

# Naturalismuskritik und Metaphorologie

Geert Keil

## 1 Einleitung

In natürlicher Sprache formulierte Theorien über welchen Gegenstandsbereich auch immer zeichnen sich wesentlich durch ihre zentralen Begriffe aus. In der Begrifflichkeit einer Theorie spiegeln sich ihre Klassifikationen und gegebenenfalls die angenommenen natürlichen Arten wider. Da von den natürlichen Arten unter anderem abhängt, welche induktiven Schlüsse möglich sind, kann man ohne Übertreibung sagen, dass die zentralen Begriffe einer Theorie einen Teil ihrer Erklärungslast tragen.

Eine *naturalistische* Theorie beansprucht, die von ihr behandelten Phänomene als Teile der natürlichen Welt verständlich zu machen, und das heißt im szientifischen Naturalismus, die Phänomene mit den Begriffen und Methoden der Naturwissenschaft zu erklären.<sup>1</sup> Bei naturalistischen Theorien über den Menschen geht es um Phänomene wie Geist, Sprache, Vernunft, Bewusstsein, Handlungsfähigkeit und Normativität. Naturalistische Erklärungen dieser Phänomene dürfen nicht auf begriffliche Ressourcen jenseits der Naturwissenschaften zurückgreifen. Was das im Einzelnen bedeutet, ist nicht leicht zu sagen. Folgendes sollte aber unkontrovers sein: Wenn gezeigt werden könnte, dass vermeintlich naturalistische Theorien humanspezifischer Phänomene an irgendeiner Stelle, und sei es an gut verborgener, unanalyzierte und nichteliminierbare anthropomorphe Metaphern enthalten, dann wäre das ein gravierender Einwand gegen den behaupteten naturalistischen Charakter der fraglichen Theorie. Wenn darüber hinaus gezeigt werden könnte, dass anthropomorphe Metaphern in naturalistischen Theorien nicht aus Unachtsamkeit und sporadisch, sondern aus Erklärungsnot und endemisch vorkommen, dann wäre das ein Grund zur Skepsis gegen die Erfolgsaussichten des Großprojekts der Naturalisierung des Menschen und seiner Fähigkeiten.

Eine entsprechende Rolle der Metaphorologie für die Naturalismuskritik habe ich vor vielen Jahren in meinem Buch *Kritik des Naturalismus* vor-

---

<sup>1</sup> Ich unterscheide zwischen drei Arten des Naturalismus (vgl. z. B. Keil 2008). Der *metaphysische* Naturalismus sieht den Menschen mit allen seinen Eigenschaften und Fähigkeiten als Teil der natürlichen Welt an, der *szientifische* Naturalismus behauptet einen Erklärungsprimat naturwissenschaftlicher Methoden, der *semantisch-analytische* Naturalismus behauptet, dass sich hinreichende Bedingungen für mentale Phänomene in nichtintentionalen, nicht-semantischen und nichtteleologischen Begriffen angeben lassen.

geschlagen.<sup>2</sup> Man wird nicht sagen können, dass sich dieser Ansatz der Naturalismuskritik in der Forschung durchgesetzt hat. Ich nutze die Gelegenheit, einige Grundgedanken dieses Ansatzes noch einmal auf den Prüfstand und zur Diskussion zu stellen.

## 2 *Metaphertheorie und Metaphorologie*

Zur Terminologie: Die *Metaphertheorie* wird arbeitsteilig von der Linguistik und der Sprachphilosophie betrieben und untersucht die Funktionsweise metaphorischen Sprachgebrauchs, wobei in der zweiten Hälfte des 20. Jahrhunderts die bedeutungstheoretische Perspektive die früher vorherrschende rhetorische weitgehend abgelöst hat. Es geht in der Metaphertheorie um Fragen wie: Was sind überhaupt Metaphern? Welche Typen von Metaphern lassen sich unterscheiden? Wie lässt sich die Bedeutungskonstitution von Metaphern erklären? Gibt es einen semantischen „Mechanismus“ der Metapher? Lässt sich metaphorische Rede durch synonyme oder kognitiv äquivalente nichtmetaphorische Rede ersetzen? Warum werden Metaphern überhaupt verstanden? Erhebt metaphorische Rede Wahrheitsansprüche? Einflussreiche sprachphilosophische Beiträge zur Semantik und Pragmatik der Metapher stammen von Max Black, Monroe Beardsley, Nelson Goodman, John Searle, Donald Davidson und Paul Ricoeur.<sup>3</sup>

Im Unterschied zur Metaphertheorie ist die *Metaphorologie* kein im engeren Sinne sprachtheoretisches Unternehmen, sondern ein erkenntnistheoretisches. Im deutschen Sprachraum ist die Bezeichnung „Metaphorologie“ mit den Arbeiten von Hans Blumenberg verbunden. Blumenbergs langjähriges Projekt einer „historischen Metaphorologie“ ist als Teil einer „Theorie der Unbegrifflichkeit“ darauf angelegt, Ideen- und Geistesgeschichte als Geschichte von erkenntnisleitenden Hintergrundmetaphern zu begreifen: die Welt als ein Buch zum Lesen, das Leben als Schifffahrt, die Zeit als Fluss, der Mensch als Maschine, die nackte Wahrheit, das Licht der Erkenntnis. Metaphern sind nach Blumenberg „eine authentische Leistungsart der Erfassung von Zusammenhängen“ (1979:77) und zeigen eine „noch nicht konsolidierte Begriffsbildung“ an (1971:171). Dabei sehe man einer Aussage ihren metaphorischen Charakter nicht immer auf den ersten Blick an: „Metaphorik kann auch dort im Spiele sein, wo ausschließlich terminologische Aussagen auftreten, die aber ohne Hinblick auf eine Leitvorstellung, an der sie induziert und ‚abgelesen‘ sind, in ihrer umschließenden Sinneinheit gar nicht verstanden werden können“ (1960:69).

<sup>2</sup> Vgl. Keil 1993, bes. 229–359.

<sup>3</sup> Eine Klassifikation und Kurzcharakteristik der in der jüngeren Sprachphilosophie vertretenen Metaphertheorien findet sich in Eckard 2005. Einen Überblick zum Metapherngebrauch in der Philosophie verschafft Konersmann 2007.

Ein allein an Blumenberg orientiertes Verständnis des Projekts der Metaphorologie empfiehlt sich nicht, weil in diesem Fall auch eigenwillige Elemente einzubeziehen wären: Blumenbergs sprachtheoretisch unbefriedigende Ausführungen zur „absoluten Metapher“, seine fragwürdige Vernunft- und Wissenschaftskritik, seine großformatigen und spekulativen anthropologischen Thesen. Leider steht neben „Metaphorologie“ kein anderer geeigneter Ausdruck zur Verfügung. Hier sei darunter das Unternehmen verstanden, die Funktion metaphorischer Darstellungsformen für unser Selbst- und Weltverständnis zu untersuchen, also ihre epistemischen und kognitiven Leistungen oder, in der emphatischen Diktion der transzendentalen Hermeneutik, ihre welterschließende Kraft.

Die Metaphorologie berührt sich mit in der Wissenschaftstheorie geführten Debatten über die epistemische oder zumindest heuristische Funktion von Analogien und Modellen in der Wissenschaft, insbesondere beim Theorienwandel. Der Buchtitel *The Role of Analogy, Model and Metaphor in Science* (Leatherdale 1974) ist in dieser Hinsicht instruktiv. Dass die Trias *Metapher, Modell, Analogie* in der wissenschaftstheoretischen Diskussion verbreitet ist, zeigt allerdings auch an, dass dort in der Regel mit einem weiten, sprachtheoretisch unkonturierten Metaphernbegriff gearbeitet wird. Dies gilt schon für Mary Hesse (1966) sowie für die Arbeiten von Lakoff und Johnson. George Lakoff ist zwar Linguist, bearbeitet aber mit Mark Johnson das Metaphernthema im Schnittfeld zwischen Linguistik von Kognitionswissenschaft, wobei sie mit einem denkbar unscharfen Metaphernbegriff arbeiten und wenig Interesse für die sprachtheoretischen Feinheiten aufbringen. Zum Begriff der Metapher erfährt man: „The essence of metaphor is understanding and experiencing one kind of thing in terms of another“ (Lakoff/Johnson 1980:5). Mit dieser Unterbestimmung der Metapher haben Lakoff und Johnson die weitere Debatte in den Kognitionswissenschaften nachteilig beeinflusst.

Es liegt auf der Hand, dass Metaphorologie nicht unabhängig von Metapherntheorie betrieben werden kann, genauer: dass eine metapherntheoretisch naive oder unterkomplexe Metaphorologie wenig aussichtsreich ist. Eine Untersuchung der kognitiven Funktionen metaphorischer Darstellungsformen, ihrer Leistungen für die Formierung unseres Selbst- und Weltverständnisses, ist auf eine Klärung der Fragen angewiesen, was Metaphern sind, wie sie semantisch funktionieren und wie metaphorische Rede sich zur wörtlichen verhält. Legt man eine mangelhafte oder oberflächliche Metapherntheorie zugrunde, wird man auch in der Metaphorologie kaum zu erhellenden Ergebnissen kommen.

Mit der gleichwohl nötigen Vergrößerung lassen sich fünf Metapherntheorien oder Theorienfamilien unterscheiden, davon drei, deren Schwächen größer sind als ihre Vorzüge, und zwei, bei denen es umgekehrt ist (vgl. Keil 1993:271–299). Zur ersten Gruppe zähle ich die *Vergleichstheorie*, die *Substitutionstheorie* und die *Filtertheorie*, zur zweiten die *Interaktionstheorie* und die *pragmatische Theorie der indirekten Mitteilung*.

### 3 *Anthropomorphe, physiomorphe, soziomorphe und technomorphe Metaphern*

Metaphern lassen sich nach verschiedenen Gesichtspunkten klassifizieren. Eine nützliche Unterscheidung ist die zwischen lexikalisierten, konventionellen und innovativen Metaphern. Sie betrifft den Grad der Konventionalität einer Metapher. *Lexikalisierte* Metaphern (Katachrese, tote Metapher) wie „Flaschenhals“, „Flussbett“, „Hochstapler“, „großspurig“, „kaltblütig“ und „niedergeschlagen sein“ sind in den Wortschatz der Sprache eingegangen und füllen Lücken im Lexikon. Sprecher verwenden sie in der Regel ohne Bewusstsein ihres metaphorischen Ursprungs. Als Jean Paul die ganze Sprache „ein Wörterbuch erblasseter Metaphern“ nannte, hatte er diesen Metapherotyp im Sinn. *Konventionelle* Metaphern wie „ein Buch verschlingen“, „hinkender Vergleich“ und „einen Computer füttern“ sind nicht als eigene Lexikoneinträge verankert, sondern als usuelle Kopplungen in der Sprachpraxis der Sprechergemeinschaft. Auch sie müssen von kompetenten Sprechern nicht eigens dekodiert werden. *Innovative* („poetische“, „emphatische“) Metaphern sind idiolektale Neubildungen oder -kopplungen. Sie kommen nicht nur in der Poesie vor, sondern auch in anderen Bereichen. Manche innovative Metaphern finden schnelle Verbreitung und werden zu konventionellen, andere verblassen nie.

Die Unterscheidung zwischen anthropomorphen, physiomorphen, soziomorphen und technomorphen Metaphern ist eine nach dem „Bildspendebereich“ (Harald Weinrich). Bei jeder Metapher ist zwischen dem *Bildspender*, von dem die Merkmale übertragen werden, und dem *Bildempfänger* zu unterscheiden, auf den sie übertragen werden. Verbreiteter sind die Termini „tenor“ und „vehicle“ (Richards) sowie „focus“ und „frame“ (Max Black), aber Weinrichs Begriffe sind anschaulicher und weniger missverständlich.

In physiomorphen Metaphern werden Merkmale aus der Sphäre der Natur auf etwas Nichtnatürliches übertragen, in technomorphen Metaphern werden Merkmale aus dem Bereich der technischen Gegenstände auf Nichttechnisches übertragen, in soziomorphen Metaphern Merkmale aus dem Bereich des sozialen Lebens auf Nichtsoziales.

Bei den anthropomorphen Metaphern können wir nicht ohne Erläuterung von der Sphäre des Menschen als Bildspender sprechen, denn der Mensch ist ja nicht nur ein vernunft-, sprach- und handlungsfähiges Wesen, sondern zugleich ein Teil der physischen Welt, ein Stück Natur. Auf Menschen können Prädikate wie „wiegt 80 Kilogramm“, „befindet sich im Treppenhaus“ oder „besteht größtenteils aus Kohlenstoff“ zutreffen. Wenn man diese Prädikate einem nichtmenschlichen Gegenstand zuspricht, wird man das keine anthropomorphe und auch keine metaphorische Beschreibung nennen. Auch die Nominaldefinition von „anthropomorph“ – „nach Gestalt des Menschen“ – ist deshalb keine große Hilfe. Wenn man

einen zu weiten Anthropomorphismusbegriff zugrunde legt, wird einem die Pointe des Anthropomorphismusvorwurfes entgehen.

Einen Naturprozess anthropomorph beschreiben heißt, ihn *metaphorisch* beschreiben, und das bedeutet, ihn so zu beschreiben, als ob er in der relevanten Hinsicht menschlich wäre. In einem interessanten Sinne anthropomorph sind Beschreibungen, in denen einem nichtmenschlichen Gegenstand ein Prädikat zugesprochen wird, das im wörtlichen Sinn allein auf Menschen zutrifft. Um die Sache etwas abzukürzen: Es gibt gute Gründe dafür, als systematischen Kern eines interessanten Begriffes von Anthropomorphismus die Zuschreibung des *Handelnkönnens* anzusehen. Handeln impliziert im Unterschied zum bloßen Stattfinden von Körperbewegungen Intentionalität, demnach wäre es genauer das Modell des absichtlichen, zielgerichteten Handelns in seiner Anwendung auf nichtmenschliche Gegenstände, das dem Anthropomorphismus zugrunde liegt. Anthropomorph beschriebene Naturdinge, Naturprozesse oder Artefakte werden als intentional Handelnde vorgestellt. Dass diese Eingrenzung plausibel und für die Naturalismuskritik hochrelevant ist, zeigen einschlägige Beispiele wie die Konzeptualisierung der Evolution als eines selektierenden Akteurs, die des Computers als eines mit Repräsentationen arbeitenden, informations- oder symbolverarbeitenden Mechanismus oder die des Gehirns als einer den sensorischen Input auswertenden und entscheidungsfähigen Instanz. Alle diese Zuschreibungen münden, wie unten näher ausgeführt wird, in die Unterstellung intentionaler Leistungen, und sei es im Modus des Als-ob.

Was das Verhältnis des Anthropomorphismus zu den anderen -morphismen betrifft, so fallen soziomorphe und technomorphe Beschreibungen physischer Prozesse klarerweise auf die anthropomorphe Seite. Technomorphismus und Soziomorphismus sind Varianten oder Spezifikationen des Anthropomorphismus.

Konventionelle Personifizierungen bzw. Agentivierungen von Naturprozessen bieten ein Bild von abgestuften Intentionalitätsgraden. Die Wendungen „der Sturm wütet“, „die Natur tut nichts umsonst“ und „die Sonne lacht“ konnotieren in stärkerem Maße Absichtlichkeit als „der Nebel steigt“ oder „der Mond wandert über den Himmel“. Dies liegt offenbar an den Verben. Je höher der implizierte Grad von Absichtlichkeit, desto eher ist ein Verb für menschliche Handlungen reserviert. Bei allen Abstufungen aber drücken alle Verben qua Verben eine *Tätigkeit* aus, die dem Subjekt des Satzes zugeschrieben wird. Insofern der Wind weht, *tut* er etwas. Die philosophische Kinderfrage, was der Wind denn tue, wenn er nicht weht (Erwachsenenantwort: Er schläft), erinnert an die Fragwürdigkeit der Anwendung des Täter-Aktions-Schemas auf Naturprozesse. Nietzsche sieht in der Subjekt-Verb-Struktur der indoeuropäischen Sprachen, die stets Akteure und Handlungen suggeriere, einen gleichsam grammatisch eingebauten Fundamentalanthropomorphismus am Werk. Unter den „großen Irrthümern“ der durch

die „älteste und längste Psychologie“ geprägten Metaphysik führt Nietzsche auf: „alles 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“. Damit habe der Mensch „den Willen, den Geist, das Ich, aus sich herausprojicirt“ (Nietzsche 1889:85).

Die Metaphysikkritik Nietzsches tritt also als Kritik anthropomorpher Projektionen auf. Zu dieser tief ansetzenden Anthropomorphismuskritik gibt es indes eine Gegenrechnung. Wenn man fragt, mit welchen Ausdrücken denn die Aktivitäten und Eigenschaften „des Willens, des Geistes oder des Ich“ im Alltagsdeutsch beschrieben werden, stößt man auf zahllose physiomorphe Wendungen: „ich begreife etwas“, „ich bin niedergeschlagen“, „ich zerbreche mir den Kopf“, „ich versteige mich zu einer These“, „etwas berührt mich“, „etwas macht einen Eindruck auf mich“, etc. In diesen konventionellen Redewendungen fungieren physische Prozesse, also Vorgänge in der Körperwelt, als Bildspender, während mentale Vorgänge die Bildempfänger sind. Einige dieser Physiomorphismen haben eine somatische Basis. Empfindungsfähige Wesen haben einen affizierbaren Körper, und was diesem geschieht, geschieht auch ihnen. Wie einem zumute ist, wenn man sich niedergeschlagen fühlt, muss man niemandem lange erklären, weil das körperliche Niedergeschlagenwerden in diesem Ausdruck noch präsent ist. Auch bei „Fingerspitzengefühl haben“ oder „ein dickes Fell haben“ liegt die somatische Basis noch an der sprachlichen Oberfläche. Steigt man tiefer in die Etymologie unserer Rede über Mentales ein, so wird man erst recht fündig. Ivor Armstrong Richards bemerkt: „We can find no word or description for any of the intellectual operations which if its history is known, is not seen to have been taken, by metaphor, from a description of some physical happening“ (Richards 1936:91). Und Lakoff und Johnson, kurz und bündig: „We typically conceptualize the nonphysical *in terms of* the physical“ (Lakoff/Johnson 1980:59).

#### **4 Das anthropomorph-physiomorphe Paradox und die Interaktionstheorie der Metapher**

Es drängt sich die Frage auf, wie anthropomorphe Redeweisen über die Natur und physiomorphe Redeweisen über mentale Vorgänge sich zueinander verhalten. Auf den ersten Blick stehen die anthropomorphe Naturinterpretation und die physiomorphe Selbstinterpretation des Menschen in einem paradoxen Verhältnis. Auf den zweiten Blick wird die Lage noch unübersichtlicher, denn es gibt komplexe Prädikationsstrukturen, in denen ein Wechselspiel von Projektionen und Rückprojektionen vorliegt. Sprachlich kann dieses Wechselspiel ganz unauffällig realisiert sein. Eine Stelle bei Homer, auf die Bruno Snell aufmerksam gemacht hat, bietet ein Beispiel: „[Die Krieger] hielten aus [...] wie ein Fels im Meer, der Wind und Wellen

zum Trotz verharrt“. Nun „trotzen“ Felsen aber nicht im wörtlichen Sinne. Snell kommentiert:

Daß der Fels ein menschliches Verhalten deutlich macht [...], beruht darauf, daß dieser tote Gegenstand anthropomorph gesehen wird: das unbewegliche Stehen der Klippe in der Brandung wird gedeutet als Ausharren, so wie der Mensch ausharrt in einer bedrohten Situation. Der Gegenstand wird also tauglich, im Gleichnis etwas zu veranschaulichen, dadurch, daß in diesen Gegenstand das hinein-gesehen wird, was er dann seinerseits illustriert. (Snell 1946:268f)

Die wechselseitige Durchdringung von anthropomorphen und physiomorphen Elementen in unserem Selbst- und Naturverständnis ist verschiedentlich bemerkt, aber sprach- und metapherntheoretisch nur selten genauer beschrieben worden. Auch ich kann das hier nicht leisten, weil in diesem Beitrag nicht die Metapherntheorie, sondern die Rolle der Metaphorologie für die Naturalismuskritik im Mittelpunkt steht. Allgemein sind die meisten Metapherntheorien nicht zur Erklärung von Rückprojektionseffekten in der Lage, weil sie einen einsinnigen Mechanismus der Merkmalsübertragung annehmen. Paradigmatisch dafür ist Karl Bühlers Theorie der „selektiven Wirkung der Sphärendeckung“, der zufolge die nicht passenden Merkmale des Bildspenders und des Bildempfängers weggefiltert werden und nicht zur Bedeutungskonstitution der Metapher beitragen (vgl. Bühler 1934:342–356). Allein die *Interaktionstheorie* der Metapher, die auf Richards zurückgeht und von Max Black weiterentwickelt wurde, rechnet mit semantischen Rückwirkungen des Bildempfängers auf den Bildspender. Den Menschen einen Wolf zu nennen rückt nicht nur den Menschen in ein besonderes Licht, sondern lässt auch „den Wolf dabei menschlicher als sonst erscheinen“ (Black 1954:75). Die Interaktionstheorie nimmt eine semantisch produktive Interaktion der beiden einander Kontext gebenden Begriffssphären an. Der Bildempfängersbereich und der weitere Kontext unterdrücken nicht bloß Merkmale des Bildspenders, sondern es kommt zu einer wechselseitigen Übertragung und Modifizierung von Merkmalen:

It is central to any interaction view of metaphor that the effect of the context is not wholly negative, but that it has an important positive role. [...] It is the virtue of good models or metaphors that they tend to exploit their novel contexts rather than be frustrated by them. (Leatherdale 1974:195)

Im Anschluss an Blacks Interaktionstheorie hat Mary Hesse ihre einflussreiche Auffassung der Metapher als einer „Neubeschreibung“ des Gegenstands entwickelt. Hesse unterscheidet zwischen einer „positiven“ und einer „negativen Analogie“, wobei die positive Analogie die Schnittmenge der Merkmale angibt, die Bildspender und Bildempfänger gemeinsam haben, also das *tertium comparationis* der klassischen Metapherntheorie. Negative Analogie sind die Restmengen der nicht übereinstimmenden Merkmale.<sup>4</sup> Zusätzlich gibt es nach Hesse einen Bereich der

<sup>4</sup> Lakoff und Johnson (1980:52ff) sprechen von „used“ und „unused parts“ der Metapher.

„neutralen Analogie“, von dem man bei der Einführung der Metapher noch nicht weiß, ob er zur Bedeutungskonstitution beiträgt oder nicht. Genau in dieser Unabsehbarkeit bestehe der heuristische Wert der metaphorischen Neubeschreibung. Wir wissen nicht vorab, wie weit sich die durch die Metapher behauptete Analogie erstreckt (vgl. Hesse 1966:162). Damit Metaphern und Modelle unser Verständnis dessen, wofür sie Metaphern und Modelle sein sollen, befördern können, müssen sie eine „richer substructure“ und dadurch einen „surplus content“ gegenüber der nichtmetaphorischen Beschreibung des Gegenstandes haben (vgl. Leatherdale 1974:60, 146). Michael Arbib spricht in diesem Sinne von „exploratory metaphors“, mit deren Hilfe „both the similarities and the differences“ erforscht werden können (Arbib 1972:4). In mancher Hinsicht mag die Welt wie ein Uhrwerk sein, in anderen Hinsichten nicht. In mancher Hinsicht gleicht die natürliche Selektion der bewussten Auslese anhand von Merkmalen, in anderen nicht. In mancher Hinsicht gleicht der menschliche Geist einem Computer, in anderen nicht.

## 5 Artefakte und Funktionen

Um die Rolle der Metaphorologie für die Naturalismuskritik genauer zu bestimmen, reicht die simple Gegenüberstellung von anthropomorpher und physiomorpher Metaphorik nicht aus. Nehmen wir die *technomorphe* Metaphorik hinzu, die in dieser Antithese nicht ohne weiteres zu verorten ist. Technik sei hier im ursprünglichen Sinn von *technê* verstanden, also als Kunstfertigkeit oder handwerkliche Kompetenz. Als Bereichsbezeichnung grenzt *technê* den Bereich des künstlich Hergestellten vom natürlich Vorfindlichen ab: Manche Dinge kommen erst durch Zutun des *homo faber* in die Welt, andere sind schon vorher da. Alle konstruktivistische Sophistik, die das Natürliche partout als Gemachtes entlarven will, kann nichts daran ändern, dass mit dieser Unterscheidung etwas getroffen wird.

Ein Produkt der Technik ist zunächst einmal ein *Artefakt* im Unterschied zu einem vorfindlichen Naturgegenstand. Artefakte stehen in der Mitte zwischen Menschen und Naturdingen. Einerseits sind sie etwas Menschengemachtes, andererseits sind sie durch ihre Materialität ein Stück Natur, und zwar ein Stück unbelebte Natur, dessen Verhalten unter Gesetze der Physik fällt. Ein technisches Artefakt im engeren Sinn ist ein Gegenstand, der mit einer bestimmten Absicht und zur Erfüllung eines bestimmten Zwecks konstruiert wurde. Da Artefakte keine Lebewesen sind, kommen ihnen im wörtlichen Sinn keine Erkenntnis-, Handlungs- oder Sprachfähigkeiten zu.

Der Mensch hat die Arbeitsweise seines Geistes stets im Hinblick auf die jeweils neuesten technischen Erfindungen interpretiert. Aristoteles ver-

wendete viele Handwerksmetaphern, die mechanistischen Anthropologien seit dem 17. Jahrhundert orientierten sich am Modell der mechanischen Räderuhr. Im 19. Jahrhundert kamen die bis heute verbreiteten Elektrometaphern hinzu („eine lange Leitung haben“, „abschalten“, „Kurzschluss“, „Kontakt suchen“), später dann Eisenbahnmetaphern. Seit der Mitte des 20. Jahrhunderts verbreitet sich die Computermetapher des Geistes, zunächst in der Begrifflichkeit der Kybernetik, später im Sinne der Software-Hardware-Analogie der computerfunktionalistischen Auffassung des Geist/Körper-Problems: der menschliche Körper entspreche der Hardware, der Geist dem Programm, welches auf einer beliebigen Hardware implementiert sein kann. Die Computermetapher des Geistes und die älteren Maschinenmetaphern sind für unseren Zusammenhang aufschlussreich, weil einerseits Naturalisierungsansprüche mit ihnen verbunden werden, andererseits aber ein technisches Artefakt alles andere als ein bloßes Naturding ist. Der allein den Gesetzen der Mechanik oder des Elektromagnetismus folgende Mechanismus ist ja zugleich ein zu einem bestimmten Zweck konstruierter und verwendeter Apparat. Wird er zur Unterstützung oder zum Ersatz einer menschlichen Leistung eingesetzt, so wird seine Funktionsweise typischerweise in Analogie zu derjenigen Leistung beschrieben, die er ersetzt. Der Mensch baut sich ein Gerät, damit er nicht mehr rechnen muss, und nennt es einen Rechner. Er kann schlecht Französisch und kauft sich einen elektronischen Übersetzer. Der Geheimdienst hat zu wenig Personal, um tausende Stunden Tonbänder abzuhören und setzt ein Spracherkennungssystem ein. Diese Apparate werden nicht über ihr Material oder ihre Struktur bestimmt, sondern über ihre Funktion. Die Leistungen, von denen der Mensch sich durch den Apparat entlasten will, werden dem Apparat selbst zugeschrieben.

Die Neuanwendung der Kognitions- und Handlungsverben, also ihre Übertragung von der menschlichen Tätigkeit auf den technischen Prozess, ist zunächst einmal nicht zu beanstanden. Wir alle reden ständig so. In den meisten Kontexten wäre jede andere, nichtfunktionale Rede über die fraglichen Artefakte umständlich, irrelevant und ein Verstoß gegen Gricesche Konversationsmaximen. Anders verhält es sich, wenn weitreichende philosophische Thesen mit dieser Übertragung verbunden werden, insbesondere Naturalisierungsansprüche. Die Annahme, funktionale oder intentionale Charakterisierung der Leistungen von Artefakten sprächen für eine naturalistische Auffassung des Geistes, ist naiv. Aus dem Umstand, dass eine menschliche Leistung wie beispielsweise das Schachspielen sich funktional äquivalent (aber was heißt das genau?!) durch einen technischen Apparat erbringen lässt, folgt nichts darüber, ob Phänomene wie Geist, Sprache, Vernunft, Bewusstsein oder Handlungsfähigkeit sich vollständig mit naturwissenschaftlichen Begriffen und Methoden erklären lassen.

## 6 Das Programm der Naturalisierung des intentionalen Idioms

In der neueren Philosophie des Geistes hat man es nicht mehr mit naiven Naturalisten zu tun, sondern mit aufgeklärten oder halbaufgeklärten. Diese Naturalisten glauben nicht mehr, ihr theoretisches Reduktionsziel schon dadurch erreicht zu haben, dass sie den fraglichen Gegenstand als zur Körperwelt gehörig erwiesen haben oder dass sie zu einer bestimmten kognitiven Leistung einen im Output hinreichend ähnlichen Algorithmus angeben können. Mit dem in der analytischen Philosophie des Geistes verfolgten Projekt einer „Naturalisierung der Intentionalität“ ist eine *begriffliche* Reduktion des Mentalen auf Nichtmentales gemeint. Es geht dort nicht darum, den Menschen mit all seinen geistigen Fähigkeiten als ein Stück Natur zu erweisen. Im Projekt der Naturalisierung der Intentionalität ist Naturalisierung eine semantische Reduktionsbeziehung, keine ontologische.

Was damit gemeint ist, macht eine programmatische Formulierung von Jerry Fodor deutlich: Eine naturalistische Erklärung oder Analyse eines intentionalen Phänomens ist nach Fodor eine solche, die in nichtmentalen, nichtsemantischen und nichtteleologischen Begriffen notwendige und hinreichende Bedingungen für das Vorliegen des intentionalen Phänomens angibt (vgl. Fodor 1987:98, 126).

Gesucht sind also nichtintentional formulierte Bedingungen für das Vorliegen eines Phänomens, das man ohne Kenntnis solcher Bedingungen mit intentionalen Ausdrücken beschreibt. Fodors Konjunktion *nichtmental*, *nichtsemantisch* und *nichtteleologisch* hat eine lange Vorgeschichte in der Philosophie des Geistes. Erklärte Naturalisten warten oft mit Charakterisierungen intentionaler Phänomene auf, die nicht unmittelbar Mentales zum Gegenstand haben, die aber intentionale Präsuppositionen besitzen, ihren Sinn also daraus beziehen, dass an anderer Stelle noch intentionale Phänomene unanalysiert geblieben sind.<sup>5</sup> Ein Indiz für versteckte intentionale Voraussetzungen ist eben die Verwendung semantischer und teleologischer Begriffe, deren Verwendung Fodor deshalb zusätzlich verbietet.

Das Verbot semantischer Ausdrücke besagt, dass Begriffe wie *meinen*, *bedeuten*, *bezeichnen* oder *repräsentieren* nicht unanalysiert vorkommen dürfen. Die enge Verwandtschaft von semantischem und mentalistischem

<sup>5</sup> „From the fact that a statement is not explicitly about anything mental it does not follow that none of its presuppositions make any reference to our cognitive interests, our way of regarding different contexts, or our intentional powers“ (Putnam 1992:57). Man kann auf das Problem der intentionalen Präsuppositionen reagieren, indem man den Begriff des Intentionalen von vornherein weiter fasst: „Say that a property is intentional if and only if either it is a propositional-attitude property – for example, the property of believing that such and such – or its instantiation presupposes instantiation of propositional-attitude properties“ (Baker 1995:193).

Idiom hat vor allem Quine herausgestellt. Sie beruht darauf, dass propositionale Einstellungen, als Paradigmen des Mentalen, semantische Identitätsbedingungen haben: Zwei Sprechern schreiben wir zum Beispiel dann dieselbe Überzeugung zu, wenn die sprachlichen Ausdrücke ihrer Überzeugungen ineinander übersetzbar sind.

Das Verbot teleologischer Ausdrücke besagt, dass *Ziele*, *Zwecke* und *Funktionen* nicht unanalysiert vorkommen dürfen. Das Teleologieverbot ist ungleich umstrittener als das Semantikverbot, was daran ersichtlich ist, dass offen oder verdeckt teleologische Charakterisierungen häufig zur angeblich naturalistischen Einführung semantischer Begriffe benutzt werden, so in „teleofunktionalistischen“ und „biosemantischen“ Analysen der sprachlichen Bedeutung. Das Teleologieverbot ist auch deshalb umstrittener als das Semantikverbot, weil die Biologie als eine respektable Naturwissenschaft mit funktionalen Begriffen arbeitet. Galt noch Kant die Kategorie des Zwecks als „Fremdling in der Naturwissenschaft“, sehen viele Naturalisten mittlerweile sowohl funktionale als auch teleologische Begriffe als unproblematisch an. Sie verweisen darauf, dass die Biologie „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). Der Hinweis darauf, dass die Biologie nun einmal teleologische Begriffe verwendet und auch nicht auf sie verzichten kann, reicht aber nicht aus, um diese Begriffe naturalistisch akzeptabel zu machen. Wer sich als Naturalist auf die Biologie beruft, muss entweder zeigen, dass funktionale Charakterisierungen nichts Teleologisches an sich haben, oder dass sie trotz ihrer teleologischen Struktur keine zwecksetzende Instanz präsupponieren. Er muss die Kantische Formel „Zweckmäßigkeit ohne Zweck“ erläutern, genauer: er muss bei Strafe des Anthropomorphismus zeigen, wie denn in der Natur Zweckmäßigkeit ohne zwecksetzende oder zweckbeurteilende Instanz möglich ist. Die schlichteren Gemüter unter den Naturalisten behaupten, Darwin habe dies längst gezeigt, indem er den Mechanismus der Evolution gefunden und Zweckmäßigkeit auf Kausalität zurückgeführt habe. Davon kann aber keine Rede sein. Mutation und Selektion sind keine kausalen Mechanismen, und die Aufklärung der Logik evolutionärer und anderer funktionaler Erklärungen ist eine diffizile Aufgabe, an der sich schon viele Wissenschaftstheoretiker die Zähne ausgebissen haben (vgl. dazu Keil 2007).

Oben habe ich vorgeschlagen, als systematischen Kern eines interessanten Begriffes von Anthropomorphismus die Zuschreibung des Handelnskönnens anzusehen. Ein „interessanter“ Begriff des Anthropomorphismus ist in unserem Zusammenhang einer, der für die Auseinandersetzung mit naturalistischen Theorien humanspezifischer Phänomene einschlägig ist, genauer: durch den sich diejenigen Metaphern auszeichnen lassen, die in naturalistischen Theorien nicht vorkommen dürfen. Theorien, die an irgendeiner Stelle auf unanalysierte anthropomorphe Metaphern oder Prä-

suppositionen zurückgreifen, erfüllen ihren selbstgesteckten Anspruch nicht, weil sie ihre Erklärungskraft aus einer ihnen verbotenen Quelle borgen.

Ausgangspunkt dieser Form der Naturalismuskritik ist der unkontroverse Befund, dass Menschen einander bestimmte Eigenschaften, Fähigkeiten und Leistungen im wörtlichen Sinn zuschreiben. Dass es wirklich das Vermögen des intentionalen, zielgerichteten Handelns ist, das als Zentrum oder Basis dieser Zuschreibungen fungiert, bedürfte einer eingehenden Begründung. Andere Philosophen sehen gegenüber der Handlungsfähigkeit die Zuschreibung mentaler Zustände (Wünsche, Überzeugungen, Absichten) als basal an. Manche sprechen von einer holistisch verfassten Zuschreibungspraxis und vom „Zirkel der Intentionalität“. Gemeint ist die These, dass Zuschreibungen intentionaler Zustände und Fähigkeiten einen geschlossenen Verweisungszusammenhang bilden, in dem jede intentionale Zuschreibung durch eine Reihe von anderen gerechtfertigt wird. Dass dieser Zirkel der Intentionalität existiert, ist die direkte Antithese zum Programm der Naturalisierung der Intentionalität. Wären hinreichende nichtintentional formulierte Bedingungen für irgendein intentionales Phänomen aufgefunden, so wäre der intentionale Zirkel durchbrochen, und der Naturalist könnte auf den Domino-Effekt rechnen.<sup>6</sup>

## 7 Das Homunkulus-Problem

In computationalen Theorien kognitiver Phänomene werden technischen Systemen und subpersonalen Instanzen vielfältige und voraussetzungsreiche Fähigkeiten zugeschrieben. Thermostaten, Computer, Roboter, Gehirne oder Neuronenverbände können angeblich Informationen gewinnen und verarbeiten, Berechnungen anstellen, Symbole manipulieren, Entscheidungen fällen, Regeln befolgen und sie auf interne Repräsentationen anwenden.

Diese Fähigkeiten und Leistungen sind nicht offen intentional, wohl aber nach Auffassung vieler Kritiker *intentionalitätspräsupponierend*. Der Einwand lautet, dass sie ihren Sinn daraus beziehen, dass in ihrer Umgebung intentionale Phänomene unanalysiert geblieben sind.<sup>7</sup> Zu den Kritikern des Symbolverarbeitungsansatzes in den Kognitionswissenschaften gehört John Searle. Er spießt Fragen auf wie diese: „Wie errechnet das visuelle System aus Schattierungen Gegenstandskonturen?“ Oder: „Welche Symbolmanipulation 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?“ Searle behauptet (vgl. 1990:28ff), dass diese Fragen ebenso sinnvoll seien wie die Frage: „Wie be-

<sup>6</sup> „Given any [...] suitably naturalistic break of the intentional circle, it would be reasonable to claim that the main *philosophical* problem about intentionality had been solved“ (Fodor 1990:52).

<sup>7</sup> Die folgende Darstellung beruht auf Keil (2003).

rechnet ein Nagel aus dem Schlag des Hammers die Distanz, die er im Holz zurücklegen muss?“ Doch während die letztere Frage offenkundig absurd ist, füllen Fragen der ersten Art kognitionswissenschaftliche Standardwerke. Nach Searle stellen weder Nägel noch Gehirne im Wortsinne Berechnungen an noch manipulieren sie mit Symbolen, beides täten allein Menschen.

Diese Behauptung ist umstritten und wirft die Frage auf, warum eigentlich bestimmte Prädikate für die Beschreibung der Tätigkeiten von Personen reserviert sein sollten. Nach Searle beruhen die zitierten Fragen auf der *versteckten* Annahme einer personalen oder personähnlichen Instanz mit intentionalen Zuständen, kurz: auf der Annahme eines Homunkulus. Als „Homunkulus“ bezeichnet man in der Philosophie des Geistes eine postulierte menschenähnliche Instanz, die ausdrücklich oder unausdrücklich zur Erklärung der Arbeitsweise des menschlichen Geistes herangezogen wird. Ein „Homunkulus-Fehlschluss“ entsteht nach Anthony Kenny (1971:155) durch „the reckless application of human-being predicates to insufficiently humanlike objects“. Aber was ist ein „nicht hinreichend menschenähnlicher“ Gegenstand, und woran bemisst sich das? Wittgenstein hat einmal dekretiert: „Wir sagen nur vom Menschen, und was ihm ähnlich ist, es denke“ (1960:§ 360). Diesem Dekret wird aus kognitionswissenschaftlicher Sicht mit einigem Recht entgegengehalten, dass der Sinn der Frage, ob die Arbeitsweise von Maschinen oder Gehirnen mit mentalen Prädikaten beschrieben werden darf, verfehlt wird, wenn man die Antwort einfach unter Verweis auf den gewöhnlichen Sprachgebrauch stipuliert. Die computerfunktionalistische Auffassung des Mentalen zeichnet 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.

Der Homunkulismus-Kritiker kann allerdings die schwierige Frage nach dem originalen Anwendungsbereich des intentionalen Idioms zunächst dahingestellt sein lassen. Er muss seinen Einwand lediglich im Sinne des *Regressarguments* formulieren. Wenn der Naturalist den Anspruch erhoben hat, die kognitive Fähigkeit F 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 F zu identifizieren oder zu postulieren.

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 Glas-

körper 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. Diese Annahme ist eine Implikation des Bildbegriffs: Ein Bild ist etwas, was gesehen werden kann. Tatsächlich befindet sich aber in meinem Kopf niemand mit einem zweiten Paar Augen, der von hinten meine Retina betrachtet. Und auch ich selbst tue das nicht. Ich kann nur betrachten, was sich *vor* meinen Augen befindet.

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 im Linsenauge der Wirbeltiere steht seit Jahrhunderten in den Lehrbüchern. An der Darstellung des Strahlengangs ist nichts auszusetzen, wohl aber an der Rede vom Netzhautbild. Den empirischen Befund 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. Auch den Vertretern der neueren computationalen Wahrnehmungstheorien gelingt es aber vielfach nicht, sich vollständig von der Bildanalogie zu befreien. Aktuelle Wahrnehmungstheorien, die unter anderem mit der Erklärung von *Konstanzphänomenen* befasst sind, formulieren 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?

Um genauer zu sehen, welche Gestalt das Homunkulus-Problem in der „computational theory of vision“ annimmt, sei abschließend ein Blick auf die Pionierarbeit von David Marr geworfen. Sehen ist nach Marr (1982:3) „the process of discovering from images what is present in the world“. Das Rohbild, ein zweidimensionales Bestrahlungsmuster auf der Netzhaut, werde in einen „primal sketch“ umgewandelt, der aus einer Beschreibung

geometrischer Eigenschaften des zweidimensionalen Bildes besteht. Über einen weiteren Zwischenschritt werde dann der Output des Sehsystems erzeugt, nämlich 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, 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 hat seinerzeit die Vorstellung kritisiert, „daß es sich bei Augen, Nasen und Ohren um eine Art Auslandskorrespondenten handelt“, die uns mehr oder weniger zutreffende Nachrichten schicken (Ryle 1954:124).

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 soll. 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. Solche Wahrnehmungsurteile hat Marr aber nicht im Sinn. Es soll sich bei der Beschreibung um den Output des Wahrnehmungsprozesses selbst handeln, den Marr als algorithmischen Prozess der Extraktion räumlicher Information aus dem Netzhautbild auffasst. Tatsächlich ist der Output des Wahrnehmungsprozesses aber keine Beschreibung und keine Abbildung, sondern – eine Wahrnehmung.

Searle hat seine Kritik am Symbolverarbeitungsansatz in den Kognitionswissenschaften mit der gebotenen Radikalität formuliert. Während das Argument des Chinesischen Zimmers zeigen sollte, dass die Semantik nicht in der Syntax enthalten ist, argumentiert Searle zehn Jahre später: „Worse

yet, syntax is not intrinsic to physics“ (Searle 1990:26). Kein physischer Prozess ist intrinsischerweise ein Berechnungsvorgang, eine Regelbefolgung oder eine Symbolmanipulation, da physische Vorgänge als solche nicht syntaktisch miteinander verknüpft sind. Berechnung ist eine bestimmte Art von Symbolmanipulation, und „symbol‘ and ‚same symbol‘ are not defined in terms of physical features“ (ebd.:35).

## 8 *Naturalismuskritik und Metaphorologie*

Der Umstand, dass angeblich naturalistische Erklärungen eines menschlichen Vermögens oft an irgendeiner Stelle auf Homunkuli, auf unanalysierte intentionale Begriffe oder auf andere unerläuterte anthropomorphe Anteile zurückgreifen, ist kein Apriori-Argument gegen Naturalisierungsprogramme, aber er verschiebt die Beweislast: Der Naturalist hat zu erklären, dass und warum die Verwendung der fraglichen Begriffe unbedenklich ist. Seine übliche Verteidigungslinie ist wohlbekannt: Die Verwendung des inkriminierten Ausdrucks sei eine bloße *façon de parler*, eine harmlose Metapher. Freilich glaube er nicht, dass beispielsweise Thermostaten im Wortsinne etwas anstreben können, aber man möge ihn doch bitte karitativ interpretieren und ihm nicht nach Philosophenart das Wort im Munde umdrehen.

Recht hat der Naturalist darin, dass eine bestimmte Form von Naturalismuskritik zu kurz greift, nämlich das bloße Aufspießen von Wörtern, die in naturalistischen Theorien und Erklärungen nicht vorkommen dürfen. Die natürliche Sprache hat offenbar kein Vokabular ausgebildet, das von Anfang an für mentale Operationen oder für absichtliche menschliche Handlungen reserviert war. Spätestens wenn wir die Etymologie einbeziehen, finden wir das reine mentalistische Idiom nicht, sondern zahllose Wendungen, „[which] have been taken, by metaphor, from a description of some physical happening“ (Richards 1936:91).

Auch aus dem bloßen Vorkommen einer Metapher lässt sich nicht schließen, dass eine Erklärung erschlichen wurde. Es gibt kein kanonisches oder gar mechanisches Testverfahren dafür, wie harmlos eine Metapher ist. Man sieht es einer metaphorischen Prädikation nicht an, worin jeweils der positive, der negative und der neutrale Analogiebereich besteht. Die sprachkritischen Instrumente, deren sich eine metaphorologisch verfahrenende Naturalismuskritik bedienen muss, erfordern eine sorgfältige Handhabung. Es braucht semantische und pragmatische Präsuppositionsanalysen und die linguistisch angeleitete Diagnose von Kategorienfehlern. Wo es sich um Metaphern handelt, ist metaphorentheoretische Expertise erforderlich. Es geht nicht darum, welche Ausdrücke vorkommen oder nicht vorkommen dürfen. Es geht darum, festzustellen, wie sie jeweils explanatorisch eingesetzt werden und welche der mit ihnen verbundenen Implikationen stillschweigend in die Erklärung eingehen.

**Literaturverzeichnis**

- Arbib MA (1972) *The Metaphorical Brain*. New York
- Baker LA (1995) *Explaining Attitudes*. Cambridge, MA
- Bennett MR, Hacker PMS (2003) *Philosophical Foundations of Neuroscience*. Oxford
- Black M (1954) *Die Metapher*. Zit. nach: Haverkamp A (Hrsg) (1983) *Theorie der Metapher*. Darmstadt, 55–79
- Blumenberg H (1960) Paradigmen zu einer Metaphorologie. In: *Archiv für Begriffsgeschichte* 6:7–142
- Blumenberg H (1971) Beobachtungen an Metaphern. In: *Archiv für Begriffsgeschichte* 15:161–214
- Blumenberg H (1979) Ausblick auf eine Theorie der Unbegrifflichkeit. In: ders., *Schiffbruch mit Zuschauer*. Frankfurt/M, 75–93
- Boolos GS, Jeffrey RC (1989) *Computability and Logic*. Cambridge MA
- Bühler K (1934) *Sprachtheorie*. Jena
- Eckard R (2005) *Metaphertheorien. Typologie – Darstellung – Bibliographie*. Berlin/New York
- Fodor JA (1987) *Psychosemantics*. Cambridge, MA
- Fodor JA (1990) *A Theory of Content and Other Essays*. Cambridge, MA
- Hesse M (1966) *Models and Analogies in Science*. Notre Dame, Indiana
- Keil G (1998) Was Roboter nicht können. Die Roboterantwort als knapp mißlungene Verteidigung der starken KI-These. In: Engel A, Gold P (Hrsg) *Der Mensch in der Perspektive der Kognitionswissenschaften*. Frankfurt/M, 98–131
- Keil G (2000) *Kritik des Naturalismus*. Berlin/New York
- Keil G (2003) Homunkulismus in den Kognitionswissenschaften. In: Köhler WR, Mutschler HD (Hrsg) *Ist der Geist berechenbar?* Darmstadt, 77–112
- Keil G (2007) *Biologische Funktionen und das Teleologieproblem*. In: Honnefelder L, Schmidt MC (Hrsg) *Naturalismus als Paradigma*. Berlin, 76–85
- Keil G (2008) *Naturalism*. In: Moran D (ed) *The Routledge Companion to Twentieth-Century Philosophy*. London, 254–307
- Kenny A (1971) *The Homunculus Fallacy*. In: Hyman J (ed) *Investigating Psychology*. London/New York 1991, 155–165
- Konersmann R (Hrsg) (2007) *Wörterbuch der philosophischen Metaphern*. Darmstadt
- Lakoff G, Johnson M (1980) *Metaphors We Live By*. Chicago/London
- Leatherdale WH (1974) *The Role of Analogy, Model and Metaphor in Science*. Amsterdam/Oxford
- Lycan W (1991) *Homuncular Functionalism Meets PDP*. In: Ramsey W et al. (eds) *Philosophy and Connectionist Theory*. Hillsdale, 259–286
- Marr D (1982) *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco
- Nietzsche F (1889) *Götzendämmerung*, Werke VI/3, hrsg. von Colli/Montinari. Berlin 1969, 49–157
- Putnam H (1992) *Renewing Philosophy*. Cambridge, MA/London
- Richards IA (1936) *The Philosophy of Rhetoric*. New York 1965
- Roth G (2004) *Worüber dürfen Hirnforscher reden – und in welcher Weise?* In: Geyer C (Hrsg) *Hirnforschung und Willensfreiheit*. Frankfurt/M, 66–85
- Ryle G (1954) *Die Wahrnehmung*. In: ders., *Begriffskonflikte*. Göttingen 1970, 117–138
- Searle JR (1990) *Is the Brain a Digital Computer?* Proceedings and Addresses of the American Philosophical Association 64/3:21–37
- Snell B (1946) *Die Entdeckung des Geistes*. Hamburg <sup>3</sup>1955
- Weinrich H (1963) *Semantik der kühnen Metapher*. Zit. nach: Haverkamp A (Hrsg) *Die Metapher*. Darmstadt 1983, 316–339
- Wittgenstein L (1960) *Philosophische Untersuchungen*. Frankfurt/M