

Geert Keil

## Homunkulismus in den Kognitionswissenschaften

### 1. Was ist ein Homunkulus-Fehlschluß?<sup>1</sup>

Das lateinische Wort „homunculus“ bedeutet „kleiner Mensch“. Es ist mit der im 13. Jahrhundert aufgekommenen alchimistischen Vorstellung verbunden, man könne ein menschenähnliches Wesen auf chemischem Wege im Labor erzeugen. In der Philosophie des Geistes wird der Ausdruck jedoch in einem anderen Sinne verwendet. Es geht dort nicht um wirkliche kleine Menschen, sondern um ausgedachte, nämlich um postulierte menschenähnliche Instanzen, die ausdrücklich oder unausdrücklich zur Erklärung der Arbeitsweise des menschlichen Geistes herangezogen werden. Daß ein geistbegabtes Wesen ein bestimmtes Vermögen besitzt, wird dadurch erklärt oder analysiert, daß einer seiner Teile oder ein Subsystem in ihm dieses Vermögen besitzt.<sup>2</sup>

Vielen Philosophen gilt das Postulieren eines Homunkulus als *Fehlschluß*, mithilfe dessen eine Erklärung erschlichen werden soll. Den Ausdruck „homunculus fallacy“ hat Anthony Kenny 1971 mit den folgenden Worten eingeführt:

„I shall call the reckless application of human-being predicates to insufficiently human-like objects the ‘homunculus fallacy’, since its most naïve form is tantamount to the postulation of a little man within a man to explain human experience and behaviour“ (Kenny 1971, 155).

<sup>1</sup> Die ersten Seiten dieses Aufsatzes enthalten einen Rückblick auf meinen Beitrag „Über den Homunkulus-Fehlschluß“ (Keil 2003).

<sup>2</sup> Es gibt noch einen weiteren Sinn von „Homunkulus“: In der Neurophysiologie ist der Homunkulus die Darstellung einer verzerrten menschlichen Figur, deren Proportionen der Menge des kortikalen Gewebes entsprechen, das für die einzelnen sensorischen und motorischen Funktionen des Menschen ‘zuständig’ ist. Auch die je nach Körperteil unterschiedlich ausgeprägten Fähigkeiten der sensorischen Unterscheidung und der Bewegungssteuerung entsprechen diesen Proportionen. Der Homunkulus hat beispielsweise eine große Zunge, große Füße und riesige Hände, aber einen kleinen Rumpf und kleine Beine. (vgl. Penfield/Boldrey 1937)

Nun könnten die „human-being predicates“, von denen Kenny spricht, auch solche sein, die sich auf äußerliche Attribute des Menschen beziehen, beispielsweise auf seine körperliche Gestalt. Da solche Attribute für die Philosophie des Geistes nicht einschlägig sind, wird häufig eine engere Charakterisierung des Homunkulus-Fehlschlusses gewählt. Der Fehlschluß bestehe darin, daß Prädikate, die auf kognitive oder perzeptive Leistungen einer ganzen Person zutreffen, auch auf Teile von Personen oder auf subpersonale Vorgänge angewendet werden.

Ob es sich bei dieser Praxis um einen Fehlschluß handelt, ist umstritten. Ein Fehlschluß im philosophischen Sprachgebrauch ist die Anwendung eines ungültigen Schlußschemas. Nun ziehen Philosophen, denen man den Homunkulus-Fehlschluß vorwirft, typischerweise nicht ausdrücklich den Schluß: Auf eine Person trifft das Prädikat F zu, also trifft es auch auf Teile der Person zu. Nach meiner Rekonstruktion (vgl. Keil 2003, 3) besteht der Kern des Homunkulus-Einwands nicht darin, daß der Opponent eine personenähnliche Instanz innerhalb einer Person annimmt, sondern daß er eine annehmen *müßte*, um die Verwendung der gewählten Prädikate zu rechtfertigen.

In der Philosophie des Geistes und in der Psychologie werden seit jeher *Instanzenmodelle der Psyche* vertreten. Platon und Freud beschrieben die Arbeitsweise einzelner Seelenteile in Analogie zur Arbeitsweise von Personen. Sie taten es sehenden Auges und glaubten, gute Gründe dafür zu haben. In Platons Gleichnis des Seelenwagens wird die Seele als aus zwei Pferden und einem Wagenlenker zusammengesetzt vorgestellt. Das eine Pferd sei gut, edel und folgsam, das andere aber „entgegengesetzter Abstammung und Beschaffenheit“, wild und starrsinnig, „taub, der Peitsche und dem Stachel kaum gehorchend“. „Schwierig und mühsam ist daher notwendig bei uns die Lenkung“ (*Phaidros* 246 b und 253 d). Nichtmetaphorisch benennt Platon die drei Seelenteile als das Begehrende, das Eifernde und das Überlegende.

Freud hatte zunächst ein hydraulisches Modell der Psyche entworfen, in dem die „psychischen Energien“ wie in einem Röhrensystem hin- und herfließen, weil sich an bestimmten Stellen Druck aufbaut, der an anderer Stelle wieder abgebaut werden muß. Die hydraulischen und mechanischen Redeweisen hat Freud nie aufgegeben, ergänzt sie allerdings später durch das topische Instanzenmodell, erst durch das Zweiinstanzenmodell von bewußtem und unbewußtem System, seit 1923 dann durch das Dreiinstanzenmodell von Ich, Es und Über-Ich. Man kann in den Texten zeigen, wie die Instanzenmodelle aus einer *soziomorphen Uminterpretation* des hydraulischen Modells entstehen (vgl. Keil 1993, 128-145). Die psychischen Energien werden nun personifiziert, und Freud spricht von den Instanzen bei-

spielsweise als von den streitenden Parteien in einem Bürgerkrieg. Ein Kritiker stellt Freuds Instanzenmodell so dar:

„Wir benennen eine Reihe von Instanzen oder Abteilungen im sogenannten psychischen Apparat einer Person und sprechen von diesen so, als ob jede von ihnen sich nach Art eines Individuums verhielte, mit einer umschriebenen Gruppe von Zielen, mit einer bestimmten Art und Menge von Energie, mit einer Strategie und einem bestimmten Einfluß. [...] Jedenfalls wird die Seele als eine potentielle oder tatsächliche Vielheit mehr oder weniger aufsässiger, mehr oder weniger weisungsgebundener oder weisungsberechtigter Individuen behandelt“ (Schafer 1982, 181).

Worin genau besteht nun der Einwand gegen homunkulare Redeweisen? Der Standardeinwand lautet, daß das Postulieren subpersonaler menschenähnlicher Instanzen nicht erklärungskräftig ist, weil es auf einen *Regreß* hinausläuft. Der frühe Dennett formuliert diesen Einwand wie folgt:

Die „roles played by the little man in the brain are merely specialized roles projected inwards from the details of our initial analysanda, the variety of affairs of a person. The solitary audience in the theatre of consciousness, the internal decision-maker and source of volitions or directives, the reasoner, if taken as *parts* of a person, serve only to postpone analysis.“

Homunkulare Redeweisen haben somit „the disadvantage of merely postponing the central problem before us by positing unanalysed man-analogues as systematic elements in that which we are trying to analyse, namely Man. The ‘little man in the brain’, Ryle’s ‘ghost in the machine’, is a notorious non-solution to the problems of mind“ (Dennett 1969, 190 und 87).

Diese klaren Worte zum Regreßproblem sind bemerkenswert, weil Dennett in späteren Jahren zu den entschiedensten Verteidigern homunkularer Erklärungen kognitiver Leistungen gehört. Auf seine offensive Rehabilitierung von Homunkuli im Rahmen des „homunkularen Funktionalismus“ werde ich unten eingehen.

## 2. Analyse des Mentalen und Naturalisierung der Intentionalität

Man kann sich fragen, was eigentlich die treibende Kraft hinter dem Postulieren personenähnlicher psychischer Instanzen ist. Kann man beispielsweise dem Schließenden die allgemeine Überzeugung zuschreiben, daß, was vom Ganzen prädiert werden kann, auch von seinen Teilen prädiierbar ist? Kenny meint in der Tat, daß die Homunkulisten schließen: ‘Was

vom Menschen präzifizierbar ist, ist auch von seinem Gehirn präzifizierbar'.<sup>3</sup> Daß ein solcher Schluß stattfindet, ist indes psychologisch wenig plausibel. Warum sollte jemand diesen Übergang für korrekt halten? Wir schließen ja auch nicht aus der Tatsache, daß Autos rollen können, darauf, daß Motorhauben oder Benzintanks rollen können. Wenn wir also bei Personen und Teilen von Personen geneigt sein sollten, so zu schließen, muß das einen besonderen Grund haben, der über den generellen und als solchen unplausiblen Schluß vom Ganzen auf die Teile hinausweist.

Ich vermute, daß die treibende Kraft hinter dem Homunkulus-Fehlschluß sich aus zwei Komponenten zusammensetzt. Die erste ist die Suche nach einer nichtintentionalen Analyse oder Erklärung eines intentionalen Phänomens, also das, was man in der Philosophie des Geistes das Programm der Naturalisierung der Intentionalität nennt. Um dabei nicht auf einen physikalistischen oder behavioristischen Begriffsrahmen beschränkt zu sein, wird das Naturalisierungsprogramm mit der repräsentationalistischen Annahme kombiniert, daß im Geist (oder im Gehirn) symbolverarbeitende Prozesse stattfinden.

Geistige Leistungen beschreiben wir gewöhnlich vom intentionalen Standpunkt aus, also im Begriffsrahmen der *belief/desire*-Psychologie. Eine tief sitzende Überzeugung vieler Kognitionswissenschaftler und ihrer Wegbegleiter in der Philosophie des Geistes ist indes, daß die intentionale Charakterisierung mentaler Leistungen und Fähigkeiten als Leistungen und Fähigkeiten ganzer Personen nicht das letzte Wort sein darf. Solche Fähigkeiten müssen sich auf zugrunde liegende natürliche Prozesse zurückführen lassen, die *in* Personen stattfinden. Warum? Weil Intentionalität einfach nicht der Stoff ist, aus dem die Welt gemacht ist. Die Annahme nicht weiter analysierbarer intentionaler Einstellungen würde die intentionale Psychologie vom Rest der Wissenschaften isolieren, und sie würde *uns* vom Rest der Natur isolieren. Das Beharren auf der nichtanalysierbaren Rede von Wünschen, Überzeugungen und Absichten würde es unmöglich machen, Personen als Teil der einen natürlichen Welt zu betrachten, deren restliche Teile schließlich nach naturalistischer Auffassung keine intentionalen Eigenschaften haben. Zu diesem naturalistisch motivierten Mißtrauen gegen den Realitätsgehalt nichtanalysierter intentionaler Zuschreibungen lassen sich zahllose Belege anhäufen; hier muß ein einschlägiges Zitat von Jerry Fodor genügen:

<sup>3</sup> „The inappropriate use of predicates, not being a form of argument, is not strictly a fallacy, as I observed. But it leads to a form of argument [...] which is fallacious in the strict sense of the word: the argument that because a certain human-being predicate attaches to a human being it attaches to his brain“ (Kenny 1971, 164).

„I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. [...] If aboutness is real, it must be something else" (Fodor 1987, 97).

Intentionalität und intentionale Einstellungen sind keine Bestandteile der physischen Welt. Wenn irgend möglich, möchte der Naturalist sie auf Konstituenten zurückführen, die nicht selbst intentionale Eigenschaften haben. Dafür müßte der sogenannte Zirkel der Intentionalität durchbrochen werden, also die Praxis, einen intentionalen Ausdruck jeweils durch andere intentionale Ausdrücke zu erläutern. Nun sind aber im Umkreis der Kognitionswissenschaften die meisten Autoren der Auffassung, daß bestimmte Arten von intentionalen Erklärungen einstweilen unersetzlich sind. Die vielbeschworene „kognitive Wende“ gegenüber dem Behaviorismus und dem Physikalismus besteht gerade darin, daß die Unersetzlichkeit von in intentionalen Begriffen formulierten Erklärungen menschlichen Verhaltens anerkannt wird.

In vielen kognitionswissenschaftlichen Theorien ist nun zu beobachten, daß Konzepte, deren Verwendung man als erklärungskräftig erkannt hat, auf einer subpersonalen Ebene noch einmal verwendet werden. Ob sie dort immer noch erklärungskräftig sind, ist natürlich nicht ausgemacht. Die Annahme, daß sie es sind, beruht auf der repräsentationalistischen Theorie des Geistes, welche die Weiterverwendung eines intentionalen oder intentionoiden Begriffsrahmens rechtfertigen soll. Subpersonalen Instanzen werden in kognitionswissenschaftlichen Theorien selten umstandslos Wünsche, Überzeugungen und Absichten zugeschrieben. Vielmehr werden Konzepte wie *Regelbefolgung*, *Repräsentation*, *Berechnung* und *Informationsverarbeitung* verwendet. Diese Konzepte sind alle nicht offen intentional, wohl aber nach Auffassung einiger Kritiker *intentionalitätspräsupponierend*, d.h. sie beziehen ihren Sinn daraus, daß in ihrer Umgebung noch intentionale Phänomene unanalysiert geblieben sind (vgl. dazu Keil 2000a, bes. 193-204). Der Vorwurf lautet dann, daß auch die genannten Begriffe im wörtlichen Sinne allein auf kognitive Tätigkeiten ganzer Personen zutreffen, bei Strafe des Homunkulus-Fehlschlusses. (Der Homunkulus-Fehlschluß wird gelegentlich auch als „intentional fallacy“ bezeichnet.)

Ein scharfer Kritiker des Homunkulismus in den Kognitionswissenschaften ist John Searle. Er spießt Fragen auf wie diese: „Wie errechnet das visuelle System aus Schattierungen Gegenstandskonturen?“ Oder: „Welche Symbol-

manipulation findet im Gehirn statt, wenn wir 6 mit 8 multiplizieren?“ Oder: „Wie berechnet das Gehirn aus zweidimensionalen Projektionen auf der Netzhaut eine dreidimensionale Beschreibung der Welt?“ Und Searle behauptet, daß diese Fragen ebenso sinnvoll seien wie die Frage: „Wie berechnet ein Nagel aus dem Schlag des Hammers die Distanz, die er im Holz zurücklegen muß?“ (vgl. Searle 1990, 28 ff.). Doch während die letztere Frage offenkundig absurd ist, füllen Fragen der ersten Art kognitionswissenschaftliche Standardwerke. Nach Searle stellen aber weder Nägel noch Gehirne im Wortsinne Berechnungen an; das täten allein Menschen. Solche Behauptungen sind natürlich heiß umstritten, und sie werfen die Frage auf, *warum* eigentlich bestimmte Prädikate für die Beschreibung der kognitiven Tätigkeiten ganzer Personen reserviert sein sollten. Soviel scheint klar zu sein, daß die von Searle beanstandeten Redeweisen nicht offen homunkular sind. Wenn sie zu beanstanden sind, dann deshalb, weil sie auf der *versteckten* Annahme einer personalen oder personähnlichen Instanz mit intentionalen Zuständen beruhen, eben auf postulierten Homunkuli in dem Sinne, daß der Opponent eine personenähnliche Instanz annehmen *mißte*, um die Verwendung der gewählten Prädikate zu rechtfertigen. Kenny hatte den Ausdruck „Homunkulus-Fehlschluß“ unter anderem mit der Formulierung eingeführt: „the reckless application of human-being predicates to insufficiently human-like objects“. Hier drängt sich die Frage auf, was ein „nicht hinreichend menschenähnlicher“ Gegenstand ist. Und woran bemißt sich dies? Wittgenstein hat einmal dekretiert: „Wir sagen nur vom Menschen, und was ihm ähnlich ist, es denke“ (PU § 360). Diesem Dekret wird aus kognitionswissenschaftlicher Sicht mit einigem Recht entgegengehalten, daß der Sinn der Frage, ob maschinales Verhalten mit mentalen Prädikaten beschrieben werden kann, verfehlt wird, wenn man die Antwort einfach unter Verweis auf den gewöhnlichen Sprachgebrauch stipuliert. Die funktionalistischen Auffassungen des Mentalen zeichnen sich gerade durch die Annahmen der *Abstraktheit* und der *multiplen Realisierbarkeit* mentaler Zustände aus, und aus dieser Sicht erscheint das Beharren auf Personen als den einzigen Trägern mentaler Prädikate als dogmatische Gegenbehauptung. Doch wenn ich recht sehe, kann der Homunkulismus-Kritiker die schwierige Frage nach dem originalen Anwendungsbereich des intentionalen Idioms zunächst dahingestellt sein lassen. Er muß lediglich daran festhalten, seinen Einwand im Sinne des Regreßargumentes zu formulieren. Wenn der Opponent den Anspruch erhoben hat, die kognitive Fähigkeit X reduktiv zu erklären oder zu analysieren, dann kann diese Analyse sich nicht darin erschöpfen, im Innern des fraglichen Systems ein Subsystem mit eben dieser Fähigkeit X zu identifizieren.

### 3. Homunkulismus in Theorien der visuellen Wahrnehmung

Ein *locus classicus* für Homunkulus-Fehlschlüsse sind Theorien der visuellen Wahrnehmung, in denen der Sehprozess als ein Auswerten oder Interpretieren von Netzhautbildern analysiert wird. Worin hier der Fehler besteht, zeigt sich am eindrucklichsten am Problem des invertierten Netzhautbildes: *Wie kommt es, daß wir die Gegenstände so sehen, wie wir sie sehen, nämlich aufrecht, wo doch das Netzhautbild auf dem Kopf steht? Warum führt ein invertiertes Netzhautbild nicht zu einem invertierten Wahrnehmungseindruck?* Johannes Kepler, der dieses Problem Anfang des 17. Jahrhunderts formulierte, vermutete zunächst, es müsse irgendwo im Glaskörper des Auges eine zweite Linse geben, die das verkehrte Netzhautbild wieder aufrecht stellt. Er bemerkte nicht, daß schon der Problemstellung ein Homunkulus-Fehlschluß zugrunde liegt. Die Frage, wo die Invertierung wieder aufgehoben wird, entsteht allein unter der Annahme, daß auf der Netzhaut etwas ist, was gesehen wird.<sup>4</sup> Diese Annahme ist eine Implikation des Bildbegriffs selbst: Ein Bild ist etwas, was gesehen werden kann. Tatsächlich befindet sich aber in unserem Kopf niemand mit einem zweiten Paar Augen, der von hinten unsere Retina betrachtet.<sup>5</sup>

Die Behauptung, daß schon die Rede von Netzhautbildern irreführend sei, empfinden viele Theoretiker der visuellen Wahrnehmung als Provokation, weil sie es für einen empirischen Befund halten, daß die von einem äußeren Gegenstand reflektierten und auf die Retina auftreffenden Lichtstrahlen dort ein Bild des Gegenstandes erzeugen. Das Auge sei eine *camera obscura* mit Linse; die entsprechende Darstellung des Strahlengangs findet sich in allen Lehrbüchern. An der Darstellung des Strahlengangs ist natürlich nichts auszusetzen, wohl aber an der Rede vom Netzhautbild. Den empirischen Be-

<sup>4</sup> Das Netzhautbild ist nicht zu verwechseln mit den in Theorien mentaler Repräsentationen postulierten „inneren“ oder „geistigen Bildern“, die ebenfalls ein Homunkulus-Problem erzeugen. Kurz gesagt: Wer sich visuell etwas *vorstellt*, stellt sich kein *Bild* vor (außer bei Gemälden), sondern den Gegenstand selbst. Zwar wird an Stelle von „sich etwas vorstellen“ manchmal der Ausdruck „ein Bild vor Augen haben“ verwendet, aber gerade weil beide Redeweisen eine ähnliche Grammatik haben, sind Bilder nicht zugleich das, *was* sich jemand vorstellt. Die Verwendung des Bildbegriffs an Stelle von „Vorstellung“ ist schon deshalb nicht angezeigt, weil es Vorstellungen auch in anderen Sinnesmodalitäten gibt: Man kann sich einen Geschmack oder einen Geruch vorstellen. Ryle weist darauf hin, daß in diesen Fällen eine Original/Kopie-Redeweise, wie sie der Rede von vorgestellten Bildern zugrunde liegt, nicht zur Verfügung steht (vgl. Ryle 1949, 345). Daß es die Original/Kopie-Redeweise bei einigen Sinnesmodalitäten nicht gibt, ist ein Indiz dafür, daß sie auch bei visuellen Vorstellungen irreführend und ein sprachliches Oberflächenphänomen ist.

<sup>5</sup> Das Problem des invertierten Netzhautbildes habe ich in (Keil 2003) genauer diskutiert.

fund sollte man neutraler so formulieren, daß auf der Netzhaut ein *Bestrahlungsmuster* entsteht. Den begrifflichen Unterschied zwischen „Bild“ und „Bestrahlungsmuster“ zu würdigen fällt vielen Wahrnehmungstheoretikern schwer. Mit dem Bildbegriff operierende Theorien fallen auch nicht zwangsläufig dem Homunkulus-Fehlschluß zum Opfer. Das vermeintliche Invertierungsproblem gibt einen guten Testfall dafür ab, ob es sich um eine bloße *façon de parler* oder um eine begriffliche Verwirrung handelt: Die fehlgeleitete Frage, warum wir trotz des invertierten Bildes die Gegenstände richtigerum sehen, entsteht allein unter der Annahme, daß das Netzhautbild ein phänomenal zugänglicher Teil des Wahrnehmungsprozesses ist. Wer hier einen Erklärungsbedarf annimmt, hat sich von der Implikation des Bildbegriffs irreführen lassen.

Daß das Invertierungsproblem nicht gelöst werden muß, wird heute weithin anerkannt. Die Homunkulismus-Kritiker Hacker und Hyman sind allerdings der Auffassung, daß es auch den Vertretern der neueren computationalen Wahrnehmungstheorien vielfach nicht gelingt, sich vollständig von der Bildanalogie zu befreien. Aktuelle Wahrnehmungstheorien, die u.a. mit der Erklärung von *Konstanzphänomenen* befaßt sind, formulierten ihre Explananda nach wie vor in homunkulusgefährdeter Weise. So wird das Problem der Größenkonstanz häufig in der Frage ausgedrückt: Wie errechnet das Sehsystem aus der veränderlichen Größe des retinalen Bildes die tatsächliche Größe des gesehenen Gegenstands?

Der Technologiesprung von der analogen zur digitalen Bildverarbeitung hat indes die Verwendung des Bildbegriffs verändert. Wurde die Funktionsweise des menschlichen Auges in Lehrbüchern bislang am Beispiel des Fotoapparates erläutert, so heißt es nun, das Auge funktioniere wie eine digitale Videokamera. In digitalen Bildaufzeichnungsgeräten wird das einfallende Licht in Form elektrischer Ladung auf einer Punktmatrix gespeichert, bevor die Ladungen der einzelnen Punkte in ein digitales Signal umgewandelt, weitergeleitet und gespeichert werden. In der Punktmatrix ist die visuelle Information noch in einer dem einfallenden Licht entsprechenden räumlichen Anordnung vorhanden, während die digitalisierte Information nur noch in Form einer Liste von Einträgen vorliegt.

Viele Autoren nennen sowohl die Punktmatrix als auch die digitale Datei wiederum ein Bild. Zur Rechtfertigung wird angeführt, daß „something does not have to *be* a picture to *function* as one“ (Kosslyn 1983, 22). Wenn in einem Computer die Liste von Einträgen zeilenweise ausgelesen und weiterverarbeitet wird, bleibe die ursprüngliche räumliche Information schließlich erhalten: „The computer identifies these elements in a way that results in their functioning *as if* they were arranged in a visual array“ (ebd., 23).

Diese Rechtfertigung für die Beibehaltung des Bildbegriffs geht indes an demjenigen Merkmal des Bildbegriffs vorbei, das für das Homunkulus-Problem verantwortlich ist. Nötig wäre eine Auseinandersetzung mit der Frage gewesen „Wer soll dieses Bild sehen?“ Etwas, was nicht gesehen werden kann, fungiert eben in entscheidender Hinsicht *nicht* als ein Bild.

Um genauer zu sehen, welche Gestalt das Homunkulus-Problem in der „computational theory of vision“ annimmt, sei ein Blick auf die Pionierarbeit von David Marr geworfen. Sie beginnt mit einer Definition des Sehens:

„What does it mean to see? The plain man’s answer (and Aristotle’s too) would be, to know what is where by looking. In other words, vision is the *process* of discovering from images what is present in the world, and where it is.“ (Marr 1982, 3)

Diesen „process of discovering from images what is present“ analysiert Marr wie folgt: Das Rohbild, ein zweidimensionales Bestrahlungsmuster auf der Netzhaut, wird in einen „primal sketch“ umgewandelt, der aus einer Beschreibung geometrischer Eigenschaften des zweidimensionalen Bildes besteht. Über einen weiteren Zwischenschritt wird dann der Output des Sehsystems erzeugt: eine Beschreibung von dreidimensionalen Objekten der Außenwelt und ihrer räumlichen Anordnung (vgl. ebd., 37).

Marr sagt nirgends, *wer* die Information aus dem Netzhautbild extrahiert und *wem* die Beschreibung des zweidimensionalen Musters gegeben wird. Benötigt er ein Satzsubjekt, so wählt er meist „the visual system“. Dieses „System“ hat beachtliche Fähigkeiten: Es kann unter anderem etwas herausfinden, Bilder auswerten (ansehen?), Eigenschaften beschreiben und etwas interpretieren. Searle, Hacker, Pessoa/Thomson/Noë und andere haben Marr deshalb Homunkulus-Fehlschlüsse vorgeworfen. Allerdings gibt es eine interessante Verschiebung: Die Rede vom Netzhautbild spielt in der computationalen Wahrnehmungstheorie keine prominente Rolle mehr. Die Arbeitsweise subpersonaler Systeme wird nun im Idiom der Informationsverarbeitung beschrieben. Mit der Rede von *Information*, *Beschreibung*, *Interpretation* und *Repräsentation* werden einzelne kognitive Module (in den Neurowissenschaften auch physische Teile des Gehirns) als Gesprächsteilnehmer aufgefaßt, die einander Nachrichten zukommen lassen, Informationen austauschen und derart ihr Verhalten aufeinander abstimmen. Diese Metaphorik ist nicht erst durch die Computertechnologie aufgekommen, sondern entstammt der Informationstheorie und Nachrichtentechnik der Jahrhundertmitte. Schon Ryle kritisiert in einem Aufsatz zur Wahrnehmung die Vorstellung, „daß es sich bei Augen, Nasen und Ohren um eine Art Auslandskorrespondenten handelt“, die uns mehr oder weniger zutreffende

Nachrichten schicken (Ryle 1954, 124).<sup>6</sup> Eklatant homunkular ist auch die in der Interpretation der Libet-Experimente zur Datierung von Willensimpulsen verbreitete Rede davon, mein Gehirn „entscheide“ sich an meiner Statt zur Ausführung einer Körperbewegung und „benachrichtige“ mich (oder mein Bewußtsein) dann von seiner Entscheidung. Der Schritt zur unfreiwilligen Komik ist überschritten bei denjenigen Autoren, die sich in die Lage ihres Gehirns versetzen und dann ihre tristen Lebensbedingungen beklagen: „So muß ich, ein Stück weiße und graue Materie, eingeschlossen in die Höhle eines knöchigen Schädels, nun meine Situation erkennen. Alles, was ich von der Welt wahrnehmen kann, ist das nervöse Ticken einiger Millionen bündelweise durch Öffnungen in meine Höhle ragender Fasern.“<sup>7</sup> Dieser Schädelhöhlenklastrophobiker verwechselt sich schlicht mit seinem Gehirn.

In Marrs computationaler Wahrnehmungstheorie bin es nicht „ich“, dem die Eigenschaften des Netzhautbildes beschrieben werden, sondern etwas *in mir*, eine postulierte Instanz im Prozeß der Bildauswertung. Man kann allgemein sagen, daß der Homunkulus in Theorien, die auf dem Symbolverarbeitungsansatz beruhen, seine Natur verändert: aus dem inneren Betrachter wird der innere Leser und Interpret. Um den metaphorischen Charakter dieser Rede zu erkennen, muß man sich nur die Frage vorlegen, in welcher Sprache Marrs „primal sketch“, der aus einer Beschreibung geometrischer Eigenschaften bestehen soll, formuliert sein mag; dazu unten.

Merkwürdig ist auch, daß der Output des Wahrnehmungsprozesses eine Beschreibung der dreidimensionalen Außenwelt sein soll. Eine Beschreibung der Außenwelt wird freilich gegeben, wenn wir ein Wahrnehmungsurteil fällen. Wahrnehmungsurteile hat Marr aber nicht im Sinn, wenn er den Output eine Beschreibung oder Repräsentation „of the three-dimensional shape and spatial arrangement of an object“ nennt (a.a.O., 37). Es soll sich dabei um den Output des Wahrnehmungsprozesses selbst handeln, den Marr als algorithmischen Prozeß der Extraktion räumlicher Information aus dem Netzhautbild auffaßt. Der Output des Wahrnehmungsprozesses ist aber keine Beschreibung und keine Repräsentation, sondern – eine Wahrnehmung.<sup>8</sup>

<sup>6</sup> Ebenso Wittgenstein: „Ist das eben nicht eine irreführende Metapher: ‘Mein Auge gibt mir Kunde davon, daß dort ein Sessel stehe?’“ (PU § 356).

<sup>7</sup> Den Hinweis auf dieses Zitat des Neurobiologen Christoph von der Malsburg verdanke ich Andreas Kemmerling (vgl. Kemmerling 2000, 224). Ähnliche Beschreibungen finden sich bei „radikalen Konstruktivisten“.

<sup>8</sup> So sieht es auch Hacker: „The output of the neuro-visual process, if it can be said to have an ‘output’, is *that the creature sees* whatever it sees. But to see something is not to construct or produce a description of anything“ (Hacker 1991, 143).

#### 4. Homunkulismus und Repräsentationalismus

Die Frage, in welcher Sprache oder welchem Repräsentationsformat die am Erbringen einer kognitiven Leistung angeblich beteiligten Instanzen oder Subsysteme miteinander kommunizieren, ist noch offen. Bei Fodor ist es „Mentalese“, eine eigens postulierte „Language of Thought“. Die Annahme, daß es eine solche Sprache geben muß, in der Teile des Geistes miteinander kommunizieren, beruht auf dem Repräsentationalismus. In Fodors „Repräsentationaler Theorie des Geistes“ wird die Arbeitsweise des Geistes als Berechnungsprozeß aufgefaßt, der mit mentalen Repräsentationen operiert, genauer: mit den syntaktischen Eigenschaften mentaler Symbole. Der Vorteil der Beschränkung auf die syntaktischen Eigenschaften liegt auf der Hand: Einer „syntaktischen Maschine“, die uninterpretierte Symbole manipuliert, braucht man nicht die Fähigkeit zuzuschreiben, Symbole zu *verstehen*. Gesteht man ihr dagegen zu, wie van Gulick (1989, 223) es tut, „to understand the content of the internal symbols or representations on which it operates“, dann ist man, so wendet Beckermann (1995, 667) ein, „einer *homunculus*-Annahme zumindest gefährlich nahe“, denn in diesem Falle müsse man ja „eine interpretierende Instanz“ annehmen. Beckermann selbst möchte in seiner Skizze zur visuellen Informationsverarbeitung „völlig ohne *homunculus*-Annahmen aus[kommen]“ (ebd.). Eine Seite weiter, nachdem er vier Stufen der Verarbeitung des visuellen Reizes unterschieden hat, heißt es: „Am Ende haben wir sogar den Eindruck, nicht mehr ein Bild, sondern direkt Objekte und Szenen zu sehen.“ (668) Soll diese merkwürdig distanzierte Beschreibung besagen, daß wir in Wirklichkeit Bilder sehen und deren kausale Quellen erst rekonstruieren müssen? Dies wäre eine Einschätzung im Geiste *phänomenalistischer* Erkenntnistheorien, die als die philosophiehistorischen Vorläufer des Repräsentationalismus angesehen werden können. Phänomenalisten sind der Auffassung, daß die Objekte unserer Erkenntnis nicht äußere Gegenstände sind, sondern mentale Entitäten wie Vorstellungen, Empfindungen oder Sinnesdaten. So erklärten sich Locke und Hume die Wahrnehmung als eine „inner perception of ideas“. Beckermann kommt es an der zitierten Stelle auf die schwächere These an, daß wir unsere Aufmerksamkeit *auch* auf unsere Sehempfindung richten können, nicht nur auf die äußeren Gegenstände. Er sagt weiter: „Wir sehen zwar normalerweise ‘durch die Bilder hindurch’ die Objekte und Szenen, auf die diese Bilder zurückgehen, doch wir können uns auch auf die Bilder und ihre Elemente konzentrieren.“ (668) Welche Bilder meint Beckermann? „Bilder in dem hier einschlägigen Sinn“, stellt er klar, seien „nichts weiter als zweidimensionale Projektionen dreidimensionaler Szenen“ (669), eben „Netzhautbilder, genauer: Verteilungen der Feuerungsraten der Photosensoren in der Netzhaut“ (664). Indes paßt vieles von dem, was er über die frag-

lichen Bilder sagt, nicht zu dieser Erläuterung. Dasjenige, worauf wir unsere Aufmerksamkeit richten können, während wir beispielsweise beim Sehen mit einem Finger seitlich auf einen Augapfel drücken (vgl. 668), ist der phänomenale Wahrnehmungsinhalt und seine Veränderung. Dieser Wahrnehmungsinhalt kann aber nicht mit dem „Netzhautbild“ im Sinne des zweidimensionalen Bestrahlungsmusters identisch sein, denn auf dieses können wir unsere Aufmerksamkeit nicht richten. Dafür müßte das Netzhautbild ja ein phänomenal zugänglicher Teil des Wahrnehmungsprozesses sein; kurz: man müßte es sehen können.<sup>9</sup> Alternierend mit „Netzhautbild“ verwendet Beckermann auch den Ausdruck „Wahrnehmungsbild“, was die Sache nicht klarer macht. Er erklärt es dann weiter als „die natürliche Aufgabe unseres Wahrnehmungssystems, Netzhautbilder zu interpretieren, d. h. aus Netzhautbildern Repräsentationen der zugrunde liegenden Szenen zu rekonstruieren“ (669). Mit dieser Aufgabenbestimmung, die derjenigen von Marr entspricht, fällt er just in die Rede vom „Interpretieren“ zurück, die er bei van Gulick beanstandet hatte.

Es stellt sich die Frage, ob repräsentationalistische Auffassungen des Geistes oder gar des Gehirns *notwendig* den Homunkulus als Betrachter, Leser oder Interpreten nach sich ziehen. Mündet die Annahme „mentaler Repräsentationen“ zwangsläufig in den Homunkulismus? Die Antwort hängt unter anderem davon ab, was mentale Repräsentationen überhaupt sind.

Das lateinische Wort *repraesentatio* wird am besten mit „Vorstellung“ übersetzt. Legen wir die anspruchslose Annahme zugrunde, daß Menschen die Fähigkeit haben, sich etwas vorzustellen. Sich etwas vorzustellen mag in einem vorthoretischen Sinne gleichbedeutend damit sein, eine Vorstellung von etwas zu haben. Dabei ist die Produkt/Prozeß-Ambiguität von „Vorstellung“ zu beachten: man kann mit diesem Ausdruck sowohl das Vorgestellte als auch den Prozeß oder die Tätigkeit des Vorstellens bezeichnen. Im kognitionswissenschaftlichen Repräsentationalismus werden die Produkte der Vorstellungstätigkeit dann weiter zu „mentalen Repräsentationen“ hypostasiert, die syntaktische Eigenschaften und semantische Gehalte haben. Zugleich sollen sie diejenigen Elemente sein, die in mentalen Prozessen interagieren. Daß mentale Repräsentationen syntaktische und semantische Eigenschaften haben, setzt klarerweise voraus, daß sie Elemente einer Sprache sind. Mentale Repräsentationen wären dann Symbole oder Symbolketten. Daß sie interagieren, setzt voraus, daß sie kausale Beziehungen eingehen können. Ontologisch werden sie von den Repräsentationalisten

<sup>9</sup> Ähnlich wendet Hacker gegen Marr ein: „Sensations are felt (one feels pain, giddiness, nausea) in particular locations (in one’s tooth, head, stomach) but one does not *feel* one’s retinal images, one is not aware of them, and they are not objects of attention“ (Hacker 1991, 124).

meist als Zustände aufgefaßt, manchmal auch als Ereignisse oder Prozesse. Nach Fodor sind mentale Repräsentationen in einen Berechnungsprozeß involviert, der zugleich ein *kausaler Mechanismus* ist. Eben dies zeichne ein System aus, dem Repräsentationen nicht bloß zugeschrieben werden, sondern das tatsächlich mit ihnen arbeitet: „What distinguishes what organisms do from what planets do is that a *representation of the rules they follow constitutes one of the causal determinants of their behavior*“ (Fodor 1975, 74). Kurz: „Mental processes are causal sequences of tokenings of mental representations“ (Fodor 1987, 76). Ist man der Auffassung, daß nur konkrete, raumzeitliche Entitäten kausal interagieren können, so kommen als diese „tokenings“ nur die physiologischen Korrelate oder Träger der mentalen Repräsentationen in Frage.

Es liegt nach dem Gesagten auf der Hand, daß mentale Repräsentationen allenfalls theoretische Entitäten sein können, die mit Vorstellungen im landläufigen Sinne wenig gemein haben. Der kognitionswissenschaftliche Repräsentationalismus handelt nicht davon, daß Menschen sich etwas vorstellen können (den Wert einer Psychologie der Vorstellungen bestreitet allein der Behaviorismus), sondern er behauptet, daß physische Strukturen, in denen Geist realisiert ist, seien es Gehirne, Computer oder andere Artefakte<sup>10</sup>, Berechnungen mit Repräsentationen ausführen: mit Symbolen, die für etwas außerhalb ihrer selbst stehen. Ob sich theoretische Entitäten konstruieren lassen, die die genannten Anforderungen für mentale Repräsentationen zugleich erfüllen, ist eine offene Frage. Nicht minder umstritten ist, wofür die Annahme mentaler Repräsentationen überhaupt benötigt wird.

Kritik am Repräsentationalismus liegt in moderater und in radikaler Form vor. Als moderat bezeichne ich den Einwand des frühen Searle, daß Syntax für Semantik nicht hinreichend sei. Radikal ist der Einwand des späten Searle, daß physische Prozesse nicht einmal Berechnungsprozesse seien, weil die Syntax der Physis nicht intrinsisch sei. An Searle halte ich mich, um die Verbindung mit dem Homunkulusproblem zu wahren. Nach Searle ist nämlich der Homunkulus-Fehlschluß dem, was er den „Kognitivismus“ nennt – der Auffassung, daß das Gehirn ein Digitalcomputer und geistige Prozesse computational sind –, endemisch.

Der Slogan „Syntax reicht nicht für Semantik“ wird durch Searles Gedankenexperiment des Chinesischen Zimmers illustriert (vgl. Searle 1980). Dieses soll zeigen, daß meine Fähigkeit, regelgerecht Symbole zu manipulieren, nicht schon garantiert, daß diese Symbole *für mich* Bedeutung haben.

<sup>10</sup> Das ständige Schwanken zwischen der Rede vom Geist, vom Gehirn und vom Computer ist in einer Darstellung des Repräsentationalismus schwer zu vermeiden. Es liegt in der Logik der funktionalistischen Auffassung des Geistes, die sich durch die Annahme der *multiplen Realisierbarkeit* mentaler Zustände auszeichnet.

Diese Überlegung überträgt Searle auf die Arbeitsweise von Computern. Die Lehre lautet, daß Computer mit Symbolen operieren, die für sie selbst keine Repräsentationen von irgendetwas sind, sondern bedeutungslose Zeichenketten. Computer verfügen deshalb nicht über echte Repräsentationen und echte Intentionalität, sondern über bloß zugeschriebene. Die einem Computer zugeschriebenen Repräsentationen und intentionalen Zustände kommen im wörtlichen Sinne nur seinen Konstrukteuren und Benutzern zu. Zehn Jahre später verschärft Searle seine Argumentation. Sein Einwand lautet nun, daß ein physischer Prozeß intrinsischerweise nicht einmal ein Berechnungsvorgang oder eine Regelbefolgung sei, da nämlich physische Vorgänge *per se* nicht syntaktisch miteinander verknüpft seien. Nicht bloß sei die Semantik nicht in der Syntax enthalten, „I am now making the separate and different point that syntax is not intrinsic to physics“ (Searle 1990, 27). Weder die elektromechanischen Vorgänge innerhalb eines Computers noch die physiologischen Vorgänge im Gehirn hätten intrinsischerweise syntaktische Eigenschaften. Die „brute, blind nonconscious neurophysiological operations of the brain“ (ebd., 33) seien nicht uninterpretierte Berechnungsprozesse, sondern überhaupt keine. Genauer: Die Frage, ob sie welche sind, ist gegenstandslos, denn wie die Semantik ist auch die Syntax stets im Auge des Betrachters. „The ascription of syntactical properties is always relative to an agent or observer who treats certain physical phenomena as syntactical“ (26). Wir können die Arbeitsweise von Gehirnen als Berechnungsprozesse *interpretieren* und auf Computern *simulieren*, aber wir können nicht *entdecken*, daß irgendetwas ein Berechnungsprozeß ist. Kein physisches System führt intrinsischerweise Berechnungen aus, nicht einmal ein Computer selbst. Ein Rechenprozeß zu sein ist keine physische Eigenschaft, denn Berechnung ist eine bestimmte Art von Symbolmanipulation<sup>11</sup>, und „‘symbol’ and ‘same symbol’ are not defined in terms of physical features“ (35). Beispielsweise können sich zwei Vorkommnisse eines Buchstabens in ihren physischen Merkmalen drastisch voneinander unterscheiden (zumal in Handschriften). Daß sie gleichwohl als Vorkommnisse desselben Buchstabens zählen, läßt sich nur mit Bezug auf Konventionen und Intentionen der Benutzer erklären. Bei digitalen Symbolsystemen ist die Irrelevanz der physischen Eigenschaften noch eindrücklicher: Buchstäblich jeder physischen Struktur, an der sich irgendetwas unterscheiden läßt, lassen sich die

<sup>11</sup> Der Zusammenhang zwischen Computationalismus und Repräsentationalismus läßt sich mit Beckermann so darstellen: „Rechenprozesse 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 auch auf die Annahme festgelegt, daß es im Gehirn ein System von Repräsentationen gibt“ (Beckermann 1997, 90).

beiden Symbole „Null“ und „Eins“ zuordnen. Die derart zu Zeichenträgern beförderten physischen Strukturen können anliegende elektrische Spannungen oder beliebige andere unterscheidbare Zustände oder Dinge sein. Daß der Zeichencharakter eines Dinges nicht durch seine physischen Eigenschaften bestimmt wird, ist der Grund in der Sache für die multiple Realisierbarkeit von Programmen; Searle spricht sogar von einer „universellen“ Realisierbarkeit.

Das Gesagte wird nicht nur für symbolische Repräsentationen behauptet, sondern in gleicher Weise für bildhafte. Die von einer Ameise in den Sand gezogene Spur, um ein berühmtes Beispiel von Putnam aufzugreifen, sei von sich aus auch bei täuschender Ähnlichkeit kein Portrait Winston Churchills.

Der Zusammenhang mit dem Homunkulus-Problem stellt sich für Searle wie folgt dar: Physische Zustände in Gehirnen oder Computern haben intrinsischerweise keine Syntax und repräsentieren *a fortiori* nichts. Wenn wir ihnen gleichwohl syntaktische Eigenschaften und Repräsentationsgehalte zuschreiben, müssen diese Züge von einer Quelle abgeleitet sein, die echte mentale Repräsentationen hat. Diese Quelle ist der menschliche Geist. Geistbegabte Wesen können physischen Zuständen und Vorgängen Bedeutung *verleihen* und sie fortan so behandeln, als ob sie von sich aus Bedeutungsträger wären (nach Searle kommen auch sprachliche Ausdrücke auf diese Weise zu ihren Bedeutungen: wir können die Wortbedeutungen als Objektivationen unserer Äußerungsintentionen ansehen).<sup>12</sup> In diesem Sinne fungierten die Benutzer eines Computers als „externe Homunkuli“: Wir interpretieren die Zustände des Computers, als ob sie etwas repräsentierten, was sie doch im Wortsinne nur täten, wenn das geistbegabte Wesen in der Maschine säße statt vor ihr. In ähnlicher Weise sprechen wir beispielsweise vom „zielsuchenden Torpedo“. Die Intentionen des Benutzers werden dem Artefakt selbst zugeschrieben; man faßt sie als ihm gleichsam eingebaut auf. Hier gibt es allerdings einen interessanten Unterschied zwischen Computer und Gehirn. Während der Computer von geistbegabten Wesen benutzt wird, kann man dies nach Searle vom Gehirn gerade nicht sagen. Darum verschärft sich hier das Homunkulus-Problem:

„For real computers of the kind you buy in the store, there is no homunculus problem, each user is the homunculus in question. But if we are to suppose

<sup>12</sup> In jüngerer Zeit wird die Ablehnung intrinsischer Repräsentationen durch *externalistische* Bedeutungstheorien begründet: Welche Bedeutung ein Begriff habe, hänge auch von den äußeren Gegenständen ab, die bei seiner Entstehung eine kausale Rolle gespielt haben (Putnam, Davidson, Burge).

that the brain is a digital computer, we are still faced with the question ‘And who is the user?’“ (ebd., 29)

Die einschlägige Rede vom „Ich“ als Benutzer seines eigenen Gehirns ist schief, denn die erforderliche minimale Unabhängigkeit von Benutzer und Werkzeug ist hier nicht gewährleistet. Allem, was ich „mithilfe“ meines Gehirns tue, gehen ja bereits Gehirnprozesse einher.

Der gemeinsame Grund für Searles moderaten und seinen verschärften Einwand gegen den Symbolverarbeitungsansatz in den Kognitionswissenschaften ist also das Junktim, das er zwischen Sprache und Geist herstellt. Nur geistbegabte Wesen kämen als Sprachbenutzer in Frage.

Natürlich ist diese Auffassung alles andere als originell. Die Kontroverse um die Zulässigkeit des Repräsentations- und des Symbolbegriffs hat ihren Vorläufer in der Debatte um den Begriff der *Information*, die ein halbes Jahrhundert zuvor stattfand. Shannon und Weaver hatten ihren quantitativen, „objektiven“ Informationsbegriff im Kontext der Nachrichtentechnik eingeführt und ihn ausdrücklich nichtsemantisch verstanden: „In particular, *information* must not be confused with meaning“ (Shannon/Weaver 1949, 8). Den Informationsgehalt einer Nachricht bestimmten sie als das Maß der Auftrittswahrscheinlichkeit einer bestimmten Zeichenkette. Grob gesprochen: Je geordneter eine Zeichenkette, desto weniger wahrscheinlich ist ihr Auftreten, desto mehr Information enthält sie. Nichtsemantisch ist dieser Informationsbegriff, insofern er es nicht erlaubt, bedeutungsvolle von bedeutungslosen Zeichenketten zu unterscheiden. Damit scheint er in dem Sinne „objektiv“ zu sein, daß der Informationsgehalt einer Aktivität im Nachrichtenkanal nicht von Intentionen und Konventionen abhängig ist. Dieser Anschein trügt aber, denn damit die Auftretenswahrscheinlichkeit einer Zeichenkette bestimmt werden kann, muß zuvor festgelegt worden sein, was als Kanal zählt und welche Aktivität im Kanal als Nachricht zählt. Es werden ja nicht alle physischen Eigenschaften dieser Aktivität als informationstragend behandelt, sondern nur ausgewählte. Sender und Empfänger müssen einen gemeinsamen Zeichenvorrat verabredet haben, ein Lexikon (vgl. dazu von Weizsäcker 1971, 348 ff.) Ein Zeichen zu sein ist, wie Searle richtig bemerkt, keine physische Eigenschaft eines Gegenstands.

Diesem Befund widersprechen die neueren *naturalistischen Bedeutungstheorien*. Die Projekte einer naturalistischen Grundlegung der Semantik, wie sie von Dretske, Fodor, Millikan und Papineau verfolgt werden, nehmen ihren Ausgang von den sogenannten „natürlichen Zeichen“ (Grice), um dann solche Zeichen, wie etwa die Jahresringe eines Baumes, als Basisfälle des Bedeutens auszuzeichnen. Die Relation der natürlichen Anzeige wird dann meist mithilfe evolutionstheoretischer Überlegungen weiter angereichert („Biosemantik“, „Teleofunktionalismus“), auf daß sich die vermeint-

lich konventionelle Bedeutung sprachlicher Symbole als Spezialfall der natürlichen Bedeutung ergeben möge. Nun sind aber die „natürlichen Zeichen“ überhaupt keine Zeichen, vielmehr sind sie *Anzeichen* oder Indikatoren. Ein wichtiger Unterschied zwischen konventionellen Zeichen und kausalen Anzeichen besteht darin, daß Zeichen fehlrepräsentieren können, Anzeichen aber nicht. Natürliche Anzeichen in dem Sinne, in dem Dretske diesen Begriff präzisiert, können *per definitionem* nicht fehlgehen. Was fehlgehen kann, ist unser *Wissen* um die Kausalbeziehungen, die den Anzeichen zugrunde liegen. Nun ist es der menschlichen Sprache als einem intentionalen Phänomen aber wesentlich, daß Sprecher sich auf Gegenstände beziehen können, über deren wesentliche Eigenschaften sie sich täuschen oder die sie mit anderen verwechseln. Die Möglichkeit der Fehlrepräsentation gehört zu den *Explananda* einer Bedeutungstheorie, kann durch die Annahme einer Kausal- oder strengen Kovarianzbeziehung zwischen Zeichen und Bezeichnetem aber nicht erklärt werden. Daß die Identifikation eines Redegegenstandes fehlschlagen kann, spielt eine wichtige Rolle bei der Ausdifferenzierung von Intension und Extension: Der sprachliche Sinn des Wortes „Pferd“ ändert sich ja nicht, wenn ein Sprecher sich im besonderen Falle über die Anwesenheit eines Pferdes täuscht. Es ergibt sich ein Folgeproblem, das als das „Disjunktionsproblem“ diskutiert wird: Wie kann eine kausale Bedeutungstheorie den Fall, daß eine Repräsentation „F“ bedeuten soll, aber manchmal auch von Gs verursacht wird, von dem Fall unterscheiden, daß eine Repräsentation „F-oder-G“ bedeutet? Das Disjunktionsproblem ist ein Spezialfall der *Unterbestimmtheit* der Bedeutung durch die Belege. Ein prominenter Ausweg ist die Erweiterung der kausalen zu einer „teleofunktionalistischen“ Bedeutungstheorie, welche Repräsentationsgehalte auf *Normalbedingungen* relativiert und auf diese Weise zu spezifizieren sucht. Die Normal- oder Optimalbedingungen werden ihrerseits evolutionstheoretisch hergeleitet. Ob diese normativen Prädikate sich tatsächlich auf naturalistisch akzeptable Weise erläutern lassen, ist umstritten. Soweit ich sehe, ist es den naturalistischen Bedeutungstheorien bisher nicht gelungen, die behauptete enge Verwandtschaft von konventioneller Bedeutung und natürlichen Anzeichen plausibel zu machen.

Halten wir als Zwischenergebnis fest, daß das Kritikziel der Homunkulismus-Kritiker nicht die Kognitionswissenschaften als solche sind, sondern diejenigen kognitionswissenschaftlichen Theorien, die dem Repräsentationalismus frönen.<sup>13</sup> Wo dem Gehirn ein Operieren mit bildhaften Repräsentation

<sup>13</sup> Fodors „Repräsentationale Theorie des Geistes“ ist freilich nur *eine* Variante des Repräsentationalismus, wenn auch eine besonders einflußreiche und weit ausgearbeitete. Eine eingehendere Auseinandersetzung mit dem Repräsentationalismus hätte u.a. Fodors Unterscheidung zwischen „weitem“ und „engem“ Inhalt, die Kon-

tionen zugeschrieben wird, müsse ein Homunkulus als innerer Betrachter postuliert werden; werden symbolische Repräsentationen in Mentalesisch angenommen, mutiere der Homunkulus zum Leser und Interpreten.<sup>14</sup>

Auf den Homunkulus-Einwand reagieren die Repräsentationalisten auf verschiedene Weise. Einige bestreiten, daß das Postulieren von Homunkuli nötig ist, andere packen den Stier bei den Hörnern und rehabilitieren den Homunkulus (dazu unten), wieder andere verteidigen die repräsentationalistische Auffassung durch den Verweis auf ihre Unumgänglichkeit. Der letztere Zug ist keine direkte Widerlegung des Homunkulus-Einwandes, wohl aber eine indirekte: Wenn es zum Repräsentationalismus keine theoretische Alternative gibt, müssen sinnkritische Gegeneinwände unzutreffend oder irrelevant sein.

Das Argument der Unumgänglichkeit besagt, daß ohne die Annahme interner Repräsentationen nicht erklärt werden kann, wie ein Computer oder ein Gehirn bestimmte Leistungen erbringen. Oft ist die Erklärung eines Systemverhaltens, wie Beckermann sagt, „nur möglich, wenn man bestimmte physische Strukturen im Inneren des Systems als Repräsentationen auffaßt“ (Beckermann 1997, 89).

Zu diesem Argument ist zunächst zu bemerken, daß die Homunkulus-Kritiker die Erklärungskraft repräsentationaler Erklärungen typischerweise nicht bestreiten. Eine Erklärung kann zu schwach sein, also etwas unerklärt lassen, sie kann aber auch zu stark sein, also zu viel erklären. Moniert wird an repräsentationalen und intentionalen Erklärungen des Verhaltens von Nichtpersonen das letztere. Man betrachte zum Beispiel die Erklärung des Verhaltens eines Tisches durch das Anführen seiner intentionalen Einstellungen: „Mir gefällt es hier gut, ich werde mich in der nächsten Stunde nicht von der Stelle bewegen“. Diese Erklärung mag prognostisch äußerst erfolgreich sein. Doch sie ist unnötig stark. Ein weniger anspruchsvolles Explanans hätte ausgereicht, nämlich ein Verweis auf das Trägheitsprinzip. (Zudem verletzt die intentionale Erklärung eine von Hempels Adäquatheitsbedingungen für wissenschaftliche Erklärungen: Das Explanans darf nur wahre Sätze enthalten.) Einschlägig ist hier Dennetts Lehre von den drei

sistenz der verschiedenen Teile seiner Theorie und dann alternative Versionen des Repräsentationalismus zu prüfen.

<sup>14</sup> Der Begriff des Lesers, wenn auch nicht des Interpreten, findet sich bereits im Konzept der universellen Turingmaschine, eines endlichen Automaten, der Symbole auf dem Eingabeband lesen und in Abhängigkeit davon in einen anderen Zustand übergehen kann. Zur Frage, wie ein solcher abstrakter Automat physisch realisiert sein soll, finden sich in der Literatur frivole Kommentare wie diese: „We are not concerned with the mechanics or the electronics of the matter. Perhaps the simplest way to picture the thing is quite crudely: Inside the box there is a man, who does all the reading and writing and erasing and moving“ (Boolos/Jeffrey 1989, 21).

Einstellungen (*physical stance, design stance, intentional stance*). Dennett behauptet zwar, daß die Einnahme des intentionalen Standpunkts durch ihren Prognoseerfolg gerechtfertigt wird, doch er knüpft diese Rechtfertigung an die Bedingung, daß andere, nichtintentionale Erklärungsstrategien versagt haben. Aus diesem Grunde schreiben wir einem Planeten keine Repräsentation seiner Umlaufbahn zu: Eine Erklärung vom „physical stance“ erklärt seine Bahnkurve ebensogut. Denselben Kommentar provoziert Searles witzige Frage, die als *reductio ad absurdum* einer computationalen Erklärung gemeint ist („Wie berechnet ein Nagel aus dem Schlag des Hammers die Distanz, die er im Holz zurücklegen muß?“).

Beckermann hat diesen Überlegungen schon Rechnung getragen, denn er spricht ja von Fällen, in denen sich das Systemverhalten *allein* durch die Annahme von Repräsentationen erklären läßt. Nun, dieser Nachweis, daß eine physikalische Erklärung nicht möglich ist, wäre im Einzelfall zu erbringen. Doch wie sollte ein solcher Nachweis aussehen? Eine untadelige Erklärung des Systemverhaltens beispielsweise eines Schachcomputers (Beckermanns Beispiel) bestünde im Anführen einer gesetzmäßigen Korrelation zwischen Input und Output auf der Ebene physischer Prozesse. *De facto* verfügen wir über solche Erklärungen des Verhaltens von Computern, auf denen komplexe Programme laufen, nicht, und über mikrophysische Erklärungen des Verhaltens von Gehirnen erst recht nicht. Der Verweis auf die faktische Undurchschaubarkeit eines Systems reicht aber nur dazu aus, den Übergang zu repräsentationalen Erklärungen *pragmatisch* zu rechtfertigen. Auf diese pragmatische Notwendigkeit wird man sich schnell einigen können. Die mechanistische Erklärung des Verhaltens eines Schachcomputers wäre immens kompliziert und würde den Rückgriff auf Wissen erfordern, das in zu vielen Köpfen verteilt ist. Alle Spielstrategien, die seine Entwickler programmiert haben, müßten Schritt für Schritt aus der Programmiersprache in elementarere Beschreibungen zurückübersetzt werden, bis die Ebene der elektromechanischen Vorgänge im Hauptprozessor erreicht ist.<sup>15</sup> Eine solche „Erklärung“ würde viele dicke Bücher füllen und wäre überdies für jemanden, der sich mit Beckermann fragt, warum der Computer „so gute Züge wählt“, uninformativ.

Der letzte Punkt verlangt eine nähere Erörterung. Beckermann bezieht seine Unumgänglichkeitsthese nicht auf Erklärungen jedweder Art, sondern auf 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“ (Beckermann 1997, 89).

<sup>15</sup> Schon den Transitionsgraphen für die Zustandsübergänge einer universellen Turingmaschine, auf der ein modernes Schachprogramm läuft, dürfte niemals jemand aufgezeichnet haben.

Für alternativlos erklärt er den Repräsentationalismus somit nur bei *funktionalen* Erklärungen *erfolgreichen Systemverhaltens*. Diese Einschränkung ist geeignet, die Unumgänglichkeitsthese zu trivialisieren. Beckermann sucht nach einer Erklärung für die „Tatsache, daß unser Schachcomputer – in der Regel – plausible oder sogar erfolgreiche Züge macht“ (ebd., 88). Dieses Explanandum ist aber schon in repräsentationalen Begriffen formuliert! Wenn wir das Verhalten des Rechners als das Ausführen „plausibler“ oder „erfolgreicher Züge“ beschreiben, ist es keine offene Frage mehr, ob wir „bestimmte physische Strukturen im Inneren des Systems [...] als Repräsentationen von möglichen Stellungen“ (ebd., 88 f.) auffassen wollen. Allgemein fällt bei funktionalen und intentionalen Erklärungen die Entscheidung über den verwendeten Begriffsrahmen, anders als bei Mikro-Makro-Erklärungen, schon bei der Formulierung des Explanandums.

Beckermanns Rechtfertigung des Repräsentationalismus ist daher in der vorliegenden Form trivial. Zu rechtfertigen gewesen wäre die im Vorfeld getroffene Entscheidung, den Output des Schachcomputers überhaupt als das „Ausführen erfolgreicher Züge“ zu beschreiben. Die „Leistungen“, so wird der Homunkulus-Kritiker einwenden, werden von vornherein der falschen Instanz zugeschrieben, nämlich dem Programm statt den Programmierern. Die letzte Behauptung greift Beckermann auch direkt an. Sie ginge, wie er sagt, „an meiner Argumentation völlig vorbei“, denn wir müßten Schachcomputer, „wenn wir verstehen wollen, wie das erfolgreiche Verhalten dieser Systeme zustandekommt“, auch dann repräsentational erklären, „wenn sie auf Bäumen wachsen würden“ (ebd., 89). Doch die Einschränkung auf das Verstehenwollen des erfolgreichen Verhaltens trivialisiert die Behauptung wiederum. Der Gegner, den es zu widerlegen gälte, wird das Flackern auf dem Bildschirm des Schachcomputers von vornherein nicht als die Ausführung erfolgreicher Spielzüge beschreiben. Er wird sogar bestreiten, daß der Schachcomputer überhaupt Schach spielt. Zur Begründung könnte er etwa anführen, daß es zum Sinn des Schachspiels gehört, daß man gewinnen möchte. Diesen Wunsch könne man dem Schachcomputer aber nicht zuschreiben, da Artefakte überhaupt keine Wünsche hätten. Auch dies wird von manchen Kognitionswissenschaftlern bestritten, doch wird die Behauptung, daß Computer sogar *mentale* Zustände hätten, nicht einfacher zu rechtfertigen sein als die, daß sie repräsentationale Zustände hätten.

Mit seinem Gegeneinwand, daß die Genese eines Schachcomputers für die Zuschreibung von Repräsentationen keine Rolle spiele, übersieht Beckermann noch ein wichtiges Detail der oben referierten Homunkulus-Kritik. Bei Searle ist es nicht der Konstrukteur eines Artefakts, sondern sein *Benutzer*, der den Systemzuständen die Repräsentationen zuschreibt, die dem jeweiligen Verwendungszweck entsprechen. Dies ist besonders plausibel

bei Fällen, in denen über den Konstrukteur nichts bekannt ist, oder bei Vielzweckgeräten, deren funktionale Charakterisierung mit der jeweiligen Verwendung variiert. Zu widerlegen wäre also die Behauptung, daß Artefakten Repräsentationen in *dem* Sinne nicht intrinsisch sind, daß sie ihnen aufgrund von Überlegungen, die durch die physische Struktur des Artefakts unterbestimmt bleiben, *entweder* vom Konstrukteur *oder* vom Benutzer zugeschrieben werden. Normalerweise werden Konstruktions- und Verwendungszweck zusammenfallen und also die gleichen Zuschreibungen rechtfertigen; sie können aber auch deutlich auseinanderklaffen. (Wer sich auf keinen Fall an Rüstungsexporten beteiligen will, darf auch keine „dual use“-Güter liefern.) In jedem Falle scheint die Frage, einen Apparat mit welchen Repräsentationen man denn „wirklich“ vor sich hat, keine Tatsachenfrage zu sein.

##### 5. Der homunkulare Funktionalismus

In den Kognitionswissenschaften liegen mittlerweile eine ganze Reihe von offensiven Auseinandersetzungen mit dem Homunkulus-Einwand vor. Als „offensiv“ bezeichne ich Reaktionen, in denen die Postulierung innerer Instanzen mit menschenähnlichen Fähigkeiten nicht geleugnet, sondern ausdrücklich zugestanden wird. Diese Stellungnahmen reichen von Versuchen, den Homunkulus als heuristisch fruchtbare, harmlose Metapher zu erweisen, bis zu einer nachdrücklichen Rehabilitierung, die dem Homunkulus eine echte Erklärungsrolle zugesteht. Letztere Strategie verfolgt der *homunkulare Funktionalismus*.

Außer Betracht lasse ich Stellungnahmen, in denen Homunkuli wörtlich als kleine menschenähnliche Wesen aufgefaßt werden. Man findet in der Literatur gelegentlich die treuherzige Versicherung, daß es im Gehirn schließlich keine Homunkuli gäbe, insbesondere bei neurowissenschaftlich orientierten Kognitionswissenschaftlern, die den Repräsentationalismus zugunsten des konnektionistischen Paradigmas der neuronalen Netzwerke aufgegeben haben. So sagt Patricia Churchland: „The cardinal background principle for the [neuroscience] theorist is that there are no homunculi. [...] There are just neurons and their connections.“ (Churchland 1986, 406). Im Gegensatz dazu fassen die Autoren, die ich im folgenden diskutiere, Homunkuli als postulierte „theoretical entities“ auf (Lycan 1987, 40). Einige Autoren sehen die Homunkulus-Hypothese als regelrechtes neurowissenschaftliches Forschungsprogramm an, nämlich als Suche nach einem Gehirnmareal, das zentrale Kontroll- und Integrationsleistungen erbringt. So zitiert Tibbets (1995, 404) den Befund, daß „there is no single cortical area to which all other cortical areas report exclusively, either in the visual or in

any other system“ (Zeki 1993, 296), und legt diesen Umstand so aus, daß es im Gehirn keinen Homunkulus gebe. Auch hier ist vom Homunkulus als einer handfesten, empirisch entdeckbaren Struktur die Rede, nicht von einer zu Erklärungszwecken postulierten theoretischen Entität.

Unter den Autoren, die den Stier bei den Hörnern packen und den Homunkulismus offensiv vertreten, sind Dennett und Lycan die einflußreichsten. Ältere Vorarbeiten zum homunkularen Funktionalismus stellen die Aufsätze von Attneave und Fodor dar.

Nach Attneave sind die Standardeinwände gegen den Homunkulus „directed at a little man of straw“ (Attneave 1961, 777). Die Verfemung des Homunkulus sei mittlerweile zu einem ernsthaften Forschungshemmnis geworden:

„The charge that an opponent is ‘assuming a little man inside the head’ has been accepted as an argument stopper for too long. It is time that we reconsider the scientific respectability of the homunculus.“ (ebd., 781)

Insbesondere greift Attneave das Argument an, die Postulierung eines Homunkulus führe notwendig in einen infiniten Regreß. Er hält dem entgegen:

„[I]t should be noted that we fall into a regress *only* if we try to make the homunculus do everything. The moment we specify certain processes that occur *outside* the homunculus, we are merely classifying or partitioning psychological functions; the classification may be crude, but it is not in itself regressive.“ (ebd., 778)

Damit gibt Attneave eine Argumentationslinie vor, die auch Fodor, Lycan und Dennett einschlagen werden: Einen Homunkulus zu postulieren ist dann nicht von Übel, wenn man ihn nicht die gesamte Arbeit tun läßt.

Fodor wendet sich ebenfalls dem Regreßeinwand zu. „Assigning psychological functions to little men“ sei zulässig und informativ, wenn die Analyse so weit vorangetrieben wird, daß sie auf keinen „unanalyzed psychological process“ Bezug nehmen muß. Zu diesem Zweck postuliert Fodor „a class of *elementary* instructions which the nervous system is specifically wired to execute“ (Fodor 1968, 629). Diese elementaren Anweisungen hätten „no theoretically relevant internal structure“, und das Gehirn benötige auch „no further instructions for carrying out the instructions on the list and no little men to supervise their execution. [...] [I]t simply performs them“ (ebd.). Gleichwohl versteht Fodor diese Anweisungen als „Regeln“, denen das Gehirn „folgt“. Den naheliegenden Einwand, daß diese Begriffe fehl am Platze seien, schreibt Fodor dem „intellektualistischen“ Vorurteil zu, daß

einer Regel zu folgen äquivalent mit der Fähigkeit sei, zu *erklären*, wie man etwas tut. Da diese Äquivalenz nicht besteht, seien wir berechtigt, einem Organismus das Befolgen derjenigen Regeln zuzuschreiben, die für eine Computersimulation seines Verhaltens benötigt werden. Wenn ein Organismus X tun kann, ohne erklären zu können, wie er X tut, „then the organism *tacitly knows* the answer to the question ‘How do you do X?’“ (ebd., 638). Fodors Verteidigung des Idioms der Regeln und Repräsentationen beruht also auf der Einführung eines „tacit knowledge of rules“.

Ausgearbeitet werden die Ideen Attneaves und Fodors im *homunkularen Funktionalismus*, demzufolge bei der funktionalen Analyse kognitiver Prozesse das Postulieren von Homunkuli unvermeidlich und unter bestimmten Bedingungen auch gerechtfertigt und erklärungskräftig ist. Der homunkulare Funktionalismus schlägt eine *rekursive Dekomposition* homunkular beschriebener kognitiver Leistungen vor.

Dennett zufolge sind postulierte Homunkuli genau dann nicht von Übel, wenn es (a) möglichst viele sind und wenn sie (b) möglichst dumm sind. Seine Begründung dafür lautet wie folgt: Wenn es gelänge, einen Homunkulus, den man zur Erklärung einer bestimmten kognitiven Leistung postuliert hat, schrittweise durch Gruppen dümmerer Homunkuli zu ersetzen, durch eine absteigende Linie von Teams, die immer anspruchlosere Aufgaben erfüllen („Subroutinen“), dann könne man am Ende die allerdümmsten Homunkuli durch einen Mechanismus ersetzen. Die zu einiger Berühmtheit gelangten Ausführungen Dennetts dazu seien im Zusammenhang zitiert:

„The AI [= Artificial Intelligence] researcher *starts* with an intentionally characterized problem (e.g., how can I get a computer to *understand* questions of English?), breaks it down into sub-problems that are also intentionally characterized (e.g., how do I get the computer to *recognize* questions, *distinguish* subjects from predicates, *ignore* relevant parsings?) and then breaks these problems down still further until finally he reaches problem or task descriptions that are obviously mechanistic. [...] When the level is reached where the homunculi are no more than adders and subtractors, by the time they need only the intelligence to pick the larger of two numbers when directed to, they have been reduced to functionaries ‘who can be replaced by a machine’.” (Dennett 1978, 80 f.)

„In fact, homunculus talk is ubiquitous in AI, and almost always illuminating. AI homunculi talk to each other, wrest control from each other, volunteer, sub-contract, supervise, and even kill. There seems no better way of describing what is going on. Homunculi are *bogeymen* only if they duplicate *entire* the talents they are rung in to explain [...]. If one can get a team or

committee of *relatively* ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress.” (ebd., 123)

Die *top-down*-Strategie der rekursiven Dekomposition ist ausdrücklich gegen das Regreßargument gerichtet. Die Behauptung Dennetts lautet, daß der Regreß nicht infinit sei, sondern auf einer untersten Stufe, auf der die Aufgabenbeschreibungen „obviously mechanistic“ sind, zu einem Ende komme. Dennett ist der Auffassung, daß auf diese Weise ein *gradueller Übergang* von einer intentionalen zu einer nichtintentionalen, mechanistischen Charakterisierung einer kognitiven Leistung stattfindet. Und das Nachvollziehbarmachen eines solchen graduellen Überganges käme einer Naturalisierung der Intentionalität gleich – oder dem, was Dennett am Naturalisierungsprojekt für vernünftig hält.

Die Frage ist, was genau „klarerweise mechanistische“ Aufgabenbeschreibungen sein sollen. In *Consciousness Explained* wiederholt Dennett die Formulierung, der dümmste Homunkulus könne „be replaced by a machine“ (Dennett 1991, 91 u.ö.). Der Verweis auf die Ersetzbarkeit durch eine Maschine erklärt aber in diesem Zusammenhang nichts, denn dies gilt ja beispielsweise bei einem Schachcomputer auch für die gesamte Leistung des Schachspielens, seien die Subroutinen so beschrieben wie es beliebt. Maschinen oder Mechanismen ersetzen sowohl komplexe als auch einfache Leistungen. Es geht beim Homunkulusproblem überhaupt nicht darum, welche Leistungen Maschinen erbringen oder simulieren, sondern darum, mit welchen Begriffen ihr Funktionieren legitimerweise beschrieben werden kann. Dennetts wiederholte Versicherung, „we have replaced homunculi with (teams of) simple mechanisms“ (1991, 176), greift deshalb zu kurz. Es kann ja nicht das Wort „Mechanismus“ sein, das den Übergang von einer intentionalen zu einer naturalistisch respektablen Charakterisierung der fraglichen Prozesse bewerkstelligt.

Lycan argumentiert etwas anders als Dennett. Er behauptet nicht, daß die rekursive Dekomposition auf einer untersten Ebene in eine nichtintentionale Charakterisierung mündet. Vielmehr sieht er homunkulare Beschreibungen als Spezialfall *teleologischer* Beschreibungen an und macht sich daran, die wissenschaftliche Respektabilität der letzteren zu erweisen. Lycan nimmt eine „continuity of levels of nature“ an (Lycan 1987, 43), die ihren angemessenen Ausdruck in Hierarchien von mehr oder weniger teleologischen Charakterisierungen findet. Die entgegengesetzte Auffassung nennt er den „two-levelism“ von Funktion und materieller Struktur, welcher auch der Dichotomie von Software und Hardware zugrunde liegt (vgl. ebd., 46). Die Unterscheidung von materieller Struktur und Funktion könne aber nicht kategorisch getroffen werden, sondern nur relativ zur jeweils gewählten Beschreibungsebene. „The theoretical point is that the teleologicalness of

characterizations is a matter of degree“ (ebd., 43). Lycans homunkularer Funktionalismus überwindet also die Software/Hardware-Dichotomie des orthodoxen Maschinenfunktionalismus zugunsten einer Pluralität von Beschreibungsebenen, die metaphysisch durch die Annahme einer „continuity of levels of nature“ abgestützt wird.

Eine bloß instrumentelle oder metaphorische Interpretation der teleologischen Beschreibungen, auf die Dennett sich gelegentlich zurückzieht<sup>16</sup>, lehnt Lycan ausdrücklich ab:

„What is homuncular about H[omuncular] F[unctionalism] is that it interprets ‘function’ very robustly, in terms of the job descriptions of sub- and sub-subpersonal agencies that compose the bureaucracy that realizes a psychological subject. Nor does HF regard this as fiction or mere metaphor, but takes its explicitly teleological characterizations seriously as such.“ (Lycan 1991, 264)<sup>17</sup>

Lycans robust teleologische Naturauffassung enthält eine Kritik an Dennetts Rede von einer „klarerweise mechanistischen“ Basisbeschreibung. Lycan fragt: „Now at what point of this descent through the institutional hierarchy [...] does our characterization stop being teleological, period, and start being purely mechanical, period?“ (1987, 44) Wir seien in großer Verlegenheit, wenn wir die unterste Ebene benennen wollen: „‘Mechanical’? Hardly – *mechanisms* are functional items *par excellence*“ (ebd., 47). Nach ihren Funktionen werden Mechanismen schließlich klassifiziert; einschlägige Beispiele finden sich in der Biologie (z.B. physiologische Mechanismen der

<sup>16</sup> „All I have done, really, is to replace one family of metaphors and images with another, trading in the Theater, the Witness, the Central Meaner, the Figment, for Software, Virtual Machines, Multiple Drafts, a Pandemonium of Homunculi. It’s just a war of metaphors [...]. No one can think about consciousness about them“ (Dennett 1991, 455). Dieser fröhliche Verweis darauf, daß die Psychologie ohne Metaphern und Modelle nicht auskommt, reicht zur Verteidigung homunkularer Redeweisen nicht aus. Freilich läßt sich aus dem bloßen Vorkommen einer Metapher nicht schließen, daß eine Erklärung erschlichen wurde; dies muß jeweils im Einzelfall geprüft werden (zur kognitiven Funktion von Metaphern und Modellen vgl. Keil 1993, 249-271). Es gibt kein kanonisches Testverfahren dafür, wie harmlos eine Metapher ist. Die sprachkritischen Instrumente, deren sich Homunkuluskritik bedienen muß, erfordern eine sorgfältige Handhabung.

<sup>17</sup> Lycan bezeichnet den Homunkularen Funktionalismus deshalb auch als *metaphysische* Doktrin, welche „solves the mind/body problem by type-identifying mental states and events with homuncular states and events“ (ebd., 263).

Temperaturregulation). Und selbst einige chemische Elemente werden nach ihren Funktionen benannt.<sup>18</sup> Lycan fährt fort:

„Purely . . . what? We are in search of prime matter here [...]. And that stuff, if there is any, can be characterized in either of only two ways: by contrast with the functional, at some chosen level of abstraction, as in ‘the stuff that cells are made of,’ or by direct reference to a specific kind of level-bound entity, such as molecules or atoms or quarks and gluons. In neither way do we succeed in isolating the desired *general* mode of purely nonfunctional characterization, the vernacular of *pure occupancy*.” (ebd., 47 f.)

Zugrunde liegende Materie – man kann hier an das *hypokeimenon* des Aristoteles denken – lasse sich durch Wechsel auf eine basalere Beschreibungsebene stets wieder funktional charakterisieren, „and so we have no ordinary word for pure occupancy. ‘Role’/‘occupant’ remains a level-relative distinction.“ (ebd., 48).

Lycan formuliert deshalb ein bescheideneres Erklärungsziel als Dennett. Er möchte den Geist nicht auf das „Mechanische“ oder „Physische“ zurückführen, weil ihm im Zuge seiner Kritik des „two-levelism“ unklar ist, was das sein soll. Statt dessen möchte er nur zeigen „that the mystery of the mental is *no greater than* the mystery of the heart, the kidney, the carburetor or the pocket calculator“ (ebd., 44). Man kann sagen, daß Lycan sich mit einem Abstieg vom intentionalen Standpunkt zum *design stance* begnügt, weil er den basalen, teleologiefreien *physical stance* für einen Mythos hält.

Lycans fröhliches Bekenntnis zu seinen „panteleological tendencies“ ist indes nicht sein letztes Wort. An einigen Stellen gesteht er zu, daß es noch einen weiteren Erklärungsbedarf geben könnte: „I hope, and I am inclined to believe, that the teleological characterizations that Homunctionalism requires can be independently explained in evolutionary terms“ (ebd., 43). Mit dieser Hoffnung ist Lycan nicht allein; der teleofunktionalistische Rückgriff auf evolutionstheoretische Erklärungen scheint die Methode der Wahl zu sein, homunkulare und teleologische Charakterisierungen wenn schon nicht loszuwerden, so doch naturwissenschaftlich respektabel zu machen. Galt noch Kant die Kategorie des Zweck als „Fremdling in der Naturwissenschaft“, so lautet das Argument der Teleofunktionalisten nun, daß die

<sup>18</sup> Lycan berichtet genüßlich, wie ihm Ned Block einmal entgegengehalten habe: „I’ll give you *neurons* and *cells* and so on as functional, but when you come to *hydrogen* and *oxygen*, when you get right down to the level of *chemistry*, there’s just nothing functional or teleological at all!!!“ Lycan: „Oh no? ‘Hydro’-*what?* ‘Oxy’-*what?* (The shot is a cheap one but immensely satisfying.)“ (Lycan 1987, 139, Anm. 19).

Biologie als eine respektable Wissenschaft „is already shot through with ascriptions of natural teleology and that such ascriptions are not going to go away, for without them we would lose valuable generalizations“ (Lycan 1991, 264).

Doch die evolutionstheoretische Rehabilitierung der „natürlichen Teleologie“ ist ein Bücherschrank für sich. Fodor, der zeitweise selbst mit einer teleofunktionalistischen Auszeichnung der „Normalbedingungen“ sympathisiert hat, unter denen ein Zustand eines Organismus erst einen *bestimmten* Repräsentationsgehalt habe, hält später eine naturalistische Erklärung oder Analyse eines intentionalen Phänomens nur dann für gelungen, wenn in nichtmentalen, nichtsemantischen *und nichtteleologischen* Begriffen hinreichende Bedingungen für das Vorliegen des Phänomens angegeben sind (vgl. Fodor 1987, 98 und 126). Große Schwierigkeiten macht auch die Parallelisierung teleologischer mit kausalen Beziehungen: Erklärbar scheint die kausale Wirksamkeit von zu erreichenden Zielzuständen nur im Falle des zielgerichteten Handelns geistbegabter Wesen zu sein, die ihre Handlungsziele mental antizipieren können, so daß *diese Antizipationen* kausal wirksam werden.<sup>19</sup>

Wir sollten noch etwas genauer und ohne Rückgriff auf Lycans panteleologische Naturphilosophie zu sagen versuchen, was an Dennetts schrittweiser Zurückführung homunkularer Charakterisierungen auf „obviously mechanistic task descriptions“ verfehlt ist. Dennetts elementarste Aufgabenschreibung, die die dümmsten Homunkuli zu „functionaries ‘who can be replaced by a machine’“ degradiert, lautete: „to pick the larger of two numbers when directed to“. Der entscheidende Punkt ist, daß die fraglichen Instanzen, wie schon bei Fodor<sup>20</sup>, *Adressaten von Anweisungen* bleiben.

Solange dies der Fall ist, ist der Übergang von einer intentionalen zu einer nichtintentionalen Charakterisierung einer kognitiven Leistung nicht vollzogen. Der entscheidende Unterschied ist nicht der zwischen anspruchsvollen und weniger anspruchsvollen Aufgaben oder der zwischen intelligenten und weniger intelligenten Funktionsträgern, sondern der zwischen der Rede von zu lösenden Aufgaben oder zu befolgenden Anweisungen und einer nichtfunktionalen, physikalischen Beschreibung der elementaren Vorgänge. Lycan leugnet, daß es die letztere überhaupt gibt; ich beschränke mich hier auf den Einwand gegen Dennett. Solange die Rolle der dümmsten Homunkuli so beschrieben wird, daß sie eine ihnen zugewiesene Aufgabe erfüllen, wie anspruchslos auch immer, kann von einer „klarerweise mechanisti-

<sup>19</sup> Zur Kritik der evolutionstheoretischen Rehabilitierung der Teleologie vgl. Keil 1993, 107-128, 299-359.

<sup>20</sup> Fodor sprach von „a class of elementary instructions which the nervous system is specifically wired to execute“ (s.o.).

schen“ Beschreibung nicht die Rede sein. Um es pointiert zu sagen: Ebenso wie dumme Menschen eben Menschen bleiben, bleiben auch dumme Homunkuli Homunkuli.

Es ist der letzte Schritt seiner rekursiven Dekomposition, den Dennett nicht plausibel machen kann. Was die graduelle Abstufung bis zu einem Nullwert im Falle intentionaler und repräsentationaler Phänomene bedeuten könnte, ist alles andere als klar. Cummins, Margolis und Haldane haben ähnliche Kritik an Dennett geübt. Cummins wendet ein, daß

„... intentionality isn't a matter of degree. If we have real homunculi – i.e., intentionally characterized capacities – making them dumber won't make them less intentional [...]. [W]e have no idea how to achieve that last crucial step: we don't know how to computationally instantiate a real homunculus, however stupid” (Cummins 1983, 92).

Cummins sieht durch Dennetts fehlschlagenden Versuch, die intentionale Erklärungsebene schrittweise zu verlassen, den bekannten Verweisungszirkel intentionaler Begriffe bestätigt:

„Intentionally characterized capacities can be explained by analysis into other intentionally characterized capacities. But the analyzing capacities will not be less intentional than the analyzed capacity.“ (ebd.).<sup>21</sup>

Daß die Rede von zu lösenden Aufgaben und zu befolgenden Anweisungen intentional charakterisierte Fähigkeiten voraussetzt, leuchtet vielleicht nicht unmittelbar ein. Das zugrunde liegende Problem ist häufig am Begriff des *Regelfolgens* diskutiert worden. Wer fähig ist, einer Regel zu folgen, dem müssen zugleich die Fähigkeiten zugestanden werden, die Befolgung zu verweigern oder die Regel mißzuverstehen. Dafür muß es gewisse Freiheitsgrade in seinem Verhalten geben; er muß die Fähigkeit der Ja/nein-Stellungnahme besitzen. Der Grund dafür ist ein begrifflicher: Würde man demjenigen, der einer Regel oder einer Anweisung folgen soll, die Fähigkeit zur Verweigerung nicht zubilligen, so gäbe es keinen Grund dafür, das

<sup>21</sup> Verwandte Kritik übt Margolis, der Dennett mit folgendem Dilemma konfrontiert: „If [the homunculi] are merely stupid, his reductive program fails, because their behavior will be explained by means of the same psychological model [...] that he wishes to eliminate [...]; and if they are less than stupid, utterly lacking in intentionally qualified states, then his reductive program will fail again, because neither he nor we know of any way to reduce the intentional to the non-intentional“ (Margolis 1980, 255). Und John Haldane merkt an: „The philosophical problem of meaning and intentionality is not that of showing how things can have *so much* of them but rather one of explaining how they can have them *at all*“ (Haldane 1988, 129).

Idiom der Regeln und Anweisungen überhaupt in Anschlag bringen. Dennett spricht von Homunkuli „who can be replaced by a machine“. Eine echte Maschine kann man aber zu nichts auffordern; hier genügt die Betätigung der Bedienungsknöpfe. Aufforderungen haben nur einen Sinn, wenn der Adressat zu dem fraglichen Verhalten nicht *physisch determiniert* ist: „The point is that if I am causally constrained from breaking a rule (or, for that matter, from obeying it), then I can no longer obey or break the rule at all“ (Hyman 1991, 16). Wenn für einen Mechanismus in einer gegebenen Situation ein anderes als sein tatsächliches Verhalten physisch unmöglich ist, kann man ihm nicht sinnvoll die Befolgung einer Regel zuschreiben.<sup>22</sup>

Nun ist wiederholt versucht worden, einen weniger anspruchsvollen Regelbegriff ins Spiel zu bringen. Ein früher Versuch stammt von Sellars, der zwischen den Phänomenen des „merely conforming to rules“ und des „obeying rules“ unterscheidet. Im Falle des bloß regelkonformen Verhaltens gelte, daß „these doings ‘just happen’ to contribute to the realization of a complex pattern“, während das echte Regelfolgen „with the intention of fulfilling the demands of an envisaged system of rules“ vollzogen werde (Sellars 1963, 325). Bloß regelkonformes Verhalten läßt sich schon konstatieren, wenn das Verhalten des Systems eine bestimmte Regelmäßigkeit zeigt. Ein intelligenter Beobachter kann dieses Verhalten so beschreiben, *als ob* es bestimmten Regeln (Anweisungen) folgte; diese sind aber im System selbst nicht repräsentiert. Mit dieser *Als ob*-Interpretation wäre allerdings die *Realismthese* des Repräsentationalismus preisgegeben. Daß kein physischer Prozeß intrinsisch repräsentierend und regelbefolgend ist, sondern erst im Auge des Betrachters als ein solcher erscheint, war ja gerade der Einwand Searles und anderer gegen Fodors repräsentationalistische Theorie.<sup>23</sup>

Manche Autoren sprechen, um die Bedingungen für Regelfolgen abzuschwächen, von „impliziten“ Regeln, die „stillschweigend“ befolgt würden. Gegen diesen Vermittlungsversuch hat schon Quine eingewandt, daß implizites Regelfolgen von bloß regelmäßigem Verhalten kaum zu unterscheiden sei. Da sich aber zu jedem Verhalten irgendeine Regelmäßigkeit (ein Muster, eine Ordnung) finden läßt, die durch es exemplifiziert wird, wäre jedes beliebige Verhalten implizit regelbefolgend. Diese Konsequenz käme einer *reductio ad absurdum* des Regelidioms gleich.

<sup>22</sup> Mir ist bewußt, daß ich mich damit auf die kontroverse Annahme verpflichte, daß Entscheidungsfreiheit Indeterminiertheit in dem Sinne erfordert, daß es vor der tatsächlichen Entscheidung keine kausal hinreichenden Bedingungen dafür gab, wie die Entscheidung ausfällt.

<sup>23</sup> Cummins (1983, 46) verwendet übrigens den Repräsentationsbegriff ausdrücklich trivialisiert im Sinne von „Exemplifikation“, und warnt davor, dieses „being represented by a device“ zu verwechseln mit „being represented for a device“ (ebd., 49).

Was die Rechtfertigung des verwendeten Begriffsrahmens betrifft, stellt der homunkulare Funktionalismus gegenüber anderen Spielarten des Repräsentationalismus keinen Fortschritt dar. Allgemein ist auf einer sorgfältigen Begründung dafür zu bestehen, ein System überhaupt als *Kandidaten* für die Zuschreibung von Repräsentationen zu akzeptieren. Es ist eine irriige Annahme, physische Zustände selbst ließen sich unmittelbar daraufhin überprüfen, ob sie repräsentationale Gehalte besitzen und gegebenenfalls welche. Vorstellungen zu haben ist keine Eigenschaft beliebiger Zustände, sondern zunächst eine Fähigkeit von Personen und vielleicht von anderen höheren Lebewesen. Ob noch weitere „Systeme“ als logische Subjekte oder Träger von Vorstellungen, Wünschen, Überzeugungen und Absichten in Frage kommen, läßt sich nicht allein durch eine Beobachtung des „Systemverhaltens“ feststellen. Man kann einem Systemverhalten nicht ansehen, als was es zu interpretieren und welches Vokabular dafür in Anschlag zu bringen ist.<sup>24</sup> Schon die Rede von erbrachten „Leistungen“ nach der Analogie menschlicher Leistungen beruht auf einer funktionalen Charakterisierung einer zu erfüllenden Aufgabe und kann insofern nicht zur einer unabhängigen *Rechtfertigung* des Repräsentationalismus dienen.

Die homunkularen Funktionalisten richten ihre Bemühungen vornehmlich darauf, den Regreßeinwand auszuräumen. Ein Erklärungsregreß ergibt sich nach ihrer Auffassung nur, wenn subpersonale Instanzen wie ganze Personen aufgefaßt werden. Eben dies wird bei der Kritik von Homunkulus-Fehlschlüssen gewöhnlich unterstellt. Doch hat diese Unterstellung einen guten Grund, der bisher noch nicht zur Sprache gekommen ist. Die Kritiker des Homunkulismus argumentieren wie folgt: Kognitive Fähigkeiten kommen in Personen nicht als einzelne vor, sondern im Verbund. Es ist schwer zu sehen, wie eine subpersonale Instanz die ihr jeweils zugeschriebene Fähigkeit besitzen kann, ohne daneben noch viele weitere Fähigkeiten von Personen zu haben. Um beispielsweise in der von Freud beschriebenen Weise mit den anderen Instanzen einen Konflikt austragen zu können, müßte eine psychische Instanz eigene Interessen oder Antriebe haben, Anweisungen befolgen, sich ihnen widersetzen sowie sich zu etwas entscheiden können. Man kann das, etwas aufwendig, die *holistischen Zuschreibungsbedingungen mentaler Fähigkeiten* nennen.<sup>25</sup> Man wollte die Instanz

<sup>24</sup> Vgl. dazu am Beispiel der „Roboterantwort“ Keil 1998, bes. 114-122.

<sup>25</sup> Der Holismus mentaler Fähigkeiten ist nicht zu verwechseln, aber verwandt mit dem von Davidson beschriebenen Holismus der Zuschreibung einzelner propositionaler Einstellungen, also der *Ausübung* mentaler Fähigkeiten. Insofern Vorstellungen, Wünsche, Absichten und Überzeugungen über ihre propositionalen Gehalte individuiert werden, können sie nicht isoliert zugeschrieben werden, sondern nur im

eigentlich nur mit bescheidenen Talenten ausstatten; um diese aber plausibel zu machen, muß man ihr noch weitere Fähigkeiten zugestehen (vgl. zu diesem Argument Hacker 1991, 128 f.). Die subpersonale Instanz müßte, mit einem Wort, eine eigene Psyche haben. Wenn sie aber eine eigene Psyche haben müßte, um überhaupt den ihr zugeordneten Beitrag leisten zu können, liegt ein Regreß vor, denn die Funktionsweise der Psyche sollte ja gerade erklärt werden.

Nehmen wir an, der homunkulare Funktionalist könne sich der Einsicht in die holistischen Zuschreibungsbedingungen mentaler Fähigkeiten nicht ganz verschließen. Er könnte gleichwohl der Auffassung sein, daß das holistische Argument seine Position nicht trifft. Ihm geht es ja um die *Analyse* mentaler oder kognitiver Fähigkeiten, also um eine gedankliche Operation, nicht um eine tatsächliche Abtrennung. Daß mentale Fähigkeiten stets im Verbund instantiiert sind, hindere nicht daran, in diesem Bereich analytische Unterscheidungen zu treffen, bestimmte elementare Fähigkeiten auszuheben und sie funktional individuierten Subsystemen zuzuordnen. Für Dennett ist *jeder* Funktionalismus in diesem Sinne eine Spielart des Homunkulismus.<sup>26</sup>

Doch mit Cummins, Margolis und Haldane habe ich oben eingewandt, daß Dennett den letzten Schritt seiner rekursiven Dekomposition der zu erfüllenden Aufgabe nicht plausibel machen kann. Der Einwand lautet, daß man selbst die *elementarste* Fähigkeit, mit der Dennett seine dümmsten Homunkuli ausstattet, Artefakten, Gehirnen und subpersonalen Instanzen nicht sinnvoll zuschreiben kann. Diese elementarste Fähigkeit ist das Befolgen einer einfachen Anweisung (Dennetts Beispiel: „to pick the larger of two numbers when directed to“). Ich habe argumentiert, daß für die Fähigkeit zum Befolgen einer Anweisung die Freiheit angenommen werden muß, sich für oder gegen etwas zu entscheiden. Diese Auffassung sei noch etwas näher ausgeführt.

Unter Verzicht auf den Freiheitsbegriff könnte man sagen, daß Artefakte, Gehirne oder Teile von Personen etwas *tun* können müßten, um einer Anweisung zu folgen oder eine Aufgabe zu erfüllen. Dies scheint eine sehr anspruchslose Bedingung zu sein. Wir reden zwanglos so, wenn wir natürliche Prozesse oder Veränderungen mithilfe von Verben beschreiben.<sup>27</sup>

Verbund mit weiteren Einstellungen, zu denen sie in logischen und inferentiellen Beziehungen stehen.

<sup>26</sup> „In fact *all* varieties of functionalism can be viewed as ‘homuncular’ functionalism of one grain size or another“ (Dennett 1991, 262).

<sup>27</sup> „Man könnte auch von einer Tätigkeit der Butter reden, wenn sie im Preise steigt; und wenn dadurch keine Probleme erzeugt werden, so ist es harmlos“ (Wittgenstein, PU § 693).

Verben stehen nun einmal für Tätigkeiten. Doch es gibt hier ein tiefes metaphysisches Problem. Eine Aussage wie „Salzsäure verätzt die Haut“ suggeriert durch ihre Parallelität mit echten Handlungssätzen eine *Aktivität* der Salzsäure, die keinen Grund in der Sache hat. Tatsächlich findet hier ein Naturvorgang statt, der nur durch die Nötigung der Grammatik in einen aktiven und einen passiven Pol aufgespalten wird. Die Nötigung, allen Vorgängen ein Agens zu unterlegen, haben Nietzsche und Mauthner auf eine „Krankheit der Sprache“ zurückgeführt, auf die der Sprache innewohnende „älteste und längste Psychologie“, die die „großen Irrthümer“ der Metaphysik erzeugt habe: „[A]lles Geschehen war ihr ein Thun, alles Thun Folge eines Willens, die Welt wurde ihr eine Vielheit von Thätern, ein Thäter (ein ‘Subjekt’) schob sich allem Geschehen unter.“ (Nietzsche 1889, 85)<sup>28</sup>

Nun spricht die Sprache anderslautenden Behauptungen zum Trotz nicht selbst. Sprecher können gemäß Searles „Prinzip der Ausdrückbarkeit“ stets die Unterschiede machen, auf die es ihnen ankommt. Allgemein werden Naturvorgänge nicht dadurch zu intentionalen Handlungen, daß sie faktisch mit Handlungsverben beschrieben werden. Es bleibt der jüngst von Peter Rohs herausgestellte Unterschied, daß echte Handlungen mit Verben beschrieben werden *müssen*, während man Naturvorgänge mit Verben beschreiben *kann*, aber auch verbfrei, ohne daß irgendetwas verlorengehe. Daß ein Vorgang stattfindet, läßt sich stets so beschreiben, daß nacheinander verschiedene Zustände vorliegen:

„Wenn irgendwelche Gegenstände durch Raum-Zeit-Gebiete individuiert werden können, dann können ihre Zustände verbfrei wiedergegeben werden, letztlich durch eine auf dem Gebiet definierte stetige Zustandsfunktion.“ (Rohs 2000, 19)

Wir sind somit nicht durch die Grammatik zu der Annahme gezwungen, daß Naturdinge handeln oder etwas tun können.<sup>29</sup> Tatsächlich ist auch das elementare Vermögen, etwas zu tun, keine Fähigkeit, die man *beliebigen* Dingen zuschreiben kann. Wenn eine nicht-agentivierende, intentionalitätsfreie Beschreibung eines Vorgangs gewünscht ist, stehen sprachliche Mittel dafür zur Verfügung.

Ich habe an anderer Stelle dafür argumentiert, das Handlungsvermögen als den Angelpunkt einer Theorie der Intentionalität anzusehen. Von „Intention-

<sup>28</sup> Zum Problemkomplex ‘Anthropomorphismus in der Sprache’ vgl. Keil 1993, bes. 238-249.

<sup>29</sup> „Handeln“ verwende ich im Sinne von „absichtlich etwas tun“. Das Tun unterscheidet sich seinerseits durch seine *potentielle Absichtlichkeit* von bloßen Körperbewegungen und Widerfahrnissen (vgl. dazu Keil 2000b, 136-143).

nalität“ spricht man ja einerseits im Sinne des Handelnkönnens nach Absichten (Intentionalität<sub>1</sub>), andererseits im Sinne des Gerichtetsein des Bewußtseins auf Gegenstände (Intentionalität<sub>2</sub>). Meiner Auffassung nach ist Intentionalität<sub>1</sub> nicht nur ein Spezialfall, sondern das Paradigma und die Explikationsbasis für Intentionalität<sub>2</sub>. Die intentionalen Akte des Bewußtseins müssen selbst nach dem Modell von Handlungen begriffen werden, weil anders das Moment von Aktivität nicht zu rekonstruieren ist, das im Begriff des Sich-auf-etwas-Richtens steckt (vgl. Keil 1993, 357 f. und 383).

#### 6. Philosophische Sinnkritik und empirische Wissenschaft

Homunkulismuskritik gegenüber kognitionswissenschaftlichen Theorien ist ein Beispiel sinnkritischen Philosophierens und zieht als solches den Unwillen derer auf sich, die für eine Nichteinmischung der Philosophie in die inneren Angelegenheiten der Wissenschaften plädieren. Es wird heute gern gesagt, die Philosophie müsse sich auf vielen Gebieten von den Einzelwissenschaften informieren lassen, wo doch auf der anderen Seite nicht zu sehen sei, was eigentlich die Wissenschaften noch von der Philosophie lernen könnten. Methodologisch sorgten die Wissenschaften längst für sich selbst, und am erfolgreichsten seien sie, wenn sie ungestört ihrer Arbeit nachgehen können, ohne Einmischung von *armchair*-Philosophen.<sup>30</sup> Wenn man indessen der altmodischen Auffassung von Wissenschaft als einer kooperativen, methodischen Wahrheitssuche anhängt, ist diese Rede von Einmischung oder Bevormundung schwer nachzuvollziehen. Im großen Reich der Wissenschaft, zu dem ich aus Gründen, auf die ich hier nicht eingehen kann, auch die Philosophie zähle, sollte es keine inneren Angelegenheiten geben, in die einzumischen man sich verbitten könnte. John Hyman hat zur Aufgabe der Philosophie folgendes gesagt: „A major part of the philosopher’s business is to disentangle conceptual puzzles that have been woven into the fabric of empirical research“ (Hyman 1989, xiv).

Das ist gut gesagt und benennt eine wichtige Rolle, die die Philosophie für die Einzelwissenschaften spielen kann. Über die empirische Psychologie seiner Zeit hat Wittgenstein boshaft angemerkt: „Es bestehen nämlich, in der Psychologie, experimentelle Methoden *und* *Begriffsverwirrung*“ (PU II, xiv, 370). Begriffsverwirrungen aber wirken sich überall nachteilig aus, und ihre Vermeidung muß in jedermanns Interesse liegen. Homunkulus-Fehlschlüsse schaden auch den empirischen Wissenschaften selbst, denn sie können gegenstandslose Fragen veranlassen und degenerierende For-

<sup>30</sup> Vgl. z.B. Tetens 2000, 287 f. Zur folgenden Kritik vgl. Keil/Schnädelbach 2000, 32-38.

schungsprogramme anregen. Das Problem des invertierten Netzhautbildes ist ein vorzügliches Beispiel dafür. Das Invertierungsproblem war ja keine wissenschaftliche Entdeckung, sondern die Folge einer undurchschauten Implikation des Bildbegriffs: Hier ist ein Bild; es zeigt die Außenwelt invertiert; wie kommen wir von diesem Bild zu einem korrekten Bild der Gegenstände? Man hat gedacht, daß dieses Problem durch die empirischen Befunde über den Strahlengang im Auge erzwungen wurde, und erst durch scharfes Nachdenken konnte man darauf kommen, daß, mit Wittgenstein gesagt, „ein Bild uns gefangen hielt“, in diesem Falle das Bild des Netzhautbildes.

Die Behauptungen über den intentionalitätspräsupponierenden Charakter der Konzepte *Regelbefolgung*, *Repräsentation*, *Berechnung* und *Informationsverarbeitung* dürften ungleich kontroverser sein. Doch wodurch sie nicht ausgehebelt werden können, ist eine generelle Inschutznahme kognitionswissenschaftlicher Theorien gegen philosophische Einmischung. Das Computermodell des Geistes ist alles andere als empirische Wissenschaft, ebensowenig ist es der Repräsentationalismus. Und ob homunkulare Redeweisen sich rechtfertigen lassen, ist keine Frage der empirischen Entdeckung, wie das Gehirn oder ein Computer funktionieren. Homunkulismuskritik betrifft nicht die empirischen Befunde, sondern die oft undurchschauten begrifflichen Voraussetzungen und Implikationen kognitionswissenschaftlicher Theorien. Die Schwierigkeit besteht darin, daß beides oft verwoben ist; so entstehen die von Hyman erwähnten unbemerkt in empirische Forschung eingeflochtenen Begriffsverwirrungen.

Die sinnkritische Begleitung wissenschaftlicher Theoriebildung ist natürlich kein Privileg professioneller Philosophen. Nichts spricht dagegen, daß Einzelwissenschaftler sich daran beteiligen. Allerdings sind Fachphilosophen dafür in der Regel besser ausgebildet.

#### Literatur

Attneave, Fred (1961): In Defense of Homunculi, in: Rosenblith, 777-782.

Beckermann, Ansgar (1995): Visuelle Informationsverarbeitung und phänomenales Bewußtsein, in: Metzinger, 663-679.

Beckermann, Ansgar (1997): Ist eine Sprache des Geistes möglich?, in: Burri, 75-92.

Boolos, G.S. / Richard C. J. (1989): *Computability and Logic*, Cambridge MA.

Burri, Alex (Hg.) (1997): *Sprache und Denken – Language and Thought*, Berlin/New York.

Churchland, Patricia (1986): *Neurophilosophy*, Cambridge MA.

- Clark, P. / Wright, C. (Eds.) (1998): *Mind, Psychoanalysis and Science*, New York.
- Cummins, Robert (1983): *The Nature of Psychological Explanation*, Cambridge MA.
- Dennett, Daniel C. (1969): *Content and Consciousness*, London.
- Dennett, Daniel C. (1978): *Brainstorms*, Hassocks, Sussex.
- Dennett, Daniel C. (1991): *Consciousness Explained*, Boston/Toronto/London.
- Dölling, E. (Hg.) (1998): *Repräsentation und Interpretation*, Berlin.
- Elsner, N./Lüer, G. (Hg.) (2000): *Das Gehirn und sein Geist*, Göttingen.
- Engel, A./Gold, P. (Hg.) (1998): *Der Mensch in der Perspektive der Kognitionswissenschaften*, Frankfurt am Main.
- Fodor, Jerrold A. (1968): The Appeal to Tacit Knowledge in Psychological Explanation, *Journal of Philosophy* 65 (1968), 627-640.
- Fodor, Jerrold A. (1975): *The Language of Thought*, New York.
- Fodor, Jerrold A. (1987): *Psychosemantics*, Cambridge MA.
- Hacker, Peter (1991): Seeing, Representing and Describing: An Examination of David Marr's Computational Theory of Vision, in: Hyman, 119-153.
- Haldane, John (1988): Psychoanalysis, Cognitive Psychology and Self-Consciousness, in: Clark/Wright, 113-139.
- Hyman, John (1989): *The Imitation of Nature*, Oxford.
- Hyman, John (Ed.) (1991): *Investigating Psychology: Sciences of the Mind after Wittgenstein*, London.
- Keil, Geert (1993): *Kritik des Naturalismus*, Berlin/New York.
- Keil, Geert (1998): Was Roboter nicht können, in: Engel/Gold, 98-131.
- Keil, Geert (2000a): Naturalismus und Intentionalität, in: Keil/Schnädelbach, 187-204.
- Keil, Geert (2000b): *Handeln und Verursachen*, Frankfurt am Main.
- Keil, Geert (2003): Über den Homunkulus-Fehlschluß, *Zeitschrift für philosophische Forschung* 57, 1-26.
- Keil, G. / Schnädelbach, H. (2000): Naturalismus, in: Keil/Schnädelbach, 7-44.
- Kemmerling, Andreas (2000): Ich, mein Gehirn und mein Geist: Echte Unterschiede oder falsche Begriffe?, in: Elsner/Lüer, 223-241.
- Kenny, Anthony (1971): The Homunculus Fallacy, zitiert nach: Hyman, 155-165.
- Kosslyn, Stephen Michael (1983): *Ghosts in the Mind's Machine. Creating and Using Images in the Brain*, New York/London.
- Lycan, William (1987): *Consciousness*. Cambridge MA.
- Lycan, William (1991): Homuncular Functionalism Meets PDP, in: Ramsey/Stich/Rumelhart, 259-286.
- Margolis, Joseph (1980): The Trouble with Homunculus Theories, *Philosophy of Science* 47, 244-259.
- Marr, David (1982): *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco.

- Metzinger, Thomas (Hg.) (1995): *Bewußtsein. Beiträge aus der Gegenwartsphilosophie*, Paderborn.
- Nietzsche, Friedrich (1889): *Götzendämmerung*, in: *Werke. Kritische Gesamtausgabe* (Colli/Montinari), Bd. VI/3, Berlin 1969, 49-157.
- Penfield, W. / Boldrey, E. (1937): Somatic Motor and Sensory Representation of Man as Studied by Electrical Stimulation, *Brain* 60, 389-443.
- Platon: *Phaidros* (Sämtliche Werke IV, Übers. Schleiermacher), Hamburg 1958.
- Platon: *Phaidon* (Sämtliche Werke III, Übers. Schleiermacher), Hamburg 1958.
- Rohs, Peter (2000): Sind Handlungen Ereignisse?, *Zeitschrift für philosophische Forschung* 54, 1-22.
- Rosenblith, Walter A. (Ed.) (1961): *Sensory Communication*, Cambridge MA.
- Ramsey, W./Stich, S.P./Rumelhart, D. (Eds.) (1991): *Philosophy and Connectionist Theory*, Hillsdale.
- Ryle, Gilbert (1949): *Der Begriff des Geistes*, Stuttgart 1969.
- Ryle, Gilbert (1954): Die Wahrnehmung, in ders., *Begriffskonflikte*, Göttingen 1970, 117-138.
- Schafer, Roy (1982): *Eine neue Sprache für die Psychoanalyse*, Stuttgart.
- Schumacher, Ralph (1998): Der Homunkulus-Fehlschluß in der Philosophie des Geistes. Eine Verteidigung des Repräsentationalismus, in: Dölling, 39-60.
- Searle, John (1980): Minds, Brains, and Programs, *Behavioral and Brain Sciences* 3, 417-458.
- Searle, John (1990): Is the Brain a Digital Computer?, *Proceedings and Addresses of the American Philosophical Association* 64/3, 21-37.
- Sellars, Wilfrid (1963): *Science, Perception and Reality*, London.
- Shannon, Claude E. / Weaver, Warren (1949): *The Mathematical Theory of Communication*, Urbana.
- Tetens, Holm (2000): Der gemäßigte Naturalismus der Wissenschaften, in: Keil/Schnädelbach, 273-288.
- Tibbetts, Paul (1995): Neurobiology and the Homunculus Thesis, *Man and World* 28, 401-413.
- van Gulick, Robert (1989): What Difference Does Consciousness Make?, *Philosophical Topics* 17, 211-230.
- von Weizsäcker, Carl Friedrich (1971): *Die Einheit der Natur*, München/Wien.
- Wittgenstein, Ludwig (1960): *Philosophische Untersuchungen* (PU), Frankfurt am Main.
- Zeki, Semir (1993): *A Vision of the Brain*, Cambridge.