

### 1. Was ist ein Homunkulus-Fehlschluß?

Aristoteles erörtert in *De Anima* die verschiedenen Vermögen und Tätigkeiten der Seele. An einer Stelle beschleichen ihn indes sprachkritische Skrupel:

[S]o ist die Aussage, daß die Seele sich erzürne, ähnlich der, wie wenn man sagte, die Seele webe ein Tuch oder baue ein Haus; denn es ist vielleicht besser, nicht zu sagen, die Seele habe Mitleid, oder lerne, oder denke, sondern der Mensch mittels der Seele [...]. (*De Anima*, 408 b)

Aristoteles weist hier auf eine Denkfigur hin, die in der neueren Philosophie des Geistes einen eigenen Namen hat; man nennt sie dort den "Homunkulus-Fehlschluß".

"Homunkulus" ist eine Diminutivform von lat. "homo" und bedeutet also wörtlich "kleiner Mensch". Die primäre Verwendung des Ausdrucks ist außerphilosophisch. Man versteht darunter einen künstlich hergestellten Menschen, gemäß der im 13. Jahrhundert aufgekommenen alchimistischen Vorstellung, daß man ein Menschenwesen im Labor aus Schwefel, Quecksilber und Salz erzeugen könne. Wenn in der Philosophie vom Homunkulus-Fehlschluß die Rede ist, wird der Ausdruck in einem anderen Sinne verwendet. Es geht dort nicht um tatsächliche kleine Menschen, sondern um ausgedachte, genauer: um *postulierte menschenähnliche Instanzen, die ausdrücklich oder unausdrücklich zur Erklärung der Arbeitsweise des menschlichen Geistes herangezogen werden*.

Den Ausdruck "homunculus fallacy" hat der Oxfordener Philosoph 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)

Die Anwendung von Prädikaten, die nur auf Menschen zutreffen, auf menschenunähnliche Dinge – das ist eine sehr weite Charakterisierung des Fehlschlusses. "Human-being predicates" könnten 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 (vgl. z.B. Schumacher 1998, 41).

Es stellt sich die Frage, inwiefern man hier von einem "Fehlschluß" sprechen kann. Ein Fehlschluß im philosophischen Sprachgebrauch ist die Anwendung eines ungültigen Schlußschemas. In der Regel ist gemeint, daß von wahren Prämissen zu einer falschen Konklusion übergegangen wird. Notwendig ist dies nicht; ein Fehlschluß kann auch zu einer wahren Konklusion führen. 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. Dieser Fehlschluß wäre zu offensichtlich, als daß man ihn offenen Auges beginge.

Fehler, die man nicht offenen Auges begehen kann, muß man geschlossenen Auges begehen. Auch die anderen Fehlschlüsse, die in den Logik- und Rhetorik-Handbüchern beschrieben sind, begeht man ja in der Regel nicht absichtlich, sondern sie unterlaufen einem. Bei einer Person, die sowohl rational als auch guten Willens ist, kann man sich das absichtliche Ziehen eines Fehlschlusses nur so vorstellen, daß die Person der Auffassung ist, daß der Schluß, den alle Welt für einen Fehlschluß hält, in Wirklichkeit keiner ist. Das kann vorkommen, aber dieser Fall ist selten. Deshalb spricht die Tatsache, daß der von Kenny beschriebene Schluß selten ausdrücklich gezogen wird, noch nicht dagegen, ihn einen Fehlschluß zu nennen. Argumentationstheoretisch betrachtet, hat die Diagnose eines Fehlschlusses fast immer ein Moment von Entlarvung. Es wird ein Fehler aufgedeckt, der nicht offen zutage lag. So ist es auch beim Homunkulus-Fehlschluß: Das, was der andere gesagt hat, wird vom Kritiker auf eine solche Weise reformuliert, daß es als Instanz des fehlerhaften Schlußschemas erscheint.

Nach Kenny ist das Begehen des Homunkulus-Fehlschlusses gleichbedeutend mit dem Postulieren eines kleinen Menschen innerhalb eines Menschen. Man kann fragen, was hier "postulieren" heißen soll. Warum sollte jemand etwas postulieren, von dem er nicht glaubt, daß es so etwas gibt? "To explain human experience and behaviour", heißt es bei Kenny. Doch wie kann jemand glauben, daß etwas, was es nicht gibt, eine Erklärungsleistung erbringt? Ich schlage vor, daß auch die Behauptung, jemand postuliere einen Homunkulus, schon zur kritischen Rekonstruktion dessen gehört, was der Opponent tatsächlich gesagt hat. Die Diagnose eines Homunkulus-Fehlschlusses hätte demnach eine kompliziertere Struktur als es auf den ersten Blick aussieht. Ich schlage die folgenden Schritte vor:

- (1.) Der Opponent erklärt oder analysiert eine kognitive oder perzeptive Leistung einer Person als einen F-Prozeß innerhalb der Person.
- (2.) Das Prädikat "F" trifft wörtlich nur auf Verrichtungen von ganzen Personen zu.
- (3.) Um einen Träger für sein Prädikat zu haben und seine Erklärung zu rechtfertigen, müßte der Opponent also etwas hinreichend Personenähnliches innerhalb der Person annehmen: einen Homunkulus.
- (4.) Tatsächlich gibt es aber nichts hinreichend Personenähnliches innerhalb von Personen.
- (5.) Also hat das Prädikat keine Anwendung und die vermeintliche Erklärung ist nur eine Pseudo-Erklärung.

Das sind wohlgermerkt die Schritte der *Diagnose* eines Homunkulus-Fehlschlusses, nicht eine Rekonstruktion der Überlegungen beim Begehen des Fehlschlusses. Der Schlußprozess selbst kann wohl nur aus der psychologisierenden Beobachterperspektive aufgeklärt werden, denn wenn sich der Schließende selbst das fehlerhafte Schlußschema vor Augen hielte, würde er das Schließen vermutlich bleiben lassen. Man denke beispielsweise an Zirkelschlüsse: Sobald mir die versteckte Prämisse vor Augen geführt wird, wird es mir nicht mehr gelingen, von der Qualität meines Schlusses überzeugt zu sein. Es liegt demnach in der Natur der Sache, daß die Schritte eines Fehlschlusses aus der Perspektive des Schließenden schlecht zu rekonstruieren sind. Deshalb habe ich versucht anzugeben, in welchen Schritten die Diagnose erfolgen könnte. Der Kern des Vorwurfs wäre demnach nicht, daß der Opponent einen Homunkulus annimmt, sondern daß er einen annehmen *müßte*, um die Verwendung der gewählten Prädikate zu rechtfertigen. Und da weitgehend Einigkeit darüber besteht, daß es Homunkuli im Wortsinne nicht gibt, ist die

Verwendung der Prädikate nicht gerechtfertigt und kann nicht zur Erklärung dienen – so das Argument.

Die philosophische Bedeutung des Homunkulus-Fehlschlusses läßt sich nicht allein an der Argumentationsfigur ablesen, wiewohl diese für sich betrachtet interessant genug ist. Der Argumentationsfigur liegt ein philosophisches *Problem* zugrunde, nämlich, in erster Annäherung, die Frage, wie der menschliche Geist funktioniert. Etwas genauer lautet die Frage, ob sich die Arbeitsweise des menschlichen Geistes analysieren oder explanatorisch auf Vorgänge zurückführen läßt, die nicht selbst schon geistige Phänomene sind. ("Geistige Phänomene" seien hier im weiten philosophischen Sinne von "geistig" verstanden, der alle psychischen Vorgänge und auch das Phänomen der Wahrnehmung mit einschließt.) Vielleicht leuchtet diese Behauptung nicht unmittelbar ein, weshalb ich sie noch etwas erläutern möchte. Man kann sich fragen, was eigentlich die treibende Kraft hinter dem Fehlschluß ist, die allen Postulierungen personenähnlicher psychischer Instanzen gemeinsam ist. Kenny meint in der Tat, daß die Homunkulisten schließen: 'Was vom Menschen prädzierbar ist, ist auch von seinem Gehirn prädzierbar' (a.a.O., 164). Daß ein solcher Schluß vom Ganzen auf das Teil 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 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 die Suche nach einer nichtintentionalen Analyse oder Erklärung eines intentionalen Prozesses ist, also das, was man in der Philosophie des Geistes das Problem der Naturalisierung der Intentionalität nennt. Intentionale Fähigkeiten und Leistungen von Personen sollen sich auf natürliche Prozesse zurückführen lassen, die innerhalb von Personen stattfinden. Aus naturalistischer Sicht würde das Beharren auf der unanalysierbaren Rede von Wünschen, Überzeugungen und Absichten es unmöglich machen, Personen als Teil der einen natürlichen Welt zu betrachten, deren restliche Teile schließlich keine intentionalen Eigenschaften haben.<sup>1</sup> Auch der Umstand, daß der Homunkulus-Fehlschluß in der kognitiven Psychologie und der Philosophie des Geistes als "intentional fallacy" bezeichnet wird, spricht dafür, ihn auf das Problem der Naturalisierung der Intentionalität zu beziehen.

## 2. Instanzenmodelle der Psyche

Man kann zunächst festhalten, daß es eine große Menge von alltäglichen Redeweisen gibt, in denen die psychische Bühne als von mehreren *dramatis personae* bevölkert aufgefaßt wird: "Zwei Seelen wohnen, ach! in meiner Brust." – "Mein Herz hat entschieden." – "Was müssen meine Augen da sehen?" – "Mein Gewissen sagt mir . . . ". Hier werden jeweils einzelne kognitive oder perzeptive Leistungen eigenen Instanzen zugeschrieben, und zwar mit Prädikaten, mit denen sonst die Leistungen ganzer Personen beschrieben werden: *sehen, sagen, entscheiden*. Es scheint eine Art Alltagspsychologie des Homunkulismus zu geben, und man wird, jenseits aller Kritik am Homunkulus-Fehlschluß, etwas dazu sagen müssen, warum diese Aufspaltungen einer Person in mehrere Instanzen so plausibel sind, warum sie sich immer wieder aufdrängen, und ob sie nicht doch einen Grund in der Sache haben.

---

<sup>1</sup> Diese Vermutung über die treibende Kraft hinter dem Homunkulus-Fehlschluß habe ich an anderer Stelle (Keil 2003) genauer ausgeführt.

Nun kann man einwerfen, daß diese personifizierenden Redeweisen doch ganz harmlos seien. Brisant werde es erst, wo Erklärungsansprüche mit ihnen verbunden werden. Eben das geschieht nach Auffassung der Homunkulus-Kritiker. Das Homunkulus-Problem wird in verschiedenen Bereichen diskutiert, von denen ich im folgenden zwei auswähle. Zunächst werde ich kurz auf die *Metapsychologie* eingehen, nämlich auf Autoren, die Instanzenmodelle der psychischen Apparates entwerfen. Dann diskutiere ich etwas ausführlicher ein Problem aus der *Philosophie der Wahrnehmung*. Nach einem Exkurs über das Verhältnis zwischen philosophischer Sinnkritik und empirischer Wissenschaft werde ich ausdrückliche *Rechtfertigungen* homunkularer Redeweisen erörtern und dafür zu den Instanzenmodellen zurückkehren.

Klassische Instanzenmodelle der Psyche sind Platons Seelenwagengleichnis und Freuds Modell von Ich, Es und Überich. Platons Seelenwagen ist bekanntlich aus zwei Pferden und einem Wagenlenker zusammengesetzt. Das eine Pferd ist gut, edel und folgsam, das andere aber ist "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. Leider kann man das ungebärdige Pferd nicht ausspannen, denn der begehrende Seelenteil (*epithymētikon*) ist eine unersetzliche Antriebsquelle für den Seelenwagen. Das Metaphernfeld der Zähmung eines ungebärdigen Pferdes ist auch heute noch Teil des alltagspsychologischen Diskurses. Man denke an Ausdrücke wie *ungezügelt*, *Benehmen*, *sich im Zaum halten*, *sich Zügel anlegen*, *ein zügelloser Mensch*, *die Pferde gehen mit ihm durch*.

Freud hatte zunächst ein hydraulisches Modell der Psyche vertreten, 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 Zweinstanzenmodell von bewußtem und unbewußtem System, seit 1923 dann durch das Dreinstanzenmodell von Ich, Es und Überich. 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 beispielsweise als von den streitenden Parteien in einem Bürgerkrieg:

Das Ich ist durch den inneren Konflikt geschwächt, wir müssen ihm zur Hilfe kommen. Es ist wie in einem Bürgerkrieg, der durch den Beistand eines Bundesgenossen von aussen entschieden werden soll. Der analytische Arzt und das geschwächte Ich des Kranken sollen [...] eine Partei bilden gegen die Feinde, die Triebansprüche des Es und die Gewissensansprüche des Überichs. (Freud, GW XVII, 98)

Mehrinstanzenmodelle der Psyche stehen von Natur aus unter Homunkulusverdacht. Der Grund dafür ist, daß eine psychische Instanz die ihr zugeschriebenen Fähigkeiten kaum 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 oder sich ihnen widersetzen können sowie sich zu etwas entscheiden können. Es ist schwer zu sehen, wie sie über diese Fähigkeiten verfügen können sollte, ohne eine eigene Psyche zu haben. Man kann das,

etwas aufwendig, die *holistischen Zuschreibungsbedingungen mentaler Fähigkeiten* nennen. Man wollte die Instanz eigentlich nur mit bescheidenen Talenten ausstatten; um diese aber plausibel zu machen, muß man ihr noch weitere Fähigkeiten zugestehen.<sup>2</sup> Wenn aber eine subpersonale Instanz eine eigene Psyche haben müßte, um überhaupt den ihr zugedachten Beitrag leisten zu können, liegt ein Regreß vor, denn die Funktionsweise der Psyche sollte ja gerade erklärt werden. *Deshalb* sind Homunkuli nach allgemeiner Auffassung nicht erklärungskräftig: Sie haben eine eingebaute Vermehrungstendenz und führen zu einer Vervielfachung des ursprünglichen Analysandums, anstatt eine Analyse zu bieten. Sie sind "a notorious non-solution to the problems of mind", "merely postponing the central problem before us by positing unanalysed man-analogues as systematic elements in that which we are trying to analyse" (Dennett 1969, 87).

Homunkulusgefährdet erscheinen neben Platons und Freuds Instanzenmodellen auch Kants vermögenspsychologische Ausführungen zum *Willen*. Kant spricht bekanntlich von der "Tätigkeit" und von "Handlungen" des Willens. Er schreibt dem Willen Prädikate zu, von denen man meinen sollte, daß sie wörtlich nur der wollenden Person zukommen. Zugleich benutzt Kant diese Prädikate weiterhin für Personen. Diesen Umstand könnte man zum Anlaß einer wohlwollenden Interpretation nehmen, nämlich einer *metonymischen*: Der Wille stünde demnach in diesen Formulierungen metonymisch für die Person, die ihn jeweils hat. Kants personifizierende Rede über den Willen wohlwollend zu interpretieren schließt freilich aus, ihr eine Analyse- oder Erklärungsleistung zuzuerkennen.

### **3. Der Homunkulus in der Wahrnehmungstheorie: Das Problem des invertierten Netzhautbildes**

Die Instanzenmodelle der Psyche sind das natürliche Habitat des Homunkulus. Mein Hauptbeispiel soll aber die Wahrnehmungstheorie sein, genauer: Theorien des Sehens als eines Verarbeitens von Netzhautbildern. Für dieses Fallbeispiel ist eine wissenschaftsgeschichtliche Rückschau erforderlich.<sup>3</sup>

Bis zum Anfang des 17. Jahrhunderts war der genaue Strahlengang des Lichtes im menschlichen Auge unbekannt, weil das Brechungsgesetz der geometrischen Optik noch nicht aufgefunden war. Bekannt war indes das Prinzip der *Camera obscura* oder Lochkamera – bei den Arabern seit dem 11. Jahrhundert, in Mitteleuropa erst um 1500. Bei der Camera obscura fällt bekanntlich das von einem Gegenstand reflektierte Licht durch eine Lochblende in einen dunklen Kasten, der als Rückwand einen transparenten Schirm besitzt. Auf diesem Schirm kann von hinten ein auf dem Kopf stehendes Bild des Gegenstandes gesehen werden.

Johannes Kepler lieferte 1603 als erster eine korrekte Darstellung des Strahlengangs im menschlichen Auge, der etwas komplizierter ist als bei der Lochkamera, weil das Auge statt des Loches eine Linse aufweist. Der Effekt ist aber insofern derselbe, als auf der Retina ein verkleinertes, invertiertes Abbild des Gegenstandes entsteht. Die untenstehende Illustration stammt von Descartes, der in seiner *Dioptrique* einige Jahre nach Kepler ebenfalls eine an der geometrischen Optik orientierte Theorie des Sehens aufstellte.

---

<sup>2</sup> Zu diesem Argument vgl. Hacker 1991, 128 f. sowie, mit Blick auf Freud, Thalberg 1982, 259.

<sup>3</sup> Der folgende Überblick orientiert sich an Hyman 1989, 1-19.

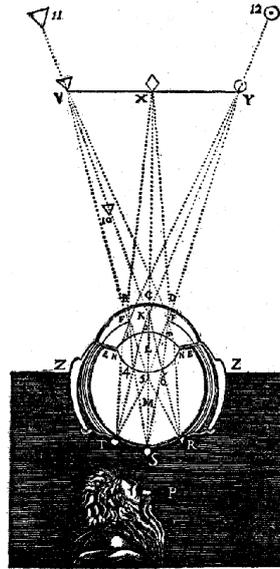


Abb. 1 Der Strahlengang im Linsenauge (nach Descartes, *Dioptrique*)

Kepler und Descartes beschrieben also das Auge als eine Camera obscura mit Linse. Aus dieser Einsicht in den Strahlengang des Lichts im Auge ergab sich nun ein Rätsel, das die Wahrnehmungstheorie auf Jahrhunderte beschäftigen sollte: *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?*

Kepler hat zunächst vermutet, es müsse irgendwo im Glaskörper des Auges eine zweite Linse geben, die uns die Welt wieder vom Kopf auf die Füße stellt. Er hat sich, wie er schreibt, 'lange gequält nachzuweisen, wo im Auge die Lichtstrahlen einander ein zweites Mal schneiden, bevor sie die Netzhaut erreichen' (vgl. Kepler 1604, 185). Als er den zweiten Schnittpunkt nicht fand, griff er zu einer extrem unplausiblen analogischen Erklärung, die hier nicht weiter von Interesse ist.<sup>4</sup> Descartes gab eine eigene Erklärung, die ebenfalls auf einer Analogie beruhte, der des Ertastens eines Gegenstands mit gekreuzten Spazierstöcken, welche 100 Jahre später von Berkeley kritisiert wurde. Immerhin hat Descartes endgültig mit der Hypothese einer zweiten Linse aufgeräumt, indem er ein echtes Auge mit freipräpariertem Glaskörper in ein Stück Karton eingesetzt hat, und tatsächlich konnte man, wenn man von hinten hindurchsah, ein auf dem Kopf stehendes Bild sehen (vgl. Descartes 1637, 115 f.). Descartes hat sich also mithilfe eines echten Auges eine Camera obscura gebaut. Die Anweisung dazu steht heute noch in manchen Lehrbüchern; man nimmt dafür wegen seiner schwach pigmentierten Häute gern das Auge eines Albino-Kaninchens.

Weder Kepler noch Descartes noch Berkeley haben gesehen, daß gar keine Antwort auf das Inversionsproblem gegeben werden muß. Sie haben nicht bemerkt, daß schon der *Problemstellung* ein Homunkulus-Fehlschluß zugrunde liegt. Die Frage, wo denn die Invertierung wieder aufgehoben wird, entsteht allein unter der Annahme, daß auf der Netzhaut etwas ist, was gesehen wird, eben ein Bild, das sogenannte Netzhautbild.<sup>5</sup> Nun ist aber das, was wir sehen, vor unseren Augen und

<sup>4</sup> Kurz gesagt: Sehen ist etwas Passives, mithin das genaue Gegenteil einer Aktivität, daher ist es kein Wunder, daß auch die Netzhautbilder gegenüber denjenigen Dingen, die sie erzeugen, verkehrt sind (vgl. ebd.).

<sup>5</sup> Das Netzhautbild ist nicht zu verwechseln mit den in Theorien mentaler Repräsentationen postulierten "inneren" oder "geistigen Bildern", die ebenfalls ein Homunkulus-Problem erzeugen (vgl. dazu Scholz 1995).

nicht auf unserem Augenhintergrund. Hinter unseren Augen sind keine weiteren Augen mehr und auch kein Menschlein mit Augen. Descartes hat dort tatsächlich einen Betrachter eingezeichnet (s. Abb. 1). Man ist versucht, den eingezeichneten Homunkulus für einen Scherz zu halten. Tatsächlich gibt aber die Zeichnung die Situation wieder, in der Descartes sich als Benutzer seiner selbstgebastelten unappetitlichen Camera obscura befindet. (Dies sieht man daran, daß an der Rückseite des Augapfels die Leder- und die Aderhaut wegpräpariert sind.) In dieser Situation befinden wir uns aber nicht, wenn wir mit unseren eigenen Augen sehen. Wir betrachten beim Sehen nicht zugleich mit einem zweiten Paar Augen unseren eigenen Augenhintergrund. Ebenso wenig tut dies unser Gehirn. Die Sehzellen absorbieren Lichtquanten, über eine Kette verstärkender chemischer Reaktionen entsteht ein elektrisches Signal, welches der Sehnerv ins Gehirn weiterleitet, doch nichts im Gehirn *sieht* das retinale "Bild". Das sogenannte Netzhautbild ist eine *physiologische Begleiterscheinung* des Sehens, und je nachdem, was mit "Netzhautbild" gemeint ist (s.u.), ist es auch eine Begleiterscheinung, ohne welche kein Sehen stattfinden würde, so wie das menschliche Auge nun einmal gebaut ist. Aber das kann man von anderen Bedingungen auch sagen: Die Lider müssen geöffnet sein, die Linse darf nicht eingetrübt, der Sehnerv nicht durchtrennt sein, *damit* wir sehen können.<sup>6</sup>

Kompliziert wird die Lage durch den Umstand, daß Descartes die mit dem Bildbegriff verbundene Gefahr durchaus gesehen hat. Er stellt fest, daß das retinale Bild den äußeren Gegenständen ähnlich ist, doch sei es nicht diese Ähnlichkeit, die unser Wahrnehmungserlebnis erzeugt; andernfalls müßte es ja, wie Descartes selbst sagt, weitere "Augen in unserem Gehirn geben" (1637, 130).<sup>7</sup> Statt dessen gibt er eine *mechanistische* Erklärung der weiteren Verarbeitung des Netzhautbildes. Es seien mechanische Reizungen, nämlich die Bewegungen kleiner Fädchen, aus denen das Netzhautbild materialiter besteht, die unmittelbar auf die Gehirnoberfläche einwirken. Dort öffnen sich dann Poren, durch die kleine Partikel fließen (die "esprits animaux") und in der Seele den Seheindruck erzeugen. Descartes versucht mithin, in der weiteren Beschreibung des Wahrnehmungsvorgangs den Bildcharakter des Netzhautbildes wieder loszuwerden. Er hat sich dabei der Suggestion des Bildbegriffs aber nur halb entziehen können. Dies wird angesichts eines

---

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* jemand sich vorstellt.

<sup>6</sup> Für eine entsprechende Kritik an Berkeleys Behandlung des Inversionsproblems vgl. Armstrong 1960, 45-52, bes. 51.

<sup>7</sup> Ein ähnlicher Hinweis findet sich bei Berkeley (vgl. 1709, § 116). Der Fall liegt bei ihm noch komplizierter als bei Descartes, denn Berkeley hat sich selbst bemüht, das Inversionsproblem als Scheinproblem zu entlarven, allerdings aus den falschen Gründen. Er stellt die Frage, in bezug auf was das Netzhautbild denn invertiert sein sollte. In gewissem Sinne sei alles am rechten Platz, denn schließlich seien die Füße einer projizierten Figur am Boden und der Kopf in der Luft (§ 114). Eine Verkehrung ergebe sich erst im Vergleich der beiden Sinnesmodalitäten des Sehens und des Tastens, und eben dieser sei nicht zulässig. Beide Sinne müßten separat betrachtet werden, denn ihre Objekte seien kategorial verschieden und stünden in keinerlei räumlicher Beziehung zueinander (§§ 111, 115). Diese extrem unplausiblen Ausführungen sind nur auf dem Hintergrund der phänomenalistisch-idealistischen Auffassung Berkeleys zu verstehen, daß "all visible things are equally in the mind, and take up no part of external space" (§ 111). Auch die Wörter "up" und "down" haben ihm zufolge nur jeweils innerhalb einer Sinnesmodalität eine Anwendung, bezeichnen also hinsichtlich des Sehens und des Tastsinns Verschiedenes (§ 97 f.). Aus nichtphänomenalistischer Sicht bezeichnen die Wörter "oben" und "unten" indes die Positionen realer Gegenstände, und es spricht nichts dafür, daß sie systematisch mehrdeutig sind.

Hilfsargumentes klar, welches Descartes für seine Annahme beibringt, daß der Ort der Interaktion zwischen Körper und Seele die Zirbeldrüse sei. Das Argument lautet:

Damit wir also nur einen einzigen und einfachen Gedanken von der gleichen Sache und zur gleichen Zeit haben, ist es notwendig, daß es eine Stelle gibt, wo die zwei Bilder, die von den beiden Augen kommen [...], sich zu einem verbinden können, bevor sie zur Seele gelangen, damit sie dieser nicht zwei anstatt einem Bild darbieten. (Descartes 1649, 55 [§ 32]).

Zu diesem Zweck sei die Zirbeldrüse als nichtpaariges Organ bestens geeignet. Sie kann die beiden Bilder zu *einem* Seheindruck vereinen: "Dadurch bilden die beiden Bilder im Hirn nur ein einziges auf der Drüse ab, das unmittelbar auf die Seele einwirkt und sie [den Gegenstand] sehen läßt" (ebd., 61 [§ 35]). Das Interessanteste an diesem bizarren Argument ist der Umstand, daß Descartes eine Erklärung dafür, daß die Seele oder der Geist aufgrund zweier Bilder einen Gegenstand wahrnimmt, überhaupt für nötig hielt. Er hält nach wie vor an der Vorstellung fest, daß der Seele *etwas zum Wahrnehmen dargeboten* wird. Die Seele wird als Subjekt einer eigenen Wahrnehmung aufgefaßt, und mit dieser Auffassung begeht Descartes nach Kenny einen Homunkulus-Fehlschluß: "To speak of the soul encountering images in the pineal gland is to commit the homunculus fallacy" (a.a.O., 156). Im Sinne der Schritte (1) und (2) meines Diagnoseschemas analysiert Descartes die Wahrnehmungsleistung einer Person als einen Wahrnehmungsprozeß, der sich innerhalb der Person zwischen ihrer Seele und dem durch die Zirbeldrüse weiterverarbeiteten Netzhautbild vollzieht.

Die *Erinnerung* erklärt Descartes analog. Die Seele erkenne eine bestimmte Bewegung der Zirbeldrüse als Erinnerungsspur des gesuchten Gegenstands. Die *esprits animaux* riefen "eine besondere Bewegung in der Hirndrüse hervor, welche der Seele gerade das Objekt vorstellt, dessen sie sich erinnern will, und welche es sie erkennen läßt" (1649, 69 [§ 42]). Ungeklärt bleibt, wie die Zirbeldrüse der Seele etwas zum Erkennen darbieten kann, wie also der Repräsentationscharakter der Erinnerungsspur zustande kommen soll, wo Descartes zuvor gerade die rein mechanische Natur der Erinnerungsspur betont hat. Ein Bild kann ja sowohl in Ansehung als auch unter Absehung von seinen repräsentierenden Eigenschaften weiterverarbeitet werden. Der Ausdruck "Bildverarbeitung" ist in dieser Hinsicht doppeldeutig: Wenn die *Mona Lisa* in eine Häckselmaschine gegeben wird, findet durchaus eine Bildverarbeitung statt, aber eben keine Verarbeitung in Ansehung der repräsentierenden Eigenschaften des Gemäldes. Es macht für die Häckselmaschine keinen Unterschied, ob das, womit sie gefüttert wird, irgendetwas darstellt oder nicht.

Bei Descartes tritt also die Seele in die Rolle des Homunkulus ein. Der Fall liegt hier anders als bei Platon, Freud und Kant, denn Descartes argumentiert ausdrücklich gegen die Aufspaltung der Tätigkeit der Seele in verschiedene Vermögen oder Instanzen. Im Zusammenhang seiner Argumentation für die Unteilbarkeit der Seele oder des Geistes sagt er: "Auch darf man nicht die Fähigkeiten des Wollens, Empfindens, Erkennens als seine [sc. des Geistes] Teile bezeichnen, ist es doch ein und derselbe Geist, der will, empfindet und erkennt" (1641, 74). Alle kognitiven und sensuellen Vermögen werden ein und derselben unteilbaren Substanz zugeschrieben; für Kants Personifizierung des Willens ist Descartes nicht anfällig. Problematisch aus Kennys Oxforder Sicht ist allein, daß Descartes als das, was diese Vermögen hat, eine Seelensubstanz fingiert, wo doch mit der *Person* schon ein bestens geeigneter Träger der mentalen Prädikate zur Verfügung steht. Nichts einzuwenden ist gegen den negativen Teil seiner Klarstellung: Es ist nicht das Auge, das

wahrnimmt, und auch nicht der Körper.<sup>8</sup> Im Rahmen des cartesianischen Substanzdualismus bleibt als Subjekt der Wahrnehmung nur die Seele übrig. Eben diesen Dualismus überbrückt aber der prägenden Auffassung Strawsons zufolge die Kategorie der Person: Personen zeichnen sich gerade dadurch aus, daß ihnen sowohl mentale als auch physische Prädikate zuschreibbar sind.

Daß eine Instanz benötigt wird, der auch physische Prädikate zukommen können, zeigt sich vor allem bei der Zuschreibung von Handlungen. Bei Aristoteles war es ja die Rede von der Seele als Handlungssubjekt, die die Absurdität offenbar werden ließ: Zu sagen, daß "die Seele sich erzürne", ist, "wie wenn man sagte, die Seele webe ein Tuch oder baue ein Haus". Die Parallelisierung mit waschechten Handlungen ist offenbar als *reductio ad absurdum* gemeint. Die Rede von psychischen Tätigkeiten der Seele kommt uns weniger absurd vor als die Rede von öffentlichen, in die Körperwelt hineinwirkenden Handlungen der Seele. Diese Unterscheidung ist aber unklar, denn auch die geläufige Auffassung der Seele als eines Steuerungszentrums für die Bewegung der Organe impliziert ja, daß sie in die Körperwelt hineinwirkt. Die Vorstellung, daß die Seele sich der Organe als ihrer Werkzeuge bedient, ist uns vertrauter als die, daß sie sich eines Webstuhls oder einer Maurerkelle bedient. Besser verständlich ist sie dadurch nicht.

Zurück zum retinalen Bild. Das überaus Irreführende ist hier, daß das Bestrahlungsmuster, das ein äußerer Gegenstand auf die Netzhaut wirft, tatsächlich eine verkleinerte Projektion ergibt, wiewohl deren Abbildungscharakter nicht zur Erklärung dafür herangezogen werden kann, wie aus der Sinnesreizung ein Wahrnehmungserlebnis und schließlich ein zutreffendes Wahrnehmungsurteil entsteht. Die Irrelevanz des projizierten Bildes, das Descartes in seiner selbstgebastelten Camera obscura gesehen hat, und das man mithilfe des von Helmholtz 1850 erfundenen Augenspiegels auch von vorn sehen kann, wird klar, wenn man den Strahlengang im Auge genauer betrachtet.<sup>9</sup> Es wird ja auf der Netzhaut ein Teil des eintreffenden Lichtes reflektiert, ein anderer Teil wird absorbiert. Dasjenige Licht, das ein äußerer Betrachter auf dem Augenhintergrund sehen kann, ist der reflektierte Teil. Dieser Teil trägt gerade *nicht* zur Reizung der Sehzellen des untersuchten Organismus bei. Dies tut allein der absorbierte Teil, den der Experimentator nicht sehen kann. Würde das gesamte Licht absorbiert, würde kein sichtbares Bild entstehen. Es ist also zu unterscheiden zwischen dem von außen sichtbaren *Bild* und dem strukturanalogen *Bestrahlungsmuster* desjenigen Lichtanteils, der tatsächlich die Zapfen und Stäbchen erregt. Diesen Unterschied zwischen dem Bild und dem Bestrahlungsmuster haben Kepler und Descartes übersehen (vgl. Hyman 1989, 18). Darin besteht die Ironie des "Netzhautbildes", das Descartes sah und das der Augenarzt mit dem Ophthalmoskop sehen kann: Dieses projizierte Bild ist für die Wahrnehmung gerade funktionslos. Der Bildcharakter der Reflexion ist eine kostenlose Zugabe, ein Epiphänomen des Sehprozesses.

Es seien zwei mögliche Einwände gegen diese Darstellung diskutiert.

Erster Einwand: "Dann nennen wir das Netzhautbild eben ein Bestrahlungsmuster auf der Netzhaut; das haben wir doch immer gemeint, und der Homunkulus-Einwand ist philosophische Beckmesserei."

---

<sup>8</sup> "c'est l'ame qui sent, & non le cors", "c'est l'ame qui voit, & non pas l'oeil" (Descartes 1637, 109 und 141).

<sup>9</sup> In der Literatur ist manchmal nur von einem auf dem Kopf stehenden Bild die Rede, manchmal von einem zusätzlich seitenverkehrten. Die Erklärung ist einfach: Der erste Fall ergibt sich, wenn man das retinale Bild wie Descartes von hinten betrachtet, der zweite, wenn man es durch das Ophthalmoskop von vorn betrachtet.

Was man immer schon gemeint hat, läßt sich *ex post* leicht behaupten. In diesem Falle kann die Behauptung zurückgewiesen werden, denn wenn man dies immer schon gemeint hätte, wäre das Inversionsproblem gar nicht erst entstanden. Die Frage, warum wir trotz des invertierten Bildes die Gegenstände richtigerum sehen, entsteht ja erst unter der Annahme, daß das Netzhautbild ein phänomenal zugänglicher Teil des Wahrnehmungsprozesses ist. Kepler und vielen anderen erschien das Inversionsproblem eben deshalb als drängend, weil sie diese Unterstellung einfach mitgemacht haben, ohne sich Rechenschaft darüber abzulegen. Freilich muß beim Sehen irgendeine strukturerhaltende Reizumwandlung stattfinden. Daß der Reiz im Linsenauge der Wirbeltiere zeitweise die Form eines verkleinerten, invertierten Bildes annimmt, bevor er weiter umgewandelt wird, ist aber ein kontingentes physiologisches Faktum. Beim Facettenauge der Gliederfüßer verhält es sich anders; dort wird das Strahlenbündel schon durch den optischen Apparat in einzelne Punkte zerlegt und nicht erst auf der Ebene der Sehzellen.

Zweiter Einwand: "Die Rede vom Netzhautbild ist *metaphorisch* gemeint. Niemand glaubt, daß das Gehirn im Wortsinne Bilder sieht, aber seine Tätigkeit läßt sich doch so beschreiben, *als ob* es das Netzhautbild sähe und auswertete."

Es muß erlaubt sein, diesen Einwand mit dem alten Witz über die Funktionsweise des Rundfunks zu kommentieren: "Stelle dir einen ganz langen Dackel vor, der von Boston nach New York reicht. Wenn man ihn in Boston am Schwanz zieht, dann bellt es in New York. Und jetzt denk dir den Dackel weg: So funktioniert Rundfunk!" (nach Hyman 1989, 72). Witze soll man nicht erklären, aber es sei doch die Parallele zu unserem Fall benannt: Es wird in beiden Fällen eine metaphorische Erklärung präsentiert, und dann werden wir aufgefordert, uns genau denjenigen Teil der Metapher wegzudenken, der die Erklärungsleistung erbringt, oder scheinbar erbringt. Metapherntheoretisch ausgedrückt, kann man mit Mary Hesse (1966) zwischen dem *positiven*, dem *negativen* und dem *neutralen* Analogiebereich einer Metapher unterscheiden. Es gibt Merkmale, in denen Bildspender und Bildempfänger übereinstimmen, es gibt irrelevante Merkmale, die nicht übertragen werden, und schließlich Merkmale, bei denen man noch nicht weiß, ob sie zum positiven oder zum negativen Analogiebereich gehören. Auf diesem dritten, *neutralen* Bereich beruht nach Hesse der heuristische Wert einer Metapher oder eines Modells. Auf den Dackel und das Radio angewandt: Zunächst wird der negative Analogiebereich als erklärungskräftig ausgegeben, danach wird die ganze Metapher zurückgezogen: "Denk Dir den Dackel weg!"

Ich stelle nicht in Abrede, daß es heuristisch fruchtbare und erkenntnisbefördernde Metaphern in der Wissenschaft gibt.<sup>10</sup> Das Postulieren von Dackeln gehört aber sicherlich nicht dazu. Und der Verdacht lautet, daß der Betrachter des Netzhautbildes in der Wahrnehmungstheorie die Rolle des Dackels spielt. Aus dem bloßen Vorkommen einer Metapher läßt sich nicht schließen, daß eine Erklärung erschlichen wurde; dies muß jeweils im Einzelfall geprüft werden. Umgekehrt reichen allgemeine Verweise darauf, daß die Psychologie ohne Metaphern und Modelle nun einmal nicht auskommt<sup>11</sup>, zur Verteidigung homunkularer Redeweisen keinesfalls aus.

Hier ist eine allgemeine Bemerkung zum Verhältnis von philosophischer Begriffsanalyse und Naturwissenschaft am Platze. Das Problem des invertierten Netzhautbildes ist ein perfektes Beispiel dafür, wie Denkfehler und konzeptuelle Verwirrungen ein fehlgeleitetes Forschungsprogramm anregen können. Das Invertierungsproblem war ja keine wissenschaftliche Entdeckung, sondern die Folge einer undurchschauten Implikation des Bildbegriffs: Hier ist ein Bild; es zeigt die

<sup>10</sup> Vgl. dazu Keil 1993, 249-271; zu Hesses Unterscheidung vgl. 254 f.

<sup>11</sup> So z.B. Freud, GW XIV, 222 und Dennett 1991, 455.

Außenwelt invertiert; Bilder sind etwas, was gesehen werden kann; 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. Freilich nicht allein durch Nachdenken: Die Einsicht in die Funktionslosigkeit des reflektierten Teils des Lichts für die Wahrnehmung nimmt ja selbst einzelwissenschaftliches Wissen in Anspruch. Begriffliche Analyse und empirische Erkenntnis mußten zusammenspielen, um das Problem aufzulösen. Dabei konnten wir die Frage, ob es eine *generelle* Implikation des Bildbegriffs ist, daß Bilder im Wortsinne gesehen werden können, auf sich beruhen lassen.<sup>12</sup> Für den Homunkulus-Fehlschluß in der Wahrnehmungstheorie spielt jedenfalls die Implikation, daß Bilder gesehen werden können, eine entscheidende Rolle. Und tatsächlich kann das Netzhautbild ja gesehen werden, nur eben nicht von demjenigen, auf dessen Netzhaut es entsteht.

Es geht bei der Diagnose von Homunkulus-Fehlschlüssen nicht darum, Diskurspolizei zu spielen und überall philosophische Sprachdisziplin durchzusetzen. Homunkulus-Fehlschlüsse schaden auch den empirischen Humanwissenschaften selbst, denn sie können gegenstandslose Fragen veranlassen, degenerierende Forschungsprogramme anregen, und das Problem des invertierten Netzhautbildes ist ein hervorragendes Beispiel dafür. Es ist nicht das einzige Beispiel. Heute werden verwandte Einwände gegen verschiedene Theorien in den Kognitionswissenschaften vorgebracht. Es wird darüber gestritten, ob nicht die Begriffe der *Repräsentation*, des *Regelbefolgens*, der *Information* und der *Berechnung* im wörtlichen Sinne allein auf Tätigkeiten von ganzen Personen zutreffen, bei Strafe des Homunkulus-Fehlschlusses.

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 sich dies bemessen soll. Bisher hieß es, daß *Teile von Personen* jedenfalls nicht hinreichend menschenähnlich seien, um ihnen mentale Fähigkeiten zuzuschreiben. Bei welchen Arten von Dingen sollen diese Zuschreibungen noch verboten sein? Wie verhält es sich mit nichtmenschlichen Säugetieren, Kindern vor dem Spracherwerb, hochentwickelten Robotern? Viele Philosophen halten die Auffassung, daß allein Personen als "intentionale Systeme" zählen, für zu stark. Dem Computerfunktionalismus zufolge sind mentale Zustände abstrakt und multipel realisierbar, und aus dieser Sicht erscheint das Beharren auf Personen als den einzigen Trägern mentaler Prädikate als dogmatische Gegenbehauptung.

Für den Kritiker von Homunkulus-Fehlschlüssen wäre es vorteilhaft, sich nicht im einzelnen darauf festlegen zu müssen, welche Wesen oder Arten von Systemen mit intentionalem Vokabular beschrieben werden dürfen und welche nicht. Und tatsächlich gibt es für ihn einen Weg, diese Frage auf sich beruhen zu lassen. Er kann gegen den Opponenten wie folgt argumentieren: "Du hast beansprucht zu erklären, worin eine bestimmte mentale Fähigkeit besteht. Nun hast Du unterstellt, im Geist gebe es etwas mit der gleichen Fähigkeit." Oder: "Du wolltest erklären, was Sehen ist; dabei hast Du auf ein Vokabular zurückgegriffen, das den Begriff des Sehens schon

---

<sup>12</sup> Verhält es sich so, wäre es beispielsweise verfehlt, Vorstellungen als "mentale Bilder" anzusehen. Vgl. dazu Scholz 1991, 180-2 und 1995, 52-6 (s.o., Fn. 5). Nach Scholz ist der Bild- oder Repräsentationscharakter keine intrinsische Eigenschaft eines Gegenstands, sondern eine, die auf einer bestimmten Verwendung des Gegenstands beruht. Wir nennen ein physisches Ding ein Bild, wenn es, im Wortsinne, als Bild *angesehen* werden kann.

voraussetzt." Wenn es dem Homunkulus-Kritiker gelingt, sein Argument im Sinne des Regreßeinwandes zu formulieren, kann er offenbar die schwierige Frage nach dem originalen Anwendungsbereich des intentionalen Idioms dahingestellt sein lassen. Dieser Zug ist freilich nur möglich, wenn der Opponent den Anspruch einer Begriffsanalyse oder einer reduktiven Erklärung erhoben hat. Andernfalls bleibt der Homunkulus-Kritiker auf die intuitive Plausibilität seiner Prämisse (4) angewiesen ("Tatsächlich gibt es nichts hinreichend Personenähnliches innerhalb von Personen"). Wir werden unten am Beispiel von Platon sehen, daß diese Prämisse auch direkt angegriffen wird.

Es gibt eine Erläuterung von "Homunkulus-Fehlschluß", die genau auf das beschriebene wahrnehmungstheoretische Problem zugeschnitten ist, nämlich die von Gareth Evans:

One commits the homunculus fallacy when one attempts to explain what is involved in a subject's being related to objects in the external world by appealing to the existence of an inner situation which recapitulates the essential features of the original situation to be explained. (Evans 1985, 397)

Diese Erläuterung ist vereinbar mit Kennys Charakterisierung des Fehlschlusses als Teil/Ganzes-Verwechslung, ist aber spezifischer als diese. Der Fehler besteht nach Evans in einer leeren Iteration: Es wird eine innere Instanz postuliert, die sich zu inneren Vorgängen so verhält wie eine Person als ganze zu ihrer Umwelt. Da nicht zu sehen ist, warum die dabei verwendeten Begriffe weniger analysebedürftig sein sollten als im ersten Durchgang, ergibt sich ein Regreßproblem.

Die ganze durch Kenny nahegelegte Vorstellung eines *verbotenen Vokabulars* ist geeignet, die Debatte auf ein falsches Gleis zu schieben. Es kann bei der Diagnose von Homunkulus-Fehlschlüssen nicht primär darum gehen, welche Wörter vorkommen. Entscheidend ist vielmehr, welche Erklärungsansprüche mit ihrer Verwendung verbunden werden. Für kognitionswissenschaftliche Theorien, die mit homunkulusgefährdeten Begriffen operieren, muß jeweils im einzelnen untersucht werden, worauf sich die Erklärungsansprüche gründen und ob ein Regreß vorliegt. Eine homunkulare Beschreibung einer kognitiven Leistung mag durchaus eine bloße *façon de parler* sein, die sich bei wohlwollender Interpretation als harmlos erweisen läßt. Ich habe nicht dafür plädiert, jedes Werk dem abgewandelten Hume-Test zu unterziehen: *Enthält es homunkulusverdächtige Ausführungen? Ja? Dann übergebe man es den Flammen, denn es kann nichts als Sophisterei und Blendwerk enthalten.* Mit Bedacht habe ich den Homunkulus-Fehlschluß nicht am Beispiel einer modernen kognitionswissenschaftlichen Theorie diskutiert, um die Einsicht in das grundsätzliche Problem des Postulierens menschenähnlicher Instanzen bei der Erklärung mentaler Fähigkeiten nicht mit hochkontroversen Behauptungen über einzelne Anwendungsfälle zu belasten. Das Problem des invertierten Netzhautbildes scheint mir ein – nach einem mühsamen, jahrhundertelangen Klärungsprozeß – hinreichend unkontroverses Beispiel zu sein.

#### 4. Offensive Rechtfertigungen von Homunkuli

Mit der Diagnose eines Homunkulus-Fehlschlusses ist eine Kontroverse selten zu Ende. Neben den Fällen, in denen der Opponent bestreitet, einen Homunkulus postuliert zu haben, gibt es auch ausdrückliche Rechtfertigungen homunkularer Redeweisen. Einschlägige Versuche, personenähnliche psychische Instanzen zu rehabilitieren, stammen von Platon, Dennett und Lycan. Platons Argument werde ich im folgenden diskutieren; auf Lycans und Dennetts Strategie der

"rekursiven Dekomposition", bei der ein Homunkulus schrittweise durch Teams immer dümmere Homunkuli ersetzt wird, kann ich hier nicht eingehen.<sup>13</sup>

Angesichts eines ausdrücklich vertretenen Instanzenmodells ist der Vorwurf eines *Fehlschlusses* deplaziert. Deshalb habe ich oben die Existenz eines echten Homunkulus-*Problems* eingeräumt, das mit dem Diagnostizieren von Fehlschlüssen nicht behoben ist und zu dem ich nun zurückkehre. Wendungen wie "zwei Seelen in meiner Brust" oder "die Stimme des Gewissens" haben etwas ungemein Plausibles. Es ist nicht leicht zu sagen, was das Gewissen eigentlich sein sollte, wenn nicht eine innere Instanz, die wie eine Person ist und sich beizeiten zu Wort meldet, kurz: "die Stimme des inneren Richters" (Kant, MS 401). Kant spricht auch von einer "zweifache[n] Persönlichkeit, in welcher der Mensch, der sich im Gewissen anklagt und richtet, sich selbst denken muß" (MS 439, Fn.).

Besonders in der Situation eines Entscheidungskonfliktes kommen personifizierende Redeweisen offenbar unserer Selbsterfahrung entgegen. Platon hat nun ein bemerkenswertes Argument für die Aufteilung der Seele in verschiedene Instanzen. Man kann es *das Argument aus der Existenz psychischer Konflikte* nennen. Es lautet, daß es in der Seele manchmal Konflikte oder widerstrebende Tendenzen gibt, während doch "dasselbe nie zu gleicher Zeit Entgegengesetztes tun und leiden" kann (*Politeia*, 436a). Platons Beispiel ist der Durst: Wenn ich durste, will das, was da in meiner Seele trinken will, trinken *und sonst nichts*. Weiß ich beispielsweise, daß das zur Verfügung stehende Getränk vergiftet ist, ändert dieses Wissen an meinem Durst nichts. Man kann zur Dramatisierung an einen dehydrierten Schiffbrüchigen denken, dessen Floß in einem Ozean von Salzwasser schwimmt. Wenn dieser Mensch sich beherrscht und nichts trinkt, dann erklären wir sein Verhalten eben nicht damit, daß er keinen Durst mehr hat. Platon sagt:

Des Durstenden Seele also, wiefern er durstet, will nichts anderes als trinken; dieses begehrt sie und danach strebt sie. [...] Und, nicht wahr, wenn jemals irgend etwas sie zurückzieht, wenn sie durstet, so wäre dies etwas anderes in ihr als das Durstende und sie wie ein Tier zum Trinken Antreibende selbst? (*Politeia* 439a-b).

Platon kommentiert in diesem Zusammenhang noch die Wendung: "Sei stärker als Du selbst!" Diese Rede – im Deutschen sagen wir "Überwinde Dich!" oder "Gib Dir einen Ruck!" – könne nicht ganz wörtlich gemeint sein, denn "wer stärker als er selbst wäre, wäre doch offenbar auch schwächer als er selbst, und der Schwächere stärker; denn es ist doch immer derselbe, der in allen diesen Redensarten auf beiden Seiten angeführt wird" (430 e-431 a). Auch diese Überlegung spricht nach Platon für die Departementalisierung der Seele. Das Argument ist dasselbe wie zuvor: Wo psychische Konflikte sind, muß es mehrere Instanzen geben, die da miteinander konfliktieren, also muß die Seele einzelne Teile enthalten. Diese Teile aber haben ihrerseits zentrale derjenigen psychischen Antriebe und Vermögen, die man sonst Personen zuschreibt, sind also in unserem Sinne Homunkuli. Es ist bei Platon klarerweise nicht so, daß ihm das Postulieren von Homunkuli im Sinne eines Fehlschlusses *unterläuft*. Vielmehr vertritt er das Instanzenmodell offenen Auges und hat ein Argument dafür.

Die Frage ist, ob es ein gutes Argument ist. Wer mit der Existenz innerpsychischer Konflikte argumentiert, sollte die Frage einbeziehen, wie solche Konflikte *beigelegt* werden. Hier gibt es die u.a. von Aristoteles und Thomas von Aquin bekannte Auffassung eines Kräftemessens zwischen

---

<sup>13</sup> Zur Homunkulismus-Debatte in der Philosophie der Kognitionswissenschaften (Stichwort "homunkularer Funktionalismus") vgl. Keil 2003.

zwei Instanzen, meist zwischen "Vernunft" und "Leidenschaft". Gegen diese Auffassung wendet Davidson ein, daß sie "keine sinnvolle Darstellung eines Konflikts im Innern einer Seele des Betreffenden erlaubt", weil sie nämlich "dem entscheidenden Vorgang der Abwägung von Überlegungen keinen Raum" läßt. Es fehle hier "der Stellvertreter des Handelnden – der Wille –", welcher "die Überzeugungskraft der Argumente von beiden Seiten beurteilen, die Entscheidung ausführen und den Rüffel einstecken" kann (Davidson 1970, 63).

Es stehen hier zwei Auffassungen davon einander gegenüber, wie innere Konflikte beigelegt werden. Der einen Darstellung zufolge handelt sich um ein bloßes Kräftemessen, um das "Kräftespiel in der Seele", wie es bei Freud heißt. Ein solches Kräftemessen wird immer der Stärkere gewinnen. Faßt man die Instanzen als mechanische Kräfte auf, lassen sich ihre relativen Größen gerade daran ablesen, welche Instanz am Ende die Oberhand behält. Der zweiten Auffassung zufolge gibt es eine weitere Instanz, die beide Parteien anhört, abwägt und dann entscheidet. Für Davidson, der dieser Auffassung zuneigt, ist diese dritte Instanz der Wille: "Ist der Wille stark, reicht er der Vernunft die Siegespalme; ist er schwach, kann es sein, daß er den Genuß oder die Leidenschaft die Oberhand gewinnen läßt" (ebd.). Diese Auffassung schreibt Davidson auch Platon zu.

Platon scheint aber die Rolle, die Davidson dem Willen zugedacht hat, mit dem *logistikon*, dem überlegenden Seelenteil, zu besetzen. Es gibt zudem einige Stellen, an denen er das *logistikon* mit der ganzen Person *identifiziert*.<sup>14</sup> Man denke an die Formulierung aus dem Wagengleichnis: "Schwierig und mühsam ist daher *bei uns* [!] die Lenkung". Diese Wendung fällt strenggenommen aus dem Gleichnis heraus und stellt den homunkularen Regreß bloß. Im *Staat* heißt es dann, daß man das *logistikon* in sich selbst stärken müsse. Man müsse, so die offen homunkulare Formulierung, "solches tun und reden, wodurch des Menschen innerer Mensch recht zu Kräften kommt" (*Politeia* 589 a).<sup>15</sup> Diese Aufforderung zur Selbsterziehung wirft die Frage auf, wer das denn tun soll. Wer soll es sein, der den überlegenden Seelenteil stärkt? Es muß eine Instanz geben, an die solche Aufforderungen adressiert sind, und diese Instanz kann kein bloßer Teil eines Kräftespiels sein. Sie kann kein bloßer Vektor in einem Kräfteparallelogramm sein, denn zu einem Kraftvektor kann man nicht sagen: "Sei größer als Du bist!" (Man kann es natürlich sagen, aber es wäre eine absurde Aufforderung.) Angesichts dieser Schwierigkeit erscheint auch Davidsons Verteidigung des Dreinstanzenmodells gegenüber dem mechanistischen Zweinstanzenmodell als halbherziges Manöver. Es wirft die Frage auf, wie sich denn die dritte Instanz, wie auch immer man sie benennt, zu der ganzen Person verhält, an die wir Aufforderungen wie "Gib Dir einen Ruck!" im wirklichen Leben adressieren. Man stelle sich einen Straftäter vor, der, gewitzigt durch seine Freud-Lektüre während der langen Untersuchungshaft, dem Richter ins Gesicht sagt: "Es ist über mich gekommen. Ich weiß wohl: Wo Es war, soll Ich werden. Aber das Es war halt stärker als das Ich, es hat gewonnen." Dann möchte der Richter vielleicht gerne antworten: "Du hättest es aber nicht gewinnen lassen *sollen*." (Kant würde genau das sagen.) Für diesen Vorwurf braucht der Richter einen Adressaten.

<sup>14</sup> Ähnlich Aristoteles: "Man darf aber wohl sagen, daß es der denkende Teil ist, der das Wesen des einzelnen Menschen, ganz oder doch in erster Linie, ausmacht." (*Eth. Nic.* 1166 a)

<sup>15</sup> Es findet sich bei Platon geradezu eine Umkehrung des von Kenny beschriebenen Schlusses vom Ganzen auf den Teil: "Und weise [nennen wir jeden einzelnen Menschen] durch jenen kleineren Teil, welcher in ihm herrscht [...], indem auch in dem einzelnen dieser Teil in sich hat die Erkenntnis dessen, was einem jeden und dem Ganzen, aus allen dreien [sc. Seelenteilen] gemeinsam zuträglich ist." (*Politeia* 442 c)

Freud nennt diese Instanz ebenfalls das Ich. Damit aber kommt das Ich zweimal vor: einmal als Bürgerkriegspartei (s.o.), und einmal als diejenige Instanz, an die zur Schlichtung des Bürgerkriegs appelliert wird – und die im Instanzenmodell unanalysiert bleibt. Appelliert wird nicht an "das Ich", sondern an *mich*, kleingeschrieben. Ebenso verhält es sich, wenn Platon zunächst die Seele im Sinne des Wagengleichnisses zerlegt und dann den Wagenlenker noch einmal mit "uns" identifiziert. Man kann Platon und Freud die Frage nicht ersparen, warum sie erst mit der Partitionierung der Psyche angefangen haben, wenn es am Ende doch wieder eine Instanz gibt, die sich die Entscheidung zurechnen lassen muß, und zwar eine unanalysierte Instanz.<sup>16</sup>

Im Lichte dieser Überlegungen verliert auch Platons Argument aus der Existenz psychischer Konflikte an Plausibilität. Daß Menschen konfligierende Wünsche beherbergen, nötigt nicht eo ipso zum Postulieren psychischer Instanzen. Nur wenn man den Wunsch zu trinken von vornherein zum "durstenden Teil der Seele" hypostasiert, ist es – analytisch! – wahr, daß der durstende Teil nicht zugleich nicht durstend sein kann. Es kann aber durchaus der Fall sein, daß ich einerseits trinken will, andererseits nicht, insofern das Trinken im gegebenen Fall das Trinken von etwas Vergiftetem wäre. Die Abneigung, mich zu vergiften, nimmt mir nicht meinen Durst, wird aber, wenn ich vernünftig bin, dazu führen, daß ich den Trinkwunsch nicht die Tat umsetze. Allgemein sollten Wünsche als Handlungsdispositionen aufgefaßt werden, zu denen eine Person sich verhalten kann, nicht als blinde Kräfte, die sie "wie ein Tier zum Trinken antreiben". Eventuelle Konflikte zwischen ihren Wünschen muß die Person erst beilegen, wenn sie ihre Entscheidung trifft, denn unvereinbare Wünsche kann man zwar hegen, aber nicht in die Tat umsetzen. Im Zuge des praktischen Schließens werden aus Wünschen Absichten: Wünsche stehen am Anfang des praktischen Überlegungsprozesses, handlungsanweisende Absichten am Ende.<sup>17</sup>

Bei irrationalen Handlungen liegt der Fall komplizierter; Phänomene wie Willensschwäche, Handeln wider besseres Wissen und Selbsttäuschung verlangen eine gesonderte Diskussion. Es sind ja gerade diese Phänomene, die Davidson und andere bewogen haben, im Anschluß an Freud eine Partitionierung des Geistes in halbautonome Abteilungen vorzuschlagen. Bei diesen Abteilungen handelt es sich nicht um regelrechte Homunkuli, sondern um teilautonome Wunsch- und Glaubenssysteme, zwischen denen die rationalen Beziehungen abgebrochen sind, nicht hingegen die kausalen. Ein in diesem Sinne abgespaltener Wunsch kann zur Handlungsursache werden, ohne ein vernünftiger Grund zu sein. Innerhalb eines Sektors besteht jeweils beträchtliche Konsistenz, während die Person als ganze eine inkonsistente Menge von Einstellungen beherbergt (vgl. dazu Davidson 1982 und Cavell 1993). Wie es der Person gelingt, die Abteilungen getrennt zu halten, bedarf natürlich der psychologischen Erklärung. Partielle Integrationsleistungen erbringen ja auch Personen, die in bestimmten Situationen irrational handeln, und Adressaten von Aufforderungen bleiben sie auch. In der sozialen und physischen Welt bleiben realitätsverleugnende Personen einem beständigen Druck ausgesetzt, die Widersprüche in ihren Einstellungen zu minimieren.

Homunkulismus-Kritik operiert innerhalb des Begriffsrahmens der intentionalen Psychologie. Sie ist in theoretischer Hinsicht der Forderung nach regreßfreier Beschreibung und Erklärung psychischer Aktivitäten verpflichtet; in praktischer Hinsicht ergibt sich daraus eine Ablehnung subpersonaler Handlungssubjekte. Weniger stark ist Homunkulismus-Kritik dem Interesse verpflichtet, in die

<sup>16</sup> Ähnlich argumentieren Roy Schafer und Habermas gegen Freud (vgl. dazu Keil 1993, 131 ff.).

Instanzenmodelle der Psyche tun allgemein gute Dienste, wenn Verantwortung abgewälzt werden soll. Man könnte sich mit Dennett eine "Kafkaesque bureaucracy of homunculi" vorstellen, "who always reply, when challenged: „Don't blame me, I just work here“" (Dennett 1991, 429).

<sup>17</sup> Zum Unterschied zwischen Wünschen und Absichten vgl. Keil 2000, 54-56 und 77-81.

Tiefen der menschlichen Seele vorzudringen und allen psychischen Phänomenen in der Grauzone zwischen rationaler und irrationaler Handlungssteuerung gerecht zu werden. Das Unternehmen, irrationale Phänomene in intentionalen Begriffen zu erklären, ist ohnehin prekär und muß angesichts massiver, pathologischer Irrationalität an Grenzen stoßen. Was die Grauzonen betrifft, dürfte die Zurechnung von Handlungen einen *normativen* Kern haben: Durch die Aufforderung, die Verantwortung für das eigene Tun zu übernehmen, werden psychische Idiosynkrasien, Störungen und Verletzungen "gänzlich bei Seite gesetzt" wie nach Kant die Einflüsse schlechter Erziehung. Zur Verantwortung gezogen und mit Sollensansprüchen konfrontiert werden nun einmal Personen, nicht psychische Instanzen. Ohne Kants Prinzip "Du kannst, weil Du sollst" bliebe Freuds Prinzip "Wo Es war, möge Ich werden" ein frommer Wunsch.

## 5. Schlußbetrachtung

Ich möchte mit einer konstruktiven Bemerkung zur Alltagspsychologie des Homunkulismus schließen. Personifizierende Redeweisen, die unserer Selbsterfahrung vornehmlich in Entscheidungskonflikten offenbar entgegenkommen, sollten nicht ausnahmslos auf bedenklichen Fehlschlüssen beruhen. Es ist einen Versuch wert, dem Fingieren von inneren Stimmen oder von "zwei Seelen in meiner Brust" eine wohlwollende Interpretation zu geben.

Betrachten wir noch einmal das Phänomen des Gewissens. Wenn wir Selbstgespräche moralischen Inhalts führen und dabei unser Gewissen als eigenen Sprecher auftreten lassen, muß dies nicht als Zerlegung unserer selbst in mehrere Teile aufgefaßt werden. Eher *vervielfältigen* wir uns: Wenn man ein Alter Ego fingiert, versetzt man sich typischerweise in die Lage dessen, der man gern sein möchte, der man einmal war oder als der man gesehen werden will. Dieses Alter Ego ist kein Teil einer realen Person, sondern eine *hypothetische ganze Person*. Es ist eine Variation meiner selbst, und anders als in den Instanzenmodellen der Psyche findet hier keine Aufspaltung des psychischen Apparates statt, sondern wir aktualisieren als ganze Personen unsere Fähigkeit der *Perspektivenübernahme*. Jemand nimmt vorübergehend die Rolle beispielsweise der Person ein, die er gern sein oder werden möchte (oder, nach Freud, die des strafenden Elternteils), und holt sich deren Rat ein. Wir schlüpfen in verschiedene Rollen, und diese *dramatis personae* sollten, ganz im Gegensatz zu Dennetts rekursiver Verdummungsstrategie, einer ganzen Person nicht möglichst unähnlich sein, sondern möglichst ähnlich; nicht möglichst beschränkt, sondern möglichst verständig.

Entscheidend ist, daß mit solchen Personifizierungen kein Anspruch auf Analyse oder reduktive Erklärung verbunden ist. Worin auch immer ihr Wert besteht, sie scheinen mir nicht auf einem Fehlschluß zu beruhen. Wenn man diese Alter Egos ebenfalls Homunkuli nennen will, dann gibt es offenbar auch harmlose Homunkuli. In dem Sinn, in dem ich den Ausdruck verwendet habe, ist aber das fingierte Gegenüber kein Homunkulus. Die Rollenprosa ist in Entscheidungskonflikten ein Kunstgriff, um sich die Einstellungen und Gründe, die gegeneinander abzuwägen sind, in ihren jeweiligen Einbettungen und Implikationen zu vergegenwärtigen. Dabei werden Gründe, von denen der Überlegende noch nicht weiß, ob sie handlungswirksam werden, und die möglicherweise die Konsistenz seiner Einstellungen gefährden, vorübergehend ausgelagert und einem Alter Ego in den Mund gelegt. Diese Dramatisierung muß indes in dem Moment aufhören, in dem es zum Handeln kommt. Handlungen müssen wir uns zurechnen lassen, während wir finstere und edle Gedanken auch einmal ausprobieren dürfen, ohne sie schließlich zu handlungswirksamen Gründen werden zu lassen. Allgemein ist das Konsistenzerfordernis auf der Ebene des Handelns stärker als auf der Ebene der Wünsche. Jemand wird nicht dadurch inkonsistent, irrational oder schizophran, daß er

Gründe erwägt, die, würde er nach ihnen handeln, nicht zu seinen sonstigen Einstellungen und zu seinem Charakter passen würden. Auch hypothetische Gründe sind aber umso verständlicher und nachvollziehbarer, je besser sie sich zu einer konsistenten Menge fügen, und deshalb wird, wer noch mit sich zu Rate geht, mitunter dazu neigen, um einige der erwogenen Gründe herum einen Charakter zu entwerfen, zu dem sie besser passen als zum eigenen.

## Literatur

- Aristoteles (*De Anima*): *Von der Seele*, übers. von H. Seidl (nach W. Theiler), Hamburg 1995.
- (*Eth. Nic.*): *Nikomachische Ethik*, übers. von F. Dirlmeier (*Werke* Bd. 6, Hg. Grumach), Berlin 1956.
- Armstrong, D. (1960): *Berkeley's Theory of Vision*, Melbourne.
- Berkeley, G. (1709): *Essay towards a New Theory of Vision* (= *The Works*, Vol. I, Ed. Luce/Jessop), London 1948, Repr. Nendeln 1979.
- Cavell, M. (1993): "Teilung des Selbst", in dies., *Freud und die analytische Philosophie*, Stuttgart 1997, 279-294.
- Davidson, D. (1970), "Wie ist Willensschwäche möglich?", in ders., *Handlung und Ereignis*, Frankfurt am Main 1985, 43-72.
- (1982): "Paradoxes of Irrationality", in: R. Wollheim/J. Hopkins (Eds.), *Philosophical Essays on Freud*, Cambridge, Mass., 289-305.
- Evans, G. (1985): "Molyneux's Question", in ders., *Collected Papers*, Oxford, 364-399.
- Dennett, D. C. (1969): *Content and Consciousness*, London.
- (1991): *Consciousness Explained*, Boston/Toronto/London.
- Descartes, R. (1637): *La Dioptrique*, Adam/Tannery-Werkausgabe Bd. VI, Paris.
- (1641): *Meditationen über die Grundlagen der Philosophie* (Übers. Buchenau), Hamburg 1994.
  - (1649): *Die Leidenschaften der Seele*, hrsg. u. übers. v. Klaus Hammacher, Hamburg 21996.
- Freud, S. (GW): *Gesammelte Werke*, hrsg. v. Anna Freud et al., London 1940 ff.
- Hacker, P. (1991): "Seeing, Representing and Describing: An Examination of David Marr's Computational Theory of Vision", in: Hyman (Ed.), 119-154.
- Hesse, M. B. (1966): *Models and Analogies in Science*, Notre Dame, Indiana.
- Hyman, J. (1989): *The Imitation of Nature*, Oxford.
- (Ed.) 1991: *Investigating Psychology: Sciences of the Mind after Wittgenstein*, London.
- Kant, I. (MS): *Die Metaphysik der Sitten* (Tugendlehre), Akademie-Werkausgabe Bd. VI, Berlin 1907.
- Keil, G. (1993): *Kritik des Naturalismus*, Berlin/New York
- (2000): *Handeln und Verursachen*, Frankfurt am Main.
  - (2003): "Homunkulismus in den Kognitionswissenschaften", erscheint in: W. R. Köhler/H. D. Mutschler (Hrsg.), *Berechenbarkeit des Geistes*, Darmstadt.
- Kenny, A. (1971): "The Homunculus Fallacy", zit. nach Hyman (Ed.) 1991, 155-165.
- Kepler, J. (1604): *Ad Vitellionem Paralipomena, quibus Astronomiae Pars Optica Traditur*, in ders., *Gesammelte Werke*, hrsg. von W. von Dyck und M. Caspar, Bd. II, München 1939.
- Platon: *Phaidros*, Sämtliche Werke IV (Übers. Schleiermacher), Hamburg 1958.
- *Politeia*, Sämtliche Werke III (Übers. Schleiermacher), Hamburg 1958.
- Scholz, O. R. (1991): *Bild, Darstellung, Zeichen*, Freiburg/München.
- (1995): "Bilder im Geiste? – Das Standardmodell, sein Scheitern und ein Gegenvorschlag", in: K. Sachs-Hombach (Hrsg.), *Bilder im Geiste*, Amsterdam, 39-61.
- Schumacher, R. (1998): "Der Homunkulus-Fehlschluß in der Philosophie des Geistes. Eine Verteidigung des Repräsentationalismus", in: E. Dölling (Hrsg.), *Repräsentation und Interpretation*, Berlin, 39-60.
- Thalberg, I. (1982): "Freud's Anatomies of the Self", in: R. Wollheim/J. Hopkins (Eds.), *Philosophical Essays on Freud*, Cambridge, Mass., 241-263.

\* \*

**Erstpublikation in: Zeitschrift für philosophische Forschung, Band 57, Heft 1 (2003), S.1-26. Nachdruck mit freundlicher Genehmigung des Verlages.**