation correctly one must *know* the conventions of representation, understand them, be able to explain them, recognize mistakes and correct or acknowledge them when they are pointed out. Whether a certain array of lines is or is not a map is not an *intrinsic* feature of the lines, nor even a *relational* feature (that is, the *possibility* of a 1:1 mapping), but a *conventional* one (that is, the *actual* employment, by a person, of a convention of mapping). (op.cit.: 497f.)

Es bleibt also nur die Schlußfolgerung, daß die Idee einer Sprache des Gehirns sinnlos ist. Im Gehirn kann es keine bedeutungshaltigen Symbole geben. Denn Bedeutung setzt das Bestehen von Konventionen voraus, und Konventionen implizieren die Existenz einer entsprechenden sozialen Praxis. Eine solche „soziale Praxis“ ist im Hinblick auf Nervenzellen aber *begrifflich* unmöglich. Die Annahme, es gebe im Gehirn ein Symbolsystem oder eine Sprache, ist daher im Wortsinn „unbegreiflich“.

### 3.

Auf den ersten Blick scheint diese Argumentation außerordentlich schlüssig. Und in der Tat bildet sie ja auch den – von vielen geteilten – Kern einer Wittgensteinianischen Bedeutungstheorie. Bei näherem Hinsehen ist sie allerdings nicht ganz so zwingend. Denn selbst die Bezugnahme auf eine soziale Praxis kann – zumindest

wenn man Kripkes diesbezüglichen Überlegungen folgt<sup>5</sup> – den normativen Charakter von Bedeutung nicht begründen. Auf jeden Fall werden Kripkes Überlegungen von Paul Boghossian so gedeutet.<sup>6</sup> Worin besteht, so fragt Boghossian, der für Bedeutung grundlegende normative Charakter? Und seine Antwort lautet:

Suppose the expression ‘green’ means *green*. It follows immediately that the expression ‘green’ applies *correctly* only to *these* things (the green ones) and not to *those* (the non-greens). The fact that the expression means something implies, that is, a whole set of *normative* truths about my behaviour with that expression: namely, that my use is correct in application to certain objects and not in application to others. ... meaningful expressions possess conditions of *correct* use. (1989: 513)

Genau aus dieser Tatsache ergibt sich das *skeptische Problem* für alle Bedeutungstheorien: „Having a meaning is essentially a matter of possessing a correctness condition. And the sceptical challenge is to explain how anything could possess *that*.“ (1989: 515) Kripkes Hauptargument gegen alle Theorien, die Bedeutung auf natürliche Eigenschaften einzelner Personen zurückführen wollen, und insbesondere gegen die dispositionale Analyse von Bedeutung lautet dementsprechend: Keine der von diesen Theorien ins Feld geführten natürlichen Eigenschaften kann die Tatsache begründen, daß mit einem Ausdruck Korrektheitsbedingungen verbunden sind; und eben deshalb sind alle diese Theorien als *Bedeutungstheorien* zum Scheitern verurteilt.

Dies ist nun der Punkt, an dem Wittgensteinianer auf einer sozialen Praxis beruhende Regeln ins Spiel bringen und argumentieren: Soweit sei zwar alles richtig; aber es zeige nur, daß Bedeutung nichts sei, was durch die Eigenschaften isolierter einzelner Personen konstituiert wird; die Bedeutung eines sprachlichen Ausdrucks ergebe sich erst aus den Regeln, auf denen der Gebrauch dieses Ausdrucks in einer Sprachgemeinschaft beruhe, und diese Regeln ergeben sich ihrerseits aus einer gemeinsamen sozialen Praxis. Aber reicht diese Antwort aus? Können Regeln und kann insbesondere eine soziale Praxis die Korrektheitsbedingungen eines sprachlichen Ausdrucks besser begründen als die natürlichen Eigenschaften von Einzelpersonen?

Daß in einer Gemeinschaft eine Regel *R* gilt, kann man im Anschluß an Hart (1961: 54ff.) folgendermaßen erläutern:<sup>7</sup>

---

<sup>5</sup> Siehe Kripke (1982).

<sup>6</sup> Boghossian (1989).

<sup>7</sup> Vgl. zu dieser Formulierung auch von Savigny (1983: p. 34).

- (1) Die Mitglieder der Gemeinschaft weichen selten von *R* ab;
- (2) wenn ein Mitglied der Gemeinschaft von *R* abweicht, dann ist es Sanktionen seitens der anderen Mitglieder der Gemeinschaft ausgesetzt;
- (3) diese Sanktionen werden im allgemeinen akzeptiert.

Wenn das so ist, dann besteht die Tatsache, daß in einer Gemeinschaft eine Regel gilt, aber auch nur in den *Dispositionen* der Mitglieder der Gemeinschaft. Und dann stellt sich natürlich die Frage, inwiefern die Dispositionen mehrerer Personen die Korrektheitsbedingungen sprachlicher Ausdrücke besser begründen können sollen als die Dispositionen einer Einzelperson.

Hierin liegt der Grund dafür, daß Kripke selbst die Bezugnahme auf die Regeln einer Sprachgemeinschaft auch nur als *skeptische* Lösung des Bedeutungsproblems akzeptiert. Eine *substantielle* Lösung ist seiner Meinung nach unmöglich. *Nichts* in der Welt kann den normativen Charakter, die Korrektheitsbedingungen sprachlicher Ausdrücke begründen. Und daher ist – in einem strikten Sinn – die Schlußfolgerung unausweichlich, daß kein sprachlicher Ausdruck die *Eigenschaft* hat, eine bestimmte Bedeutung zu haben. Es hat daher keinen Sinn zu fragen, worin diese Eigenschaft besteht. Das einzige, was wir tun können, ist zu *beschreiben*, unter welchen Bedingungen wir welchen Wörtern welche Bedeutungen *zuschreiben*, und vielleicht zu fragen, warum wir das tun.

Dabei finden wir Kripke zufolge, daß wir bei der Zuschreibung von Bedeutungen in der Tat auf die Handlungen und Dispositionen der Mitglieder von Sprachgemeinschaften Bezug nehmen. Doch daraus folgt nicht, daß es die Eigenschaft, eine bestimmte Bedeutung zu haben, wirklich gibt und daß diese Eigenschaft in diesen Handlungen und Dispositionen begründet ist.

Auf der anderen Seite meint Kripke jedoch ähnlich wie viele Wittgensteinianer, daß es einfach keinen Sinn macht, d.h. keinem verstehbaren Zweck dient, den Äußerungen einer isolierten Einzelperson Bedeutungen zuzuschreiben und daß deshalb unsere Bezugnahme auf soziale Praktiken nicht zufällig, sondern in gewisser Weise zwingend ist. Wenn sich die Frage nach der Bedeutung aber nur noch so stellt, daß es um die *Beschreibung* einer Zuschreibungspraxis und um eine *Erklärung* für diese Praxis geht, dann sind vielleicht doch noch Alternativen denkbar.

In diesem Sinn will ich im folgenden untersuchen, ob es nicht doch gute Gründe für die Praxis vieler Kognitionswissenschaftler gibt, bestimmte physische (z.B. neuronale) Strukturen als Repräsentationen aufzufassen, die eine bestimmte Bedeutung haben. Wenn sich herausstellen sollte, daß das in der Tat so ist, wäre meiner Meinung nach damit zugleich gezeigt, daß die Rede von einer Sprache des Geistes (oder

Gehirns) einen vernünftigen Sinn hat – trotz aller Argumente, die Hacker und andere vorgebracht haben.

## 4.

Beginnen möchte ich jedoch mit einem Zugeständnis. Hacker hat in seinen Überlegungen sehr klar gemacht, daß es unserem normalen Gebrauch des Ausdrucks „Sprache“ zufolge eine Sprache nur geben kann, wenn es Wesen gibt, die diese Sprache sprechen, und daß man von einem Wesen nur dann sagen kann, daß es über eine Sprache verfügt, wenn es über ein bestimmtes breit gefächertes Verhaltensrepertoire verfügt.<sup>8</sup> Eines seiner Argumente gegen die Idee einer Sprache des Gehirns war gerade, daß weder das Gehirn noch seine Teile über ein solches Verhaltensrepertoire verfügen *können*. Und damit hat er sicher recht.

Eine Sprache des Geistes kann es daher nur geben, wenn sie sich in bestimmter Hinsicht radikal von allen normalen Sprachen unterscheidet. Eine Sprache des Geistes ist, wenn es sie gibt, nämlich eine Sprache, die von niemandem gesprochen und auch von niemandem verstanden, ja nicht einmal gehört wird. (Wenn manche sagen, das Gehirn spreche oder verstehe diese Sprache, so ist das in der Tat nur metaphorisch zu verstehen.) Eine Sprache des Geistes ist sozusagen eine Sprache, die einfach geschieht. Token von Sätzen dieser Sprache entstehen unter bestimmten Bedingungen im Gehirn, werden dort verändert und bewirken zusammen mit anderen Satztoken bestimmte Handlungen. Die Satztoken müssen nicht geäußert werden, um zu existieren, und sie müssen nicht gehört und verstanden werden, um Wirkungen hervorzubringen. Alles das geschieht – fast möchte man sagen – wie von selbst.

Wenn das so ist, liegt aber in der Tat die Frage nahe, inwiefern man unter diesen Bedingungen überhaupt noch von einer Sprache reden kann. Diese Frage ist sicher berechtigt; und ich bin mir nicht ganz sicher, ob man sie wirklich überzeugend beantworten kann. Aber versuchsweise möchte ich von der folgenden Überlegung ausgehen: Sprache kann man zunächst einmal einfach auffassen als ein System von strukturierten Sätzen mit einer kombinatorischen Semantik. Die Sätze haben eine Bedeutung (Wahrheitsbedingungen), und diese Bedeutung hängt in regelhafter Weise ab von der Bedeutung ihrer Teilausdrücke. Man kann unterscheiden zwischen Satztypen und Satztoken, wobei Satztoken physikalische Strukturen sind, für die man angeben kann, welchen Satztyp sie realisieren. Wenn das so ist, dann kann man

---

<sup>8</sup> Meiner Meinung nach ist es allerdings eine interessante Frage, ob tatsächlich alle Verhaltensweisen, die Hacker anführt, notwendige Bedingungen für das Haben einer Sprache sind oder ob wir nicht auch von Wesen, die nur über einen Teil der von Hacker angeführten Fähigkeiten verfügen, sagen würden (oder sogar müßten), daß sie eine Sprache haben. Auf diese Frage kann ich hier aber leider nicht eingehen.

vielleicht aber auch sagen: Wenn es in einem System eine Menge von physikalischen Strukturen gibt, für die es ausreichende Gründe gibt, sie als Token bestimmter Satztypen mit bestimmten Bedeutungen aufzufassen, dann gibt es in diesem System eine interne Sprache der oben erläuterten Art.

Vielleicht könnte man an dieser Stelle einwenden, daß der Ausdruck „Sprache“ in diesem Zusammenhang unangemessen sei und daß man deshalb eher von „Systemen interner Repräsentationen“ reden solle. Gegen diesen Vorschlag habe ich keine Bedenken. Denn Systeme interner Repräsentationen sind alles, was ein Kognitionswissenschaftler benötigt. Und ich bin sicher, daß kein Kognitionswissenschaftler je die Auffassung vertreten hat, die Sprache des Geistes sei mehr als ein System interner Repräsentationen. Auf der anderen Seite ändert diese sprachliche Klarstellung aber nichts an der argumentativen Situation. Denn Hacker macht mehr als deutlich, daß sich seine Argumente gegen die Annahme von Systemen interner *Repräsentationen* ebenso richten wie gegen die Annahme einer *Sprache* des Geistes.<sup>9</sup> Hacker will ganz allgemein bestreiten, daß es im Gehirn oder in Computersystemen *Bedeutung tragende Strukturen* geben könne. Ich möchte im folgenden dagegen für die These argumentieren, daß es sehr wohl gute Gründe dafür geben kann, bestimmte interne Strukturen als Systeme von Repräsentationen aufzufassen, und daß die Idee einer Sprache des Geistes in diesem Sinn daher durchaus nicht sinnlos ist. Beginnen möchte ich allerdings mit einer sehr allgemeinen wissenschaftstheoretischen Bemerkung.

## 5.

Wenn es darum geht, das Verhalten komplexer Systeme zu erklären und zu verstehen, reicht es häufig nicht aus, nur die – von Dennett<sup>10</sup> so genannte – *physikalische Einstellung* einzunehmen. Ein angemessenes Verständnis ergibt sich vielmehr oft erst, wenn wir auch die *funktionale Organisation* dieser Systeme verstehen. Besonders deutlich wird diese Tatsache im Bereich der Biologie, wo Erklärungen häufig nur auf der funktionalen Ebene gegeben und anatomische und physiologische Details entsprechend kaum mehr erwähnt werden. Nehmen wir irgendein Beispiel – etwa die Temperaturregelung im menschlichen Körper, die im Lehrbuch für *Biologische Psychologie* von Birbaumer und Schmidt so erklärt wird. (Ich fasse diese Darstellung hier stark zusammen.)

---

<sup>9</sup> Vgl. z.B. Hackers These: „Nothing in the cortex constitutes a ‘symbolic representation’ of the creature’s environment.“ (1984, 497).

<sup>10</sup> Die Unterscheidung zwischen physikalischer, funktionaler und intentionaler Einstellung findet sich erstmals in Dennett (1971).

Die Thermoregulation kann formal als ein kreisförmig geschlossenes Regelsystem mit negativer Rückkopplung angesehen werden. Die Körpertemperatur wird von Meßfühlern, nämlich den Thermorezeptoren überwacht, die ihre Meldungen dem zentralen Regler zuführen. Dieser stellt fest, ob die Körpertemperatur (der *Istwert*) von ihrem Sollwert abgewichen ist und verstellt entsprechend über die Aussendung von *Steuersignalen* die Stellgrößen solange, bis die Meßgrößen den Ausgleich der Abweichung signalisieren. (Birbaumer/Schmidt 1990, 117f.)

Die Körperkerntemperatur wird an verschiedenen Stellen durch temperaturempfindliche Nerven- bzw. Sinneszellen gemessen, die Körperschalentemperatur durch Thermosensoren in und unter der Haut. Der Hypothalamus, insbesondere die *Area hypothalamica posterior* wird als Integrationszentrum für die Thermoregulation angesehen. Die zentralen Effektorneurone steuern (wahrscheinlich über eine Kette von Interneuronen) die Stellglieder zur Wärmebildung und -abgabe (Wärmebildung, Isolation der Körperschale, Schweißsekretion, Verhalten). Sie erhalten ihre afferenten Zuflüsse von den äußeren und inneren Thermosensoren. Die Kältesensoren wirken direkt aktivierend auf die Effektorneurone für Wärmebildung und über Interneurone hemmend auf die Effektorneurone für die Wärmeabgabestellglieder. Die Warmsensoren sind genau umgekehrt mit den zwei verschiedenen Typen von Effektorneuronen verschaltet. (op.cit., 119ff.)

Die fast ausschließliche Verwendung funktionalen Vokabulars ist unübersehbar. Es wird von Meßfühlern, Stellgrößen und Regelkreisen geredet ebenso wie von Integrationszentren, Thermosensoren und Effektorneuronen. Die einzigen physiologischen Begriffe scheinen anatomische Bezeichnungen wie „Area hypothalamica posterior“ zu sein. Dabei ließe sich die Geschichte auch ganz anders erzählen. Vereinfacht z.B. so: Wenn im Körper – sagen wir, aufgrund von körperlicher Arbeit – die Temperatur über einen Wert von 37°-38° Celsius steigt, bewirkt das eine erhöhte Entladungsrate bestimmter Neuronen im Körperinneren, die mit ihren Axonen bis in den Hypothalamus reichen. Im Hypothalamus wird durch die erhöhte Feuerungsrate dieser Neurone die Aktivität bestimmter sympathischer und parasympathischer Neurone gedämpft, die über Axone und neuromuskuläre Synapsen mit der glatten Muskulatur der präkapillaren Gefäße verbunden sind. Dies führt zu einer Erschlaffung dieser Muskulatur und damit zur Erweiterung der entsprechenden Gefäße.

Aber ganz unabhängig davon, daß uns diese Geschichte in ihren Einzelheiten nicht vollständig bekannt ist – sie allein würde uns auch nicht reichen. Was uns interessiert, ist nämlich die Frage, wie es unser Körper schafft, einen bestimmten *Zielzu-*

*stand* zu erreichen, wie er es schafft, seine Kerntemperatur *unter sehr verschiedenen Bedingungen* relativ *konstant* zu halten. Und *das* verstehen wir erst, wenn wir sehen, daß die physiologischen Prozesse in Form eines Regelkreises zusammenwirken und daher auch mit Hilfe des entsprechenden Begriffsystems beschrieben werden können. Funktionale Begriffe kommen also besonders dann ins Spiel, wenn es nicht mehr darum geht, einzelne physische Zustände oder Aktivitäten zu erklären, sondern darum, zu verstehen, wie erfolgreiches Handeln zustande kommt, d.h. wie es ein System schafft, unter den verschiedensten Umständen ein Verhalten zu produzieren, das gewissen Standards entspricht. Insgesamt gilt also:

These 1:

Das *erfolgreiche* Verhalten von Systemen können wir häufig nur dann angemessen verstehen und erklären, wenn wir bezüglich dieser Systeme von der physikalischen zur *funktionalen* Einstellung übergehen.

Zur funktionalen Analyse an dieser Stelle noch eine kurze Zusatzbemerkung: Wichtig ist in diesem Zusammenhang auch, daß Eigenschaften wie ein-Meßfühler-zu-sein oder ein-Stellglied-zu-sein keine im herkömmlichen Sinne „natürlichen“ Eigenschaften sind, d.h. daß wir die entsprechenden Begriffe „Meßfühler“ und „Stellglied“ – anders als z.B. Begriffe wie „Pyramidenzelle“ oder „neuromuskuläre Synapse“ – nicht aufgrund von normalen beobachtbaren oder meßbaren neurobiologischen Merkmalen zuschreiben. Denn die Anwendbarkeit dieser Begriffe auf bestimmte neuronale Phänomene hängt davon ab, ob diese Phänomene mit anderen so zusammenwirken, daß sich ein Verschaltungsmuster ergibt, das als „kreisförmig geschlossenes Regelsystem“ *interpretiert* werden kann. Um es auf eine sehr saloppe (und sicherlich auch etwas irreführende) Weise zu formulieren: Funktionale Eigenschaften finden sich nicht in der Welt; sie werden von uns in die Welt hineininterpretiert.

## 6.

Das Beispiel der Thermoregulation im menschlichen Körper ist jedoch zu unspezifisch, als daß man aus ihm etwas über den Sinn oder Unsinn der Idee einer Sprache des Geistes entnehmen könnte. Mit einem zweiten Beispiel kommen wir der Sache aber schon näher – dem Beispiel eines Schachcomputers, das auch von Dennett häufig zur Veranschaulichung herangezogen wird.<sup>11</sup>

Auch bei einem solchen elektronischen Gerät ist es zumindest im Prinzip möglich, jeden einzelnen Zug rein physikalisch zu erklären: Man kann feststellen, wie sich

---

<sup>11</sup> Erstmals in Dennett (1971).

durch den Druck bestimmter Buchstaben- und Zahlentasten bestimmte Teilzustände von Siliziumchips verändern; man kann aus der Verschaltung und den Anfangszuständen dieser Chips ableiten, welche Abfolge von Zuständen sie durchlaufen, nachdem die „Enter“-Taste gedrückt wurde; und auf dieselbe Weise läßt sich schließlich ermitteln, in welchem Zustand das Gesamtgerät am Ende stehen bleibt und welche der Leuchtdioden, aus denen das Display besteht, dann leuchten bzw. nicht leuchten. Was man auf diese Weise erreichen kann, ist aber immer nur die Erklärung konkreter einzelner Endzustände auf der Grundlage des Wissens um konkrete einzelne Anfangsbedingungen. Völlig unzureichend ist dieses Verfahren, wenn man verstehen will, aufgrund welcher Mechanismen es das Gerät schafft, immer wieder outputs zu liefern, denen Züge entsprechen, die in der jeweiligen Spielsituation plausibel oder sogar erfolgreich sind.

Ein solches Verständnis ergibt sich wiederum erst, wenn man von der physikalischen zur funktionalen Einstellung übergeht, was in diesem Fall heißt, daß man das *Programm* analysiert, das dem Verhalten des Schachcomputers zugrundeliegt. Denn erst dann ist es möglich, das, was zwischen input und output passiert, nicht mehr nur als eine Abfolge von Zuständen von Siliziumchips zu sehen. Erst in der funktionalen Einstellung kann man bestimmte Teilzustände dieser Chips als Repräsentationen von möglichen Konfigurationen der Figuren auf dem Schachbrett interpretieren. Und nur unter dieser Voraussetzung kann man das Geschehen zwischen input und output so beschreiben, wie es uns allen inzwischen geläufig ist: Das Gerät berechnet zuerst für die aktuelle Stellung die Repräsentationen aller Folgestellungen, die sich aus den für es selbst möglichen Zügen ergeben; dann zu jeder dieser Folgestellungen die Repräsentationen aller Folgestellungen, die sich aus den jeweils möglichen Zügen des Gegners ergeben; weiter zu jeder dieser Folgestellungen wieder die Repräsentationen aller Folgestellungen, die sich aus den eigenen möglichen Zügen ergeben, usw. bis zu einer durch heuristische Kriterien bestimmten Anzahl von Zügen und Gegenzügen; die einzelnen Folgestellungen werden nach vorgegebenen Kriterien bewertet; und schließlich gibt das Gerät den Zug aus, der bei (nach seinen Kriterien) optimalen Gegenzügen zu der Stellung mit der höchsten Bewertung führt.

Diese Art der Beschreibung ermöglicht nun erstmals ein Verständnis der Tatsache, daß unser Schachcomputer – in der Regel – plausible oder sogar erfolgreiche Züge macht. Denn es läßt sich zeigen, daß die Bewertungsfunktion, die der Zugauswahl zugrundeliegt, unter den jeweiligen Bedingungen tatsächlich zu plausiblen oder gar guten Stellungen führt. Wenn das Gerät zum Schluß den Zug ausgibt, der bei „optimalem“ Gegenspiel zu der am höchsten bewerteten Stellung führt, müssen seine Züge daher in der Regel recht gute Züge sein. Das gilt – wie gesagt – in der Regel; denn es gibt Stellungen, die trotz hoher Bewertung objektiv ungünstig sind, und wenn ein

solcher Fall vorliegt, wählt der Computer häufig keinen besonders guten Zug. Aber das überrascht auch gar nicht. Denn wir wissen natürlich, daß der Computer manchmal Fehler macht. Für die Erklärung des Verhaltens des Computers ist die Beschreibung der Vorgänge zwischen input und output mit Hilfe des skizzierten Programms somit doppelt hilfreich: Sie erklärt uns, warum der Computer in der Regel gute Züge wählt, und sie erklärt uns auch, warum er manchmal furchtbare Fehler macht.

Aber zurück zum Hauptpunkt. Ich hoffe, es ist deutlich geworden, daß wir das Verhalten komplexer Systeme häufig erst dann richtig verstehen und erklären können, wenn wir von der physikalischen zur funktionalen Einstellung übergehen, und das dies insbesondere dann der Fall ist, wenn es sich dabei um ein Verhalten handelt, das – gemessen an bestimmten Standards – als erfolgreich eingestuft werden kann. Wichtiger noch als dieser allgemeine Punkt ist jedoch ein Punkt, der sich ergibt, wenn wir genauer nachfragen, worauf wir denn eigentlich festgelegt sind, wenn wir bezüglich bestimmter Systeme die funktionale Einstellung einnehmen. Auch für diese Frage ist das Schachcomputer-Beispiel wieder außerordentlich instruktiv.

Ich hatte schon gesagt, daß im Hinblick auf einen Schachcomputer die funktionale Einstellung einzunehmen heißt, das Programm zu analysieren, daß in diesem Computer implementiert ist. Und dies wiederum bedeutet zweierlei: 1. daß man bestimmte Teilprozesse, die zwischen input und output ablaufen, als die Ausführung bestimmter Anweisungen auffaßt und 2. daß man rekonstruiert, wie das System die Abfolge dieser Teilprozesse organisiert. Die Ausführung einer Anweisung besteht in der Regel jedoch darin, daß eine bestimmte Datenstruktur geschaffen oder verändert wird. Und das bedeutet, daß wir bestimmte physische Prozesse nur dann als die Ausführung einer Anweisung auffassen können, wenn wir zugleich bestimmte physische Strukturen als Datenstrukturen auffassen. Für die funktionale Analyse unseres Schachcomputers heißt das konkret: Wir können das Programm, das seinem in der Regel erfolgreichen Verhalten zugrundeliegt, nur rekonstruieren, wenn wir bestimmte physische Strukturen im Inneren des Systems – die Teilzustände bestimmter Siliziumchips – als Repräsentationen von möglichen Stellungen und andere physische Strukturen dieser Art als Repräsentationen der Bewertungen von Stellungen auffassen. Wenn wir dieses Ergebnis verallgemeinern, ergibt sich die

These 2:

Die funktionale Analyse eines Systems, die allein einen Beitrag zur Beantwortung der Frage leisten kann, wie es das System schafft, sich in den unterschiedlichsten Situationen erfolgreich zu verhalten, ist in einigen Fällen nur möglich, *wenn man bestimmte physische Strukturen im Inneren des Systems als Repräsentationen auffaßt.*

An dieser Stelle könnte man – im Anschluß an die Argumente Hackers – versucht sein einzuwenden, daß die bisherige Argumentation die Tatsache völlig außer acht lasse, daß es sich bei Schachcomputern um Artefakte handelt, die von ihren Herstellern tatsächlich zu einem bestimmten Zweck programmiert wurden. Von solchen Artefakten könne man deshalb in der Tat sagen, daß in ihnen Programme ablaufen und daß es in ihnen daher auch so etwas wie Repräsentationen gebe; denn in diesem Falle gebe es jemanden – nämlich den Programmierer, der mit bestimmten physischen Strukturen bestimmte Konfigurationen von Schachfiguren repräsentieren wolle. Repräsentationen ohne eine Person, die sie verwende, seien jedoch begrifflich unmöglich.

Dieser Einwand gehe an meiner Argumentation jedoch völlig vorbei. Denn der Hauptpunkt meines Argument ist, daß wir bei manchen Systemen – *ganz unabhängig davon, wie sie entstanden sind* – annehmen müssen, daß es in ihnen Repräsentationen gibt, wenn wir verstehen wollen, wie das erfolgreiche Verhalten dieser Systeme zustandekommt. Schachcomputer müßten wir unter dieser Voraussetzung in funktionaler Einstellung also auch dann genau so beschreiben, wie ich es oben erläutert habe, wenn sie auf Bäumen wachsen würden.

In der Tat kann man an vielen Beispielen nachweisen, daß in der Neurobiologie genau diese Erklärungsstrategie verfolgt wird. Ich erinnere mich noch sehr gut an eine Diskussion, in der ich den Göttinger Physiker und Akustikexperten Manfred Schroeder einmal gefragt habe, welche neuronalen Mechanismen denn für die Lokalisation von Schallquellen verantwortlichen seien. Schroeders Antwort begann mit dem Satz: „Zunächst einmal wird im Gehirn die Kreuzkorrelation der Signale der beiden Hörnerven errechnet.“ Ein anderes Beispiel derselben Art findet sich in einem Artikel von J. Koenderink mit dem Titel „The Brain a Geometry Engine“. Denn soweit ich diesen Artikel verstanden habe, ist seine Hauptthese, daß man die Mechanismen des visuellen Kortex am besten versteht, wenn man von der zweidimensionalen Intensitätsverteilung des auf die Retina fallenden Lichts ausgeht und die anschließende neuronale Verarbeitung so interpretiert, daß in ihr die ersten, zweiten und weitere höherstufige Ableitungen dieser Verteilung errechnet werden.

... you may understand a large part of the structure of the front-end visual system as an embodiment of differential geometry of the visual field. ... Instead of the concrete ‘edge detectors’ and ‘bar detectors’, one speaks of the abstract first- and second-order directional derivatives. (1990: 125)

Ich kann hier auf weitere Details nicht eingehen; aber ich denke, daß schon an diesen nur angedeuteten Beispielen klar wird, daß viele Neurobiologen bei dem Versuch, die erstaunlichen Leistungen des Gehirns zu erklären, tatsächlich die funktionale

Einstellung einnehmen und daß sie dabei darüberhinaus tatsächlich Erklärungen auf der Basis der Annahme anstreben, daß im Gehirn bestimmte *Berechnungen* durchgeführt werden. Rechenprozesse aber sind Prozesse der Veränderung von Zahlzeichen; also Prozesse, bei denen Repräsentationen numerischer Werte manipuliert werden. Wenn man annimmt, daß im Gehirn Rechenprozesse stattfinden, ist man damit also auch auf die Annahme festgelegt, daß es im Gehirn ein System von Repräsentationen gibt.

## 7.

Damit ist im Grunde alles Wesentliche gesagt. Genauso wie es bei Schachcomputern notwendig ist anzunehmen, daß es in ihnen Repräsentationen von Stellungen und Repräsentationen von Bewertungen gibt, wenn man verstehen will, wie diese Geräte es fertigbringen, einigermaßen erfolgreiche Züge zu produzieren, kann es im Hinblick auf andere Systeme notwendig sein anzunehmen, daß es in ihnen satzartige Repräsentationen von ihrer Umwelt oder von allgemeinen Gesetzmäßigkeiten gibt, wenn man verstehen will, worauf das erfolgreiche Verhalten dieser Systeme beruht. Dies würde z.B. für alle KI-Systeme gelten, deren Problemlösungsverhalten auf der Ableitung von Formeln aus einer Menge von Axiomen beruht. Denn das Verhalten dieser Systeme können wir nicht angemessen verstehen, wenn wir nicht einige der in ihnen ablaufenden Prozesse als Inferenzprozesse interpretieren. Und Inferenzprozesse sind Prozesse der Ableitung von satzartigen Repräsentationen aus satzartigen Repräsentationen. D.h., wir können nicht einige der in einem System ablaufenden Prozesse als Inferenzprozesse auffassen, wenn wir nicht zugleich einige der in ihm vorkommenden physischen Zustände als satzartige Repräsentationen interpretieren. Hier zeigt sich ebenso wie bei dem zuvor erläuterten Schachcomputer-Beispiel ein Primat der Interpretation von Prozessen. Bestimmte Prozesse in einem System können wir nicht angemessen verstehen, wenn wir nicht zugleich bestimmte Zustände als Repräsentationen auffassen.

Und damit ist, denke ich, auch klar, unter welchen Bedingungen wir gewissermaßen sogar gezwungen sind anzunehmen, daß es in bestimmten Systemen satzartige Repräsentationen oder Symbole gibt, d.h. daß diese Systeme eine Sprache des Geistes enthalten. Dies ist nämlich genau dann der Fall, wenn wir nur dann richtig verstehen können, wie das erfolgreiche Verhalten dieser Systeme zustandekommt, wenn wir einige der in ihnen ablaufenden physischen Prozesse als funktionale Prozesse auffassen, die ihrerseits nur als Prozesse der Erzeugung und Veränderung satzartiger Repräsentationen verstanden werden können. Noch einmal in einer These zusammengefaßt:

These 3:

Die Annahme *satzartiger Repräsentationen* ist nicht nur plausibel, sondern in gewisser Weise sogar unumgänglich, wenn wir das *erfolgreiche* Verhalten eines Systems nur durch die Annahme erklären können, daß es auf *funktionalen* Prozessen beruht, die nur als *Prozesse der Erzeugung und Veränderung satzartiger Repräsentationen* verstanden werden können.

Die Redeweise von satzartigen Repräsentationen bzw. von einer Sprache des Geistes ist also weder eine Marotte gewisser Kognitionswissenschaftler noch gar eine Marotte, die auf einer fundamentalen begrifflichen Konfusion beruht. Sie ist vielmehr eine zwingende Konsequenz, die sich bei dem Versuch ergeben kann, die funktionale Architektur eines Systems zu verstehen, auf der das erfolgreiche Verhalten dieses Systems beruht.

Zum Abschluß möchte ich aber sehr nachdrücklich betonen, daß in der These 3 nur eine *Bedingung* formuliert ist. *Wenn* diese Bedingung erfüllt ist, dann kann man davon sprechen, daß es in einem System eine Sprache des Geistes gibt. Aus dieser These ergibt sich also *nicht*, daß Fodor recht hat oder daß die Kognitionswissenschaftler recht haben, die glauben, daß intelligentes Verhalten nur im Rahmen des Symbolverarbeitungsansatzes adäquat erklärt werden kann. Aber mir ging es auch nicht darum, diesen Ansatz als *sachlich angemessen* zu verteidigen, sondern nur darum, ihn vor dem Vorwurf der Begriffsverwirrung in Schutz zu nehmen.

## Literatur

- Beckermann, A. (1991): Der endgültige Todesstoß für den Repräsentationalismus? – Eine Replik auf Andreas Kemmerlings Artikel ‘Mentale Repräsentationen’. In: *Kognitionswissenschaft* 2, 91-98.
- Birbaumer, N./Schmidt R.F. (1990): *Biologische Psychologie*. Berlin/Heidelberg/New York.
- Boghossian, P. (1989): The Rule-Following Considerations. In: *Mind* 98, 507-549.
- Dennett, D. (1971): Intentional Systems. In: *Journal of Philosophy* 68, 87-106. Wiederabgedr. in: Dennett, D. (1978): *Brainstorms*. Montgomery, Verm., 3-22.
- Fodor, J.A. (1975): *The Language of Thought*. New York.
- Fodor, J.A. (1978): Propositional Attitudes. In: *The Monist* 64, 501-523. Wiederabgedr. in: Fodor (1981a), 177-203.
- Fodor, J.A. (1981a): *Representations*. Cambridge, Mass.

- Fodor, J.A. (1981b): Introduction – Something on the State of the Art. In: Fodor (1981a), 1-31.
- Fodor, J.A. (1987): *Psychosemantics*. Cambridge, Mass.
- Hacker, P. (1987): Languages, Minds and Brains. In: C. Blakemore and S. Greenfield (eds.) *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*. Oxford, 485-505.
- Harman, G. (1973): *Thought*. Princeton, NJ.
- Hart, H.L.A. (1961): *The Concept of Law*. Oxford.
- Hyman, J. (ed.) (1991): *Investigating Psychology*. London/New York.
- Koenderink, J.J. (1990): The Brain a Geometry Engine. In: *Psychological Research* 52, 122-127.
- Kripke, S. (1982): *Wittgenstein on Rules and Privat Language*. Cambridge.
- Savigny, E. von (1983): *Zum Begriff der Sprache*. Stuttgart.
- Stillings, N.A. et al. (1987) *Cognitive Science – An Introduction*. Cambridge, Mass.