

Wolf-Jürgen Cramm

Repräsentation oder Verständigung?

Eine Kritik naturalistischer Philosophien der Sprache und des Geistes

Vorwort

Die Arbeit wurde 2003 am Fachbereich für Philosophie und Geschichtswissenschaften an der Johann Wolfgang Goethe-Universität Frankfurt am Main als Dissertation im Fach Philosophie angenommen und wurde für die vorliegende Fassung nur geringfügig verändert. Ich danke meinen Betreuern Jürgen Habermas und Friedrich Kambartel für die freundliche Unterstützung und Geduld.

Inhaltsverzeichnis

Einleitung	1
Teil I NATUR, GEIST UND BEDEUTUNG	8
1. Naturalismus: Motive und Abgrenzungen	8
<i>NANÜCS und DOGNA</i>	13
<i>Schritte zu einer genaueren Bestimmung des Gegenstandes und der gemeinsamen Voraussetzungen naturalistischer Theorien</i>	18
<i>Grundvarianten des Naturalismus</i>	25
2. Zur kategorialen Differenz von Natur und Sinn	35
<i>Kategorienfehler – Die Ausgangslage</i>	35
<i>Geist: Eigenschaft oder Sinnpotential?</i>	42
<i>Die Dichotomie von Verantwortung und Geschehen</i>	48
<i>Zur Frage der Normativität von Rationalität und Bedeutung</i>	52
<i>Exkurs: Zu Robert Brandoms normativ-pragmatischer Grundlegung der Semantik</i>	64
<i>Zusammenfassung und weitergehende Erläuterung der Argumentationsstrategie</i>	71
Teil II REDUKTIVER MATERIALISMUS	81
1. 'Intention Based Semantics' – Der Versuch einer Reduktion von Semantik auf ‚materialistische Psychologie'	85
2. Erklärung, Kausalität, Reduktion	91
<i>Was und wie erklären Erklärungen?</i>	92
<i>Die Anwendung des deduktiv-nomologischen Erklärungsschemas auf Handlungen</i>	104
3. ‚Funktion' als Schlüsselbegriff einer materialistischen Theorie des Geistes	110
<i>Das Computermodell des Geistes</i>	115
<i>Zur Frage einer angemessenen Definition des Funktionsbegriffs und zum Stellenwert teleologischer Erklärungen</i>	137

4. Nicht-teleologischer Funktionalismus und kausalen Rollen	147
<i>„Enger Gehalt“ und die Präsuppositionen empiristischer Theorien der Bedeutung</i>	171
<i>Theorien direkter Bezugnahme und die externalistische Kritik am Konzept „engen Gehaltes“</i>	178
5. Teleofunktionaler Externalismus: Dretske und Millikan	189
<i>Dretske</i>	191
<i>Millikan</i>	209
<i>Lässt sich propositionaler Gehalt und sprachliche Bedeutung auf der Grundlage einer evolutionistisch begründeten Korrespondenztheorie erklären?</i>	216
Teil III INTERPRETATIVER MATERIALISMUS	238
1. Quine, Davidson und Dennett	238
<i>Quine als Hintergrund</i>	240
<i>Quines Dogmen</i>	244
<i>Davidsons „integrierte Theorie der Bedeutung und des Handelns“</i>	256
<i>Dennetts Theorie der Perspektivität des „Intentionalen“</i>	263
2. Argumentationslinien gegen eine Reduzierbarkeit des „Intentionalen“	272
3. Davidson über Voraussetzungen für „Gedanken“	279
4. Geteilte Bedeutung? – Zum sozialen und normativen Charakter von Bedeutung	286
<i>Welcher Gemeinsamkeit bedarf Verstehen? – Und was ist eigentlich eine Sprache?</i>	287
<i>Der institutionelle Charakter sprachlicher Bedeutung und der normative Status sprachlicher Handlungen</i>	301
5. Bedeutungsverstehen, Wahrheitsfähigkeit und Propositionalstruktur	310
<i>Zur kommunikativen Struktur propositionalsprachlicher Verständigung</i>	320
<i>Der produktive Charakter von Verstehens- und Verständigungsbemühungen</i>	328
Inhaltsverzeichnis	331

Der erste Schritt ist der ganz unauffällige. Wir reden von Vorgängen und Zuständen, und lassen ihre Natur unentschieden! Wir werden vielleicht einmal mehr über sie wissen – meinen wir. Aber eben dadurch haben wir uns auf eine bestimmte Betrachtungsweise festgelegt. Denn wir haben einen bestimmten Begriff davon, was es heißt: einen Vorgang näher kennenzulernen. (Der entscheidende Trick im Taschenspielerkunststück ist getan, und gerade er erschien uns unschuldig).

(L. Wittgenstein)

EINLEITUNG

Wie fügen sich solche, an sich recht vertraute, Phänomene wie Verständigung, sprachliche Bedeutung, Begrifflichkeit, Aussagegehalt (Propositionalität), Wissen, Gründe, Vernunft und ähnliches in ein modern verstandenes Bild einer natürlichen Ordnung der Welt? Ich werde mich hier kritisch mit neueren philosophischen Positionen auseinandersetzen, die auf diese Frage in einer, wie ich denke, *methodisch vereinsseitigenden* Weise Antworten zu finden suchen. Damit meine ich eine grundsätzliche Orientierung an Formen *naturwissenschaftlicher Theoriebildung* und *Erklärung*. Positionen, welche die Eingangs genannten Phänomene so deuten, als hätte unser Wissen über sie die Form einer *empirischen Theorie über Dinge, Ereignisse oder Zustände innerhalb einer kausalen Weltordnung* – bzw. als könne oder müsse jedes ‚echte‘ diesbezügliche Wissen eine solche Form annehmen – bezeichne ich als *naturalistisch*. Die jeweilige Festlegung darauf, welche *Arten von Eigenschaften* in entsprechenden Theorien oder Erklärungen eine Rolle spielen, kann dabei durchaus unterschiedlich ausfallen. Positionen eines solchen, wie ich auch sagen möchte, *dogmatischen* Naturalismus soll hier mit einem (noch genauer zu bestimmenden) normativ-pragmatistischen Verständnis unserer sprachlichen Praxis im Zusammenhang mit den eingangs genannten Begriffen oder Phänomenen entgegengetreten werden.

Themen, die heute im Umkreis des Themas ‚Bewusstsein‘ – insbesondere im Sinne eines ‚qualitativen Erlebens-Wie‘ – diskutiert werden, werden in meinen Überlegungen keine Rolle spielen. Ich werde mich auf diejenigen Fragen konzentrieren, die in den neueren Debatten unter den Labels *propositionaler Gehalt* oder (etwas missverständlich) ‚*Intentionalität*‘¹ firmieren.

Ausgangspunkt naturalistischer Philosophien sprachlicher Bedeutung und des Geistes ist die Annahme, dass, wenn unsere ‚intentionalen‘ Redeweisen in dem Sinne wahrheitsfähig sein sollen, dass sie wirklich Existierendes zum Gegenstand haben, ihre Aussagen sich auf Dinge, Ereignisse oder Zustände beziehen müssen, die kausale Relevanz besitzen. Kausale Relevanz aber kann etwas nur auf Grund seiner physischen Eigenschaften haben. Demzufolge könnten Geist und Bedeutung nur dann ‚wirklich‘ sein, wenn sie in irgendeinem Sinne als kausal relevante Eigenschaften von Dingen, bzw. als kausale Ereignisse oder Zustände verstanden werden können, die raum-zeitlich lokalisiert sind. Meinen (naturalismus-) kritischen Überlegungen wird demgegenüber die Überzeugung zu Grunde liegen, dass diese Annahmen im Hinblick auf unsere Redeweisen über Geistiges und Bedeutungshaftes *in einer grundsätzlichen Weise unangemessen* sind.

Nun scheint es aber auch unabhängig von solchen Annahmen auf den ersten Blick durchaus sinnvoll, Fragen etwa der folgenden Art zu stellen: Wie ist es möglich, dass bestimmte Folgen von Geräuschen oder Linien auf dem Papier eine Bedeutung haben können – etwas sein können, das wahr oder falsch ist bzw. etwas in einem propositionalen Sinne repräsentiert? Wie können bestimmte materielle Organisationsformen innerhalb der natürlichen Welt, wie etwa menschliche Organismen, Überzeugungen über die Welt haben, beim Überlegen richtige Schlüsse ziehen und gar vernünftig sein? Und wie erklären wir die Möglichkeit des Zustandekommens von *wahren* Überzeugungen in solchen Organismen, das heißt, was ‚garantiert‘ oder sorgt dafür, dass Überzeugungen *richtig* repräsentieren, also irgendwie jener äußeren Welt entsprechen, die sie zum Gegenstand haben?

Der Versuch, auf solche oder ähnliche Fragen Antworten zu finden, kann philosophisch ebenso stimulierend wirken, wie die Fragen selbst als unverfänglich erscheinen können. Und warum auch sollten wir hier nicht zu erhellenden Antworten

¹ Ich verwende den Begriff ‚Intentionalität‘ hier im Sinne der in der Philosophie des Geistes üblich gewordenen, an Brentano anschließenden, laxen Redeweise, mit der ursprünglich so etwas wie eine (Auf-Welt-) ‚Gerichtetheit‘ mentaler Zustände gemeint war, der inzwischen aber meist im Sinne von ‚Repräsentationalität‘, bzw. ‚Semantizität‘ verwendet wird. Da ich zum einen die ganze Rede von ‚Intentionalität‘ als *Eigenschaft* von Zuständen oder Ereignissen für zumindest missverständlich halte, zum anderen Verwechslungen mit der ursprünglichen Bedeutung des Terms ‚intentional‘ im Sinne von absichtlich vermeiden will, bleibt mir nichts anderes übrig, als den Ausdruck nur im Zusammenhang mit Anführungszeichen zu verwenden.

gelangen können, die sich innerhalb einer Begrifflichkeit natürlicher Dinge, Eigenschaften oder Vorgänge bewegen? Haben sich nicht zahllose Phänomene, die im Rahmen überkommener kultureller Deutungsmuster noch als irgendwie übernatürlich, geheimnisvoll und mysteriös anmuteten, durch Erklärungen auf der Grundlage natürlicher, physischer Gesetzmäßigkeiten aufklären lassen? Und leben wir nicht in einer Zeit, in der eine nüchtern-wissenschaftlich inspirierte Sichtweise auf die Welt – trotz aller Kritik an einer möglichen Einseitigkeit ‚technisch halbirter Vernunft‘ – noch immer unbestrittene Triumphe feiert; und zwar sowohl im Hinblick auf die systematische Erklärung von Vorgängen in der Natur, als auch hinsichtlich ihrer technisch-praktischen Beherrschbarkeit? Müsste es da nicht auch möglich sein, jene Rätselhaftigkeiten zu entschleiern, die ‚den Geist‘ bzw. Geistiges, zumindest seit der Vorherrschaft des kartesischen Substanzdualismus von Geistigem und Materiellem, zu umgeben scheinen? Altbekannte philosophische Fragen nach dem Wesen bzw. den Bedingungen der Möglichkeit sprachlicher Bedeutung, von Geist, Erkenntnis und Wissen könnten sich doch so reformulieren lassen, dass sie einer explanativ fruchtbaren Bearbeitung im Rahmen des etablierten Methodenkanons empirischer Wissenschaften zugänglich werden. Eine naturalistische Integrationsstrategie verspricht jedenfalls, solchen Fragen wie „Was ist Bedeutung?“, „Was ist der Geist (das Geistige)?“ oder „Was ist Wissen?“ den Unterton ontologischer Verwunderung zu nehmen, die einen überkommen kann, wenn man über diese Fragen in einer bestimmten Weise nachdenkt. Denn durch die Integration von Bedeutung und Geist in den Rahmen empirischer Theoriebildung und kausaler Erklärung würde der Blick auf die kausal relevanten und empirisch messbaren Eigenschaften solcher Phänomene gelenkt und diese damit in das Bezugssystem einer als unproblematisch gemutmaßten Ontologie von Zuständen und Ereignissen einfügt.

Eine solche Perspektive mag durchaus verlockend klingen. Wir sollten allerdings bedenken, dass wir schon durch die Art und Weise, in der grundsätzliche Fragen nach dem Wesen und der Möglichkeit von Bedeutung, Vernunft, Wahrheit oder Wissen gestellt werden, in irreführender Weise in eine bestimmte Richtung gedrängt werden könnten, in der die Antworten zu vermuten wären. Wir tun also gut daran, die mit bestimmten Formulierungen (zu denen auch die oben genannten gehören) oder Reformulierungen einhergehenden Unterstellungen nicht umstandslos und vorschnell als für bestimmte explanative Zwecke angemessen oder als überhaupt sinnvoll zu akzeptieren – also etwa, dass es Geräusche sind, die eine Bedeutung haben oder dass es Organismen sind, die Schlüsse ziehen oder die die Welt ‚da draußen‘ repräsentieren.

Es ist nun eines der Ziele dieser Untersuchung, zu zeigen, dass solche Bedenken gerade im Hinblick auf die methodischen und begrifflichen Voraussetzungen, unter

denen naturalistische Theorien des Geistes und der Bedeutung entsprechende Fragen angehen, tatsächlich begründet sind. Als beispielhaft für eine solche, wie ich meine, in die Irre führende Art des Herangehens, möchte ich eine Stelle aus Fred Dretskes *Explaining Behaviour* (1988) anführen:

„It is the business of this book to show how this apparent conflict, a conflict between two different pictures of how human behaviour is to be explained, can be resolved. The project is to see how reasons – our beliefs, desires, purposes and plans – operate in a world of causes, and to exhibit the role of reasons in the *causal* explanation of human behaviour. In a broader sense, the project is to understand the relationship between the psychological and the biological – between, on the one hand, *reasons* that people have for moving their bodies and, on the other, the *causes* of their bodies‘ consequent movements.“ (Preface, S. X)

Dieses Zitat ist besonders instruktiv, da hier gleich mehrere für den dogmatischen Naturalismus typische Voraussetzungen im Spiel sind. So wird zunächst Geistiges (im hier in Frage stehenden Sinne) mit propositionalen Einstellungen im Sinne von Zuständen oder Ereignissen an oder in Individuen identifiziert und zugleich unterstellt, die Kenntnis oder Zuschreibung solcher Zustände oder Ereignisse diene dem Zweck der *Erklärung* von *Verhalten* durch *Gründe*, die die *Ursache* eines Verhaltens sind. Geist entspricht gemäß Dretske denjenigen Phänomenen, mit denen sich *psychologische* Erklärungen durch Gründe beschäftigen, und die Frage sei dann, *in welchem Sinne Erklärungen durch Gründe als Kausalerklärungen verstanden werden können*.

Ohne an dieser Stelle bereits im Einzelnen auf die Problematik solcher oder ähnlicher Voraussetzungen eingehen zu wollen, möchte ich doch schon so viel verraten, dass die Gründe, weshalb ich denke, eine solche Herangehensweise sei einem angemessenen Verständnis von Geistigem und sprachlicher Bedeutung abträglich, von grundlegender Natur sind und mit fundamentalen begrifflichen Gegensätzen zu tun haben. Dies bedeutet auch, dass die auftretenden Schwierigkeiten naturalistischer Theorien nicht durch lediglich *interne* Modifikationen oder Korrekturen zu beheben sind. Allerdings kann uns eine entsprechende, grundsätzlich ansetzende Kritik nicht die inhaltliche Auseinandersetzung mit paradigmatischen Positionen des Dogmatischen Naturalismus ersparen. Ansonsten besteht nämlich die Gefahr, nur zu einer weiteren Kompartimentierung philosophischer Diskurse beizutragen, die sich an unterschiedlichen Bestimmungen dessen orientieren, was überhaupt als akzeptable Begrifflichkeit oder sinnvolle Problemstellung gilt, und die sich dabei schließlich soweit gegeneinander abschotten, dass nicht mal mehr der Versuch einer übergreifenden rationalen Auseinandersetzung unternommen wird. Eine zwar kritisch-distanzierte, sich aber gleichwohl im Einzelnen mit naturalistischen Theorien der

Bedeutung und des Geistes auseinandersetzungsvorgehensweise soll deshalb zeigen, dass diese bereits ‚diskursintern‘ – in je spezifischer Weise – mit einer ganzen Reihe hartnäckiger Schwierigkeiten konfrontiert sind. Dabei wird allerdings stets das weitergehende Ziel verfolgt, zumindest plausibel zu machen, dass diese Schwierigkeiten am Ende einen gemeinsamen Ursprung haben, nämlich eine Konfusion hinsichtlich grundlegender begrifflicher Verhältnisse. Eine kurze systematische Vergegenwärtigung einiger für unser modernes Selbstverständnis wesentlichen Unterscheidungen und Voraussetzungen soll deshalb zunächst die allgemeinen Gründe offenlegen, welche die von mir vertretene anspruchsvollere These stützen sollen, nämlich dass Naturalisierungsbemühungen der Art, wie sie heute insbesondere in der angelsächsischen Philosophie des Geistes üblich sind scheitern *müssen*; es soll gezeigt werden, warum man bereits in die Irre geht, wenn man etwa Zeichen oder ihren Gebrauch im Sinne von Dingen oder Ereignissen versteht, die unter gewissen physischen Bedingungen oder kausalen Umständen Eigenschaften erlangen können, die sie zu ‚bedeutungsvollen Dingen‘ oder ‚Ereignissen‘ machen; oder wenn man davon ausgeht, dass Überzeugungen oder Wissen subjektinterne Zustände sind, die dann Welt repräsentieren, wenn sie spezifischen kausalen Bedingungen genügen müssen.

Im Sinne der anspruchsvolleren These werde ich eine Version derjenigen Form von Argumentation verteidigen, gemäß der wir im Falle solcher Voraussetzungen Verständnisirrtümern unterliegen, die sich als *Kategorienverstoß* bezeichnen lassen. Dieser Terminus ist, zumindest im Zusammenhang mit der philosophischen Diskussion um eine angemessene Analyse mentaler Begriffe, bekanntermaßen durch Ryles‘ Verwendung in *The Concept of Mind* (1949) populär geworden². Als Kategorienverstoß möchte ich, allgemein gesprochen, Redeweisen bezeichnen, bei denen Terme in einer Weise kombiniert werden, die zwar ‚oberflächengrammatisch‘ als korrekt und unproblematisch erscheinen kann, jedoch ‚logisch‘ bzw. in begrifflicher Hinsicht unangemessen oder sinnwidrig ist, da mindestens einer der Terme relativ zu den anderen sozusagen in einen falschen kategorialen Rahmen gestellt wird. In Bezug auf die uns interessierenden Fälle lässt sich das vorläufig so verstehen, dass Redeweisen über Bedeutungen, Begriffe, Behauptungen, Urteile, propositionale

² Der Einwand des Kategorienverstoßes ist eng verwandt mit der auf Wittgenstein zurückgehenden Position, die – etwas missverständlich – als ‚*Sprachdualismus*‘ bezeichnet wird. Die entsprechenden Argumente werden aber seit einiger Zeit in weiten Kreisen analytisch orientierter Philosophen als desavouiert betrachtet. Dies wird meist mit Quines erfolgreicher Kritik an der Unterscheidung zwischen analytischen und synthetischen Aussagen begründet. Ich möchte deshalb schon hier darauf hinweisen, dass ein Argument des Kategorienverstoßtyps nicht voraussetzen muss, dass es eine *scharfe* Grenze zwischen begrifflichen (bzw. die Bedeutung betreffenden) und empirischen Fragen gibt.

Gehalte, Überzeugungen, Wünsche, Absichten, Erkenntnisse, Wissen u. ä. m. innerhalb einer (wie wir in Anlehnung an eine Carnapsche Ausdrucksweise sagen können³) ‚begrifflichen Sphäre‘ angesiedelt sind, deren kategorialer Rahmen durch eine (Form der) *Beurteilbarkeit* im Sinne von *richtig* und *falsch* bestimmt ist, die direkt oder indirekt *auf Handlungsalternativen bezogen* ist, wobei die faktischen Stellungnahmen, insgesamt gesehen, einen für die beurteilten Verhältnisse in gewissem Sinne (eingeschränkt) *konstituierenden Charakter* haben. Die entsprechenden Begriffe oder Verhältnisse können daher nicht sinnvoll in Kontexte einer begrifflichen Sphäre übertragen bzw. innerhalb dieser erklärt werden, die durch Begriffe wie ‚Ding‘, ‚Ereignis‘, ‚Zustand‘, ‚Ursachen‘, ‚Naturgesetz‘ u. ä. m. bestimmt ist, *wenn* diese in den – für naturwissenschaftliche Redeweisen konstitutiven – kategorialen Rahmen *natürlicher Vorkommnisse* oder *Geschehnisse* eingeordnet werden. Wenn dies richtig ist, können beispielsweise in einem solchen Sinne verstandene *Zustände*, unter welcher Beschreibung auch immer, nichts sein, was originär (nicht-metaphorisch) *wahr* oder *falsch* ist. Ich werde begriffliche Verständnisse, die wir dem erstgenannten Rahmen zuordnen, *kategorial normative Begriffe* nennen und begriffliche Verständnisse, die wir dem zuletzt genannten Rahmen zuordnen, *kategorial deskriptive Begriffe*.

Dies gilt es natürlich genauer zu erläutern. Doch falls der Vorwurf des Kategorienverstoßes berechtigt ist, müssen alle Formen theoretischer Bemühungen scheitern, die sich Fragen nach dem Wesen und den Bedingungen von Bedeutung und Geist vorrangig im Rahmen einer natürlichen Zustands- oder Ereignisontologie einerseits und einer methodologischen Orientierung an empirischer Beobachtung und Kausalerklärung andererseits, annähern.

In Teil I dieser Arbeit soll einerseits das thematische Feld abstecken, andererseits zugleich die nachfolgende kritische Auseinandersetzung mit den heute dominierenden naturalistischen Theorien der Bedeutung und des Geistes systematisch vorbereiten. Dazu werden zunächst kurz die allgemeinen Motive für naturalistische Einstellungen nachgezeichnet und deren unproblematische Formen von den hier zu kritisierenden, philosophischen Theorien des Naturalismus unterschieden. Im zweiten Abschnitt des I. Teils werde ich dann die begrifflichen Grundlagen meiner Kritik naturalistischer Positionen entwickeln.

Teil II setzt sich mit den heute einflussreichsten Formen eines – im weiten Sinne gedeuteten – *Reduktiven Materialismus* auseinander. Dazu werden zunächst wichtige bedeutungs- und wissenschaftstheoretische Voraussetzungen entsprechender Theorien oder Positionen dargestellt bzw. kritisch diskutiert. Schwerpunkt dieses Teils

³ R. Carnap: ‚Überwindung der Metaphysik durch logische Analyse der Sprache‘ (1931/1932).

bildet die kritische Auseinandersetzung mit sowohl kausalen, als auch teleologischen Varianten des *Funktionalismus*.

Teil III besteht in der ausführlichen Auseinandersetzung mit einer grundsätzlichen, gleichwohl naturalistischen Alternative zum Reduktiven Materialismus, die ich ‚*Interpretativen Materialismus*‘ nenne. Den Schwerpunkt wird die zu den einflussreichsten philosophischen Positionen der letzten ca. 30 Jahre zählende Theorie Donald Davidsons bilden, wobei ich mich wiederum auf sprachphilosophische Fragen konzentrieren werde. Die von mir kritisierte Position Davidsons dient dann auch als Kontrastfolie für einige konstruktive Überlegungen zum Thema Bedeutung und Verständigung, die ich am Schluss der Arbeit anstellen werde.

I. NATUR, GEIST UND BEDEUTUNG

Naturalismus: Motive und Abgrenzungen

Der moderne Naturalismus nimmt klassische Motive des Materialismus und ähnlicher ‚Ismen‘, wie des Physikalismus oder des Mechanismus auf, ist jedoch in seiner Ausrichtung allgemeiner und daher unabhängiger von historischen Selbstverständnissen der Wissenschaften⁴. Da Naturalismus heute üblicherweise im Sinne einer umfassenden *methodischen* Orientierung an den Naturwissenschaften bestimmt wird, scheint eher eine weitgehende Deckung mit dem Begriff des *Szientismus* gegeben. Dass ich den Begriff des Naturalismus dennoch dem des Szientismus vorziehe, hat im Wesentlichen zwei Gründe: Zum einen wird letzterer normalerweise pejorativ verwendet und dient daher auch nicht der Selbstbeschreibung von Naturalisten. Ich möchte die folgenden Darlegungen aber nicht von vornherein mit eindeutigen Wertungen belasten. Vor allem aber ist meine Verwendung von „Naturalismus“ eine noch allgemeinere, da ich mit diesem Begriff auch Positionen umfassen will, deren methodische Orientierung sich nicht mehr als eine – im üblichen Sinne – naturwissenschaftliche verstehen lässt – wenngleich naturwissenschaftlich orientierte Positionen den Schwerpunkt meiner Auseinandersetzung bilden werden.

Jener jüngste Naturalisierungsschub, auf den ich mich in dieser Arbeit konzentrieren will, entfaltete seine Wirkung ungefähr seit den 70er Jahren des 20. Jahrhunderts. Naturalistische Denkweisen begannen sich zunehmend der Bestimmung von Problemstellungen, Begrifflichkeiten und Methodologien solcher Themen zu bemächtigen, die üblicherweise unter Überschriften wie „Philosophie der Sprache“, „Handlungstheorie“ oder „Erkenntnistheorie“ bearbeitet wurden. Manch einem mag jenes, sich in ungemein dynamischer Weise entwickelnde, ‚allgemeine Naturalisierungsprojekt‘ dabei wie der Rettungsanker für ein universalistisch-objektivistisches Weltbild aufgedrängt haben – angesichts des drohenden Relativismus nach-kuhnscher Entwicklungen in der Wissenschaftstheorie, des aufkommenden ‚Postmodernismus‘ und dem abermals postulierten ‚Ende der Philosophie‘ zum Ausklang der 70er Jahre⁵. Die Vielzahl von entsprechenden Publikationen, die jeweils kritisch-konstruktiv aufeinander Bezug nehmen, scheint mir die Redeweise von einem ‚allgemeinen Projekt‘ zu rechtfertigen, auch wenn die maßgeblichen Gemeinsamkeiten

⁴ Vergl. G. Keil: *Kritik des Naturalismus* (1993).

⁵ Vergl. z. B. R. Rorty: *Der Spiegel der Natur* (1979/1981).

unterschiedlicher Positionen innerhalb des naturalistischen Lagers nicht immer auf den ersten Blick zu erkennen sind.

Geläufige Formen der Abgrenzung naturalistischer von nicht- (oder ‚anti‘-) naturalistischen Betrachtungsweisen oder Positionen bestehen darin, letztere, im Gegensatz zu ersteren, als solche zu charakterisieren, die in Kontrastierungen wie Regelmäßigkeit *versus* Regelhaftigkeit, kognitives System *versus* erkennendes Subjekt, Verhalten *versus* Handlung, faktische Akzeptanz (oder pointierter: ‚Zustimmungsverhalten‘) *versus* rationale Akzeptabilität, Disposition *versus* Intention oder etwa auch (Soziales) System *versus* Lebenswelt *begrifflich inkompatibel, zugleich aber doch unverzichtbare* Unterscheidungen sehen. Bei vielen der heute vorherrschenden Debatten sind die Frontlinien jedoch – zumindest in terminologischer Hinsicht – nicht so sauber zu ziehen, wie es diese Unterscheidungen nahelegen. Nicht alle derjenigen Positionen, die ich als „naturalistisch“ bezeichnen will, halten sich, zumindest auf den ersten Blick, bzw. in ihrer Ausdrucksweise, an die genannten Kontrastierungen. Häufig wird auch in naturalistischen Kontexten von einem ‚Regelwissen‘ oder von ‚Handlungen‘ gesprochen. Damit scheint man dem naheliegenden Einwand entgegentreten zu wollen, schlichtweg das Thema zu wechseln. Wenn ich mit der Behauptung der Inkompatibilität der entsprechenden begrifflichen Sphären recht habe, kann es sich hierbei allerdings lediglich um einen Etikettenschwindel handeln. Manche der von mir zu kritisierenden Positionen betonen sogar einige der genannten Unterscheidungen als notwendig, weshalb sie meist nicht als naturalistisch eingestuft werden. In Bezug auf solche Positionen müsste ich also plausibel machen können, dass sie Begriffe der Sphäre kategorial normativer Begriffe in unzulässiger Weise mit kategorial deskriptiven Begriffen in Zusammenhang bringen.

Andererseits muss aber auch der Naturalismus, um als philosophische These überhaupt interessant zu sein, auf mehr festgelegt sein, als das, was heutzutage unter intellektuell ernstzunehmenden Leuten ohnehin als selbstverständlich gilt. Dies gilt etwa für Überzeugungen, wie die, dass alles in der Welt mit rechten Dingen zugeht, bzw. es keinen übernatürlichen ‚Hokuspokus‘ gibt⁶. Um also eine Bestimmung dessen

⁶ Nur darauf scheint sich etwa noch McDowells Bestimmung von „Natur“ zu stützen, wenn er den Begriff durch das Konzept einer ‚zweiten Natur‘, die all jenes umfassen soll, was wir im weiten Feld der ‚Kultur‘ ansiedeln können, aufbläht. Unabhängig davon, ob es überhaupt sinnvoll ist, bei der Charakterisierung des Faktums kulturell geprägter Vermögen (wie besonders des sprachlichen Weltzugangs) als „unsere Natur“ von „Natur“ in einem Sinne sprechen zu wollen, der beansprucht, noch Gemeinsamkeiten mit dem aufzuweisen, was McDowell selbst als den Bereich des *naturgesetzlich* Geregelten bestimmt hatte, entbehrt ein solch weites – eigentlich anti-naturalistisches – Konzept von Natur jeder interessanten Kontrastierung (J. McDowell: *Mind and World* (1994)). Ähnliches gilt für frühere Überlegungen von P. Strawson (ders.: *Naturalismus und Skeptizismus* (1985/1987)).

vornehmen zu können, was ich für den eigentlichen Kern derjenigen Positionen halte, gegen die sich meine kritischen Überlegungen richten, sollten wir zunächst die eher unproblematischen, das heißt, weitgehend geteilten – oder zumindest von mir nicht zu problematisierenden – Grundintuitionen von denen, nach meinem Dafürhalten, unzulässigen Verallgemeinerungen oder Übertragungen durch einen Naturalismus im Sinne eines *philosophisch begründeten Dogmas* abgrenzen.

Eine eher grundsätzliche, philosophische Motivation vieler Naturalisten besteht fraglos in der Hoffnung, die Erfolgsgeschichte neuzeitlicher Naturwissenschaften auch in Bereiche ausdehnen zu können, für deren Bearbeitung nach dem Urteil von Philosophen wie Kant, Frege, Husserl, Heidegger oder Wittgenstein – wegen des eigentümlichen (transzendentalen, apriorischen, normativen, ‚grammatischen‘ oder reflexiven) Charakters der für sie spezifischen Fragen – eigenständige, gerade nicht rein erfahrungswissenschaftlich zu verstehende Methoden geboten seien. Warum, so können wir die oben bereits angedeutete naturalistische Intuition formulieren, sollte es eigentlich nicht möglich sein, unser Verständnis der Welt und unserer Stellung in ihr auch in solchen Bereichen in einer dem naturwissenschaftlichen Erkenntnisfortschritt analogen Weise zu erweitern, die klassischerweise als eine Domäne der Philosophie galten. Schließlich sind wir doch ein Teil der natürlichen Welt. Da muss es doch auch zu Fragen nach unseren Erkenntnismöglichkeiten, nach den natürlichen Bedingungen von Wahrnehmungen, wahren Überzeugungen, Wissen, gelingender sprachlicher Kommunikation, vernünftigem Denken und dergleichen erfolgversprechende, empirische Untersuchungen geben können! Offenbar sind wir doch schon jetzt, auch aufgrund empirisch gewonnener Erkenntnisse, durchaus berechtigt, etwa folgendes, plausibel erscheinendes Bild der Situation zu zeichnen: Allgemein anerkannte Theorien über die Naturgeschichte des Menschen haben uns unseres Selbstverständnisses als besondere Geschöpfe Gottes beraubt und stattdessen in eine gemeinsame Evolutionsgeschichte mit Affen und Einzellern eingereiht. In diesem Sinne muss sich auch die letzte Bastion unserer Dignität gegenüber dem Rest der Natur, nämlich unsere Vernunftbegabung – nüchtern betrachtet – als Folge einer Tendenz des naturgeschichtlich-evolutionären Prozesses verstehen lassen, biologische Organismen

Keil und Schnädelbach gelangen sogar zu der Auffassung, dass der propagierte Naturalismus, im Sinne eines philosophischen Projektes, entweder offensichtlich falsch oder trivial ist (G. Keil/ H. Schnädelbach: ‚Naturalismus‘ (2000)). Keil und Schnädelbach beziehen sich in ihrer triftigen Argumentation für die offensichtliche Falschheit nicht-trivialer Positionen allerdings vorwiegend auf eine extreme Form des erkenntnistheoretischen Naturalismus, wie sie vor allem durch Quine vertreten wird.

bzw. sich selbst reproduzierende Systeme mit bestimmten überlebensrelevanten Eigenschaften, hervorzubringen. Denn soviel wir heute wissen, führte ein evolutionärer Prozess, unter jenen kontingenterweise gegebenen natürlichen Bedingungen auf der Erde, zu einer allmählichen Komplexitätssteigerung des inneren Aufbaus von bestimmten biologischen Systemen im Sinne einer Erhöhung ihrer internen, funktionalen Differenzierung und der damit einhergehenden Verbreiterung der Reaktionsvarianz auf Umwelteinflüsse. Solche Steigerungen der Komplexität entsprechen einer sich erweiternden Fähigkeit, Abhängigkeiten von kontingenten Umwelteinflüssen zu reduzieren, eine Entwicklung, die biologische Systeme schrittweise in die Lage versetzt, der unmittelbaren ‚Herrschaft‘ physikalischer oder chemischer Gesetzmäßigkeiten zu enttrinnen, indem sie (im Rahmen eben dieser Gesetze) höher stufige Mechanismen entwickeln, die eine Selektion reaktionsrelevanter Außeneinwirkung über die Unterscheidbarkeit von ‚Information‘ und ‚Hintergrundrauschen‘ ermöglichen⁷. In diesem Sinne müsste sich doch schließlich auch dasjenige, was wir mit „Geist“ (also das Vermögen zu Vernunft, Überzeugungen, Wissen, Wünschen usw.) oder „sprachlicher Bedeutungshaftigkeit“ meinen, als Resultat einer Entwicklung verstehen lassen, die, im Rahmen evolutionärer Gesetzmäßigkeiten und unter bestimmten Bedingungen, immer komplexere Möglichkeiten reaktiven Verhaltens von Organismen in der Auseinandersetzung mit ihrer Umwelt hervorbringt. Am Anfang stehen – in einem ‚technischen‘ Sinne – kognitiv sehr einfache Systeme, wie zum Beispiel die Zecke, deren reaktive Fähigkeiten mehr oder minder darauf beschränkt sind, sich bei der Wahrnehmung von Buttersäuregeruch von ihrem Standort zu lösen und sich an einem potentiellen Wirtstier festzubeißen. Am (bisherigen) Ende der Entwicklung steht der Mensch, der, unter Zuhilfenahme eines komplexen Kommunikationsmediums, arbeitsteilig Wissen produzieren kann und dadurch auf fortgeschrittenen kulturellen Entwicklungsniveaus sogar in der Lage ist, seine angestammte Biosphäre zeitweise zu verlassen und Ereignisse für eine ferne Zukunft mit erstaunlicher Präzision vorherzusagen. Realisierungen von „Geist“ sind demnach eigentlich nichts anderes als eine höher stufige, in Gehirnen und Nervensystemen materialisierte Form der Vermittlung zwischen spezifischen Formen sensorischer Stimuli (bzw. Information für eine Spezies oder einen Organismus) und einer für eine Spezies oder einen Organismus jeweils adäquaten (den Bedürfnissen entsprechenden) Reaktion. Auch menschlicher Geist

⁷ Für eine anschauliche, allgemeine Darstellung, s. H. von Ditfurth: *Im Anfang war der Wasserstoff* (1972); Für eine detailliertere Ausarbeitung, s. B.-O. Küppers: *Der Ursprung biologischer Information* (1986). Bekanntermaßen wird ein analoger, über ein Prinzip von informationsrelativer Innen-Außen-Differenzierung bestimmter Systembegriff in der Soziologie auch auf soziale Organisationsformen und Institutionen angewendet.

bestünde, so gesehen, lediglich in komplexen Formen von biologischen ‚Reiz-Reaktions-Vermittlungsmechanismen‘ in Gestalt von Gehirn und Nervensystem, welche sich durch eine ausgeprägte Steigerung der Fähigkeit zur Unterdrückung unmittelbaren Reaktionsverhaltens (zum Erreichen von nicht-unmittelbaren, aber möglicherweise ‚höherwertigen Zielen‘ und den damit verbundenen Vorteilen) auszeichnen und so eine enorme Steigerung von kognitiven Fähigkeiten im Sinne von Leistungen wie Schlussfolgern, Planen, Gedächtnis und Reflexion ermöglichen. Von einem Organismus, der äußere Einwirkungen in nicht-zufälliger, systematischer Form selektiert, um auf bestimmte Bedingungen in bestimmter Weise reagieren zu können, könnten wir schließlich auch sagen, dass er die Welt, bzw. den Teil oder Aspekt der Welt, der seine Umwelt oder ‚Wirkwelt‘ (von Uexküll) ausmacht, repräsentiert. Entsprechende (repräsentationale) psychische Zustände wären, so gesehen, nichts anderes als physisch komplex organisierte Wirkungen aufgrund ‚gewöhnlicher‘ physischer Ursachen und physisch komplex organisierte Ursachen für ‚gewöhnliche‘ physische Wirkungen.

Ich halte solche Formen von Überlegungen nicht nur für unproblematisch, sondern auch für erhellend, wenn sie als Erklärungsansatz dafür verstanden werden, wie sich kognitive Vermögen – am Ende sogar die Befähigung zur Konsumtion und Produktion sprachlicher Bedeutung – im Rahmen einer Naturgeschichte (im Sinne naturgesetzlicher Rahmenbedingungen) entwickeln konnten, ohne dass es dazu besonderer – etwa final-wirksamer – Kräfte oder des Plans eines göttlichen Schöpfers bedurft hätte (und auch die ‚Nichts-anderes-als-Redeweise‘ in diesem Sinne gedeutet wird)⁸. Es ist, in gewissem Sinne, eine durchaus erstaunliche Erkenntnis, dass es offenbar im Wesentlichen die aus einer sowohl phylogenetischen, als auch ontogenetischen Entwicklungsgeschichte hervorgehenden Eigenschaften des Gehirns und Nervensystems, also von Neuronen und ihre komplexen ‚Verdrahtungen‘ sind, welche die materielle Grundlage für das mentale Potential von menschlichen Individuen bilden. Die Probleme fangen da an, wo entsprechende Erklärungsversuche mit dem Anspruch auftreten, Geist und Bedeutung als solche erklären zu können, bzw. uns zu sagen, was Geist oder Bedeutung – in seinen jeweiligen Ausprägungsformen – (eigentlich) ist. Unter naturalistischen Voraussetzungen könnte dieser Anspruch nur eingelöst werden, wenn es gelänge – evolutionstheoretisch oder auf eine andere Weise

⁸ Allerdings stellen biologistisch-evolutionäre Erklärungsbemühungen im angedeuteten Sinne keine ‚harte‘, mechanistische, bzw. physikalistische Form von Naturalisierung dar, da sie am Ende, auf die ein oder andere Weise, auf eine teleologische bzw. normative Begrifflichkeit zurückgreifen. Wie wir noch sehen werden, können sie dennoch als naturalistisch eingestuft werden, da sie *Erklärungsweisen* darstellen, die sich auf natürliche Vorgänge als solche beziehen.

–, physische bzw. kausale Bedingungen für Instantiierungen spezifischer Bedeutungen, Typen sprachlicher Handlungen oder propositionaler Einstellungen anzugeben. Genau dies aber ist, wie ich in dieser Arbeit zeigen will, aussichtslos.

Im Sinne der oben bereits erwähnten Möglichkeit, jene allgemeinen Intuitionen, auf die sich Naturalisten mit einer gewissen Berechtigung berufen können, von solchen zu trennen, für die dies nicht ohne weiteres gilt, ist eine Unterscheidung zwischen einem ‚Dogmatischen Naturalismus‘ (kurz: DOGNA), als einer philosophisch (und nicht etwa naturwissenschaftlich) begründeten Einstellung einerseits und einer allgemeinen Form der Einstellung gegenüber unserer natürlichen Welt andererseits, die ich ‚Naturalismus des (modernen) nüchternen Common Sense‘ (kurz: NANÜCS) nennen will, hilfreich⁹. Ersterer bildet den ausschließlichen Gegenstand meiner kritischen Auseinandersetzung. Die unten folgende, genauere Charakterisierung dessen, was ich unter ‚DOGNA‘ verstehe, umfasst auch Positionen, die von manch einem vielleicht nicht ohne weiteres als ‚naturalistisch‘ eingestuft würden.

NANÜCS und DOGNA

Als grobe Kennzeichnung für den ‚*Naturalismus des nüchternen Common Sense*‘ lässt sich zunächst jenes, für eher nüchtern eingestellte und wissenschaftlich gebildete Menschen der ‚westlichen‘ Moderne mehr oder minder selbstverständlich gewordene Bild einer ‚entzauberten Welt‘ (Max Weber) anführen. Was wir darunter genauer zu verstehen haben, ist allerdings kontrovers. Ich möchte hier folgende, recht allgemein

⁹ Diese Unterscheidung entspricht – wie noch deutlich werden wird – in wichtigen Punkten nicht der Unterscheidung eines ‚*Manifest Image of Man*‘ von einem ‚*Scientific Image of Man*‘, die Sellars in seinem einflussreichen Aufsatz ‚*Philosophy and the Scientific Image of Man*‘ (Sellars (1960), in: ders. (1963)) vorgenommen hat. Sellars‘ interessante, aber eher cursorische Überlegungen zum Begriff von Person und Gemeinschaft am Ende des genannten Aufsatzes (die offensichtlich auch einen Ausgangspunkt der brandonschen Überlegungen (in: R. Brandom: *Making in Explicit* (1994)) darstellen) weisen allerdings in eine Richtung, die der von mir intendierten Argumentationsweise grundsätzlich entgegenkommt. Überlegungen zum philosophischen Aspekt des Problems jenes ‚*clashes*‘ unserer ‚*conflicting Images*‘, also unserer ursprünglichen Selbstverständnisse mit naturwissenschaftlich bestimmten, ‚entpersonalisierenden‘ Sichtweisen, finden sich auch bei Strawson (1985/1987)).

gehaltene, Charakterisierungsvorschläge machen, die sicherlich nicht nur deskriptive, sondern auch normative Elemente im Sinne eines *angemessenen* Selbstverständnisses enthalten.

Nach NANÜCS entspricht die materielle Welt zunächst – je nach Betrachtungsweise – der Gesamtheit von Dingen und Stoffen, Zuständen und Ereignissen oder Systemen (bestimmten besonderen Relationen von Dingen, Zuständen oder Ereignissen). Dinge, Stoffe oder Systeme haben Eigenschaften, verändern sich, entstehen und vergehen. Die phänomenalen Eigenschaften von sinnlich wahrnehmbaren Dingen, Stoffen oder Systemen werden durch nicht unmittelbar wahrnehmbare (mikrostrukturelle) Eigenschaften derjenigen Dinge oder Systeme hervorgerufen, aus denen sie bestehen (z. B. molekulare oder energetische Zustände)¹⁰. Die Eigenschaften von ding-, stoff- oder systemkonstituierenden Dingen oder Systemen wiederum können derart sein, dass wir sie sinnlich nicht unmittelbar wahrnehmen. Ob die Dinge und Systeme bleiben wie sie sind – ob sie als *Zustände* ‚verharren‘ – oder ob sie sich ändern – sich etwas *ereignet* –, hat Ursachen, die in der Natur ihrer Eigenschaften (bzw. ihrer makrostrukturellen Eigenschaften) und den Eigenschaften derjenigen Dinge und Systeme liegen, mit denen sie in Wirkverhältnissen stehen. Was auf der Ebene des Natürlichen geschieht, ist nicht in irgendeinem Sinne gewollt oder beabsichtigt und kann insofern weder richtig oder falsch beziehungsweise gut oder schlecht (‚böse‘) sein. Die Natur ist *kein verantwortbares Subjekt*. *Ihr* ist es *egal*, so möchte man im Sinne des NANÜCS ‚paradoxisch‘ sagen, was geschieht.

Ein Vertreter des NANÜCS sieht in teleologisch geprägten Weltbildern, in magischen Vorstellungen und Praktiken eine ‚personalisierende‘ Sicht der Natur, einen anthromorphisierenden Umgang mit Kontingenz, den Versuch einer wertenden Oktroyierung von Sinnhaftigkeit auf nicht verfügbares Sich-Ereignen, dem an sich weder ein Wollen, noch Werte oder Zwecke innewohnen. Er/Sie sieht vielleicht darüber hinaus den Grund für das Bestehen von magischen Vorstellungen oder religiösen Orientierungen vorrangig in ihrer *Funktion* der ‚Rationalisierung‘, also einer Möglichkeit begründeter Handlungsentscheidung angesichts einer gegebenenfalls geradezu lähmenden Unüberschaubarkeit möglicher Handlungsalternativen oder als Möglichkeit der Kontrolle unserer natürlichen Emotionen im Hinblick auf ein Gefühl des Ausgeliefertseins gegenüber anderen, der Natur oder einem letztlich unverfügbaren, persönlichen Schicksal (besonders der Unentrinnbarkeit der Endlichkeit aller

¹⁰ Eine allgemeine Auffassung, die Sellars als wesentliches Merkmal des ‚*scientific image of man*‘ ansieht (Sellars (1960)).

menschlichen Existenz)¹¹. Doch selbst solche Sichtweisen führen mit der Rede von Funktionen keine *Wertungen* des Geschehens als solchem, beziehungsweise *Projektionen* bezüglich umfassender oder zukünftiger Ziele oder Zwecke natürlicher Entwicklungen mit sich.

Der vom NANÜCS unterstellten Wert- und Zweckfreiheit aller Vorgänge in der natürlichen Welt lässt sich nur mit Beschreibungen, Erklärungen oder Theorien entsprechen, die sich *wertender Begriffe enthalten*¹². Zwar stützen wir uns auch alltäglich ‚immer schon‘ in vielfältiger Weise auf Einsichten in oder Vermutungen über generelle Zusammenhänge zwischen verschiedenen Typen von Zuständen oder Ereignissen in der Welt. Erst im Rahmen bestimmter kultureller Entwicklungen, sind wir aber dazu übergegangen, eine Vermehrung solcher Einsichten durch die *Institutionalisierung* von – zunächst in bestimmten Einzelfällen erfolgreichen – wissenschaftlichen Methoden systematisch voranzutreiben. Naturwissenschaftlicher Erkenntnisfortschritt besteht grundsätzlich in experimentell gestütztem Erwerb von Kenntnissen über die Mikroeigenschaften von Dingen oder Stoffen, die in Form von gesetzesartigen (konditionalen) Aussagen formuliert werden. Dadurch können bestimmte Ereignisse in einem verallgemeinernden Sinne erklärt und so auch mit größerer Genauigkeit vorhergesagt und schließlich kontrolliert herbeigeführt werden (häufig unter Zuhilfenahme technischer Entwicklungen, die durch solche Kenntnisse ermöglicht wurden). Wenn bestimmte Ereignisse mit Hilfe von Gesetzen *kausal* erklärt werden, dann sagt uns eine solche Erklärung, warum das Ereignis – unter den entsprechenden Umständen – durch ein bestimmtes anderes Ereignis oder durch bestimmte andere Ereignisse hervorgerufen werden *musste*¹³.

Wir haben sicherlich schon vortheoretisch eine gewisse Tendenz, auch uns selbst in einer *entpersonalisierenden Perspektive*, als in gewissem Sinne natürlich gesteuerte, beziehungsweise kausal determinierte Wesen mit typischen

¹¹ Eine weitere gesellschaftliche Funktion von Religion mag im Rahmen von NANÜCS als in der Legitimation von Herrschaft oder der Herstellung von Gemeinschaftsidentität bestehend gesehen werden.

¹² Diese Bemerkung ist allerdings nicht mit einem Postulat der (vermeintlichen) ‚Wertfreiheit der Wissenschaften‘ zu verwechseln, jedenfalls dann nicht, wenn dieses so verstanden wird, wie es häufig in entsprechenden Diskussionen der Fall zu sein scheint, nämlich als ob bei der *Entscheidung über die Gültigkeit oder Angemessenheit von Theorien oder Bestätigungsverfahren* im weitesten Sinne wertende (etwa normative oder ästhetische) Faktoren keine Rolle spielen dürften. Ich vermute, ein Großteil der einschlägigen Debatten lebt von dieser Verwechslung.

¹³ Wir müssen den NANÜCS für den hier intendierten Zweck einer allgemeinen Charakterisierung nicht auf eine Interpretation des Begriffs der Notwendigkeit von Naturgesetzen festlegen und können vorläufig auch Schwierigkeiten, wie den scheinbar teleologischen Charakter bestimmter Naturerklärungen oder den Status von Gesetzen mit *ceteris-paribus*-Klausel außen vorlassen.

Verhaltensmustern im Sinne *natürlicher Regelmäßigkeiten* zu verstehen. Aus dieser Perspektive geschieht alles, also auch das, was uns selbst und den sozialen Verkehr mit unseren Mitmenschen betrifft, mit natürlicher Notwendigkeit, aufgrund intrinsischer Eigenschaften der Welt¹⁴. NANÜCS lässt die Einnahme dieser entpersonalisierten (distanzierenden) Perspektive auf uns selbst allerdings nur unter eingeschränkten Bedingungen und bei bestimmten Gelegenheiten zu. Er gibt sich damit zufrieden, dass wir uns *sowohl* als Personen, *als auch* als Teil des Naturgeschehens (etwa als Tiere oder biologische oder soziale Systeme) verstehen können, ohne das eine gegen das andere ausspielen zu können. Eine entpersonalisierende Perspektive auf uns selbst zu verabsolutieren, scheint – so fühlt zumindest der prototypische Vertreter des NANÜCS – keine Möglichkeit zu sein, die uns wirklich zur Verfügung stünde¹⁵. Allerdings kann es Situationen geben, in denen wir in gewisser Weise unsicher darüber werden, inwieweit wir die entpersonalisierende Perspektive zum Tragen kommen lassen können oder sollen, wie etwa im Falle von Fragen zur Schuldfähigkeit von Personen.

Als Positionen, die ich dem *Dogmatischen Naturalismus* (DOGNA) zurechne, möchte ich zunächst allgemein solche verstehen, die (anders als der NANÜCS) voraussetzen, dass unsere soziale *Verstehens-* und *Handlungskompetenz* – in der ein oder anderen Weise – in (mehr oder minder erfolgreichen) *empirischen Theorien über kausale Zusammenhänge* besteht und auch ihrerseits einen Gegenstand solcher Theorien bildet. Dazu passt unter anderem auch, dass der DOGNA typischerweise alle Formen assertorischer Redeweise unterschiedslos als *Beschreibungen* von Zuständen oder Ereignissen deutet. Ob bestimmte Eigenschaften, die wir von Dingen oder Stoffen präzisieren, wirklich instantiiert sind (bzw. sein können), ob es also bestimmte Arten von Dingen und Stoffen, Zuständen oder Ereignisse *wirklich gibt*, wird davon abhängig gemacht, ob den entsprechenden Ereignissen eine *genuine* kausale Rolle zukommt. ‚Harte‘ (reduktionistische) Varianten des DOGNA glauben darüber hinaus, dass sich Verstehens- und Handlungskompetenz in einer physisch-kausalistischen Begrifflichkeit reformulieren bzw. erklären lassen muss.

In Bezug auf *Bedeutung* und *Geist* im hier gemeinten engeren Sinne lässt sich DOGNA nun über die Voraussetzung spezifizieren, dass alles (bzw. alles in einem respektablen Sinne Wissenswerte), was *common-sensisch* mit den Phänomenen

¹⁴ Die deterministische ‚Intuition‘ besagt, richtig verstanden, dass ein bestimmter Gesamtzustand der physischen (‚kausal relevanten‘) Welt einen zeitlich nachfolgenden Gesamtzustand in allen Hinsichten festlegt, was natürlich nicht bedeutet, dass wir prinzipiell zu vollkommenen Prognosen über zukünftige Zustände oder Ereignisse in der Welt in der Lage sein müssten.

¹⁵ In diesem Sinne etwa P. F. Strawson: ‚Freedom and Resentment‘ (1962).

sprachlicher Bedeutung und des Geistigen zu tun hat, unter dem Gesichtspunkt einer *empirischen Beschreibung oder Erklärung von Zuständen oder Ereignissen* innerhalb des kausalen Gefüges einer natürlichen Welt erfassbar sein muss, also Gegenstand einer Form von empirischer Psychologie bzw. empirischer ‚Bedeutungs-‘ oder ‚Handlungstheorie‘ ist¹⁶. Diese Form der ausschließlichen Orientierung am Faktischen, beziehungsweise Kausal-Relevanten, gilt übrigens nicht nur im Hinblick auf Fragestellungen der Philosophie des Geistes im engeren Sinne, sondern auch für *naturalistische Theorien der Erkenntnis*. Die Idee ist hier etwa, dass Fragen nach den Bedingungen von Erkenntnis als nicht auf eine normative Analyse idealer, sondern auf eine beschreibende Analyse *faktischer* Begründungspraxis der Wissenschaft abzielend verstanden werden sollten. – Oder es soll schließlich sogar ‚die‘ Wissenschaft selbst (etwa in Form einer empirischen Psychologie) sein, die uns sagt, wie Erkenntnistheorie auszusehen hat¹⁷. In eine ähnliche Richtung weisen auch Versuche, den Wissensbegriff im Rahmen einer Begrifflichkeit kausaler Bedingungen zu analysieren¹⁸. Die hier auftauchenden spezifischen Fragen bzw. Schwierigkeiten – etwa nach der Möglichkeit nicht-normativer Bestimmungen dessen, was als respektable Wissenschaft gelten kann – werden mich allerdings eher am Rande, im Zusammenhang mit kritischen Überlegungen zur Möglichkeit naturalistischer Theorien des Geistes im engeren Sinne, beschäftigen.

Eine wesentliche Komponente naturalistischen Selbstverständnisses, die sich aus dem bisher Gesagten eigentlich schon ergibt, ist der *Materialistische Monismus*, das heißt, die Ablehnung des kartesischen *Substanzdualismus* von *res cogitans* und *res extensa* im Sinne der Ablehnung einer eigenständigen *res cogitans*. Geist – oder besser Geistiges – kann nur dadurch Wirklich(es) sein, dass er (es) *kausale Relevanz* besitzt. Diese kausale Relevanz kann aber am Ende nur aufgrund bestimmter physischer (auch energetischer) Eigenschaften von Ereignissen oder Zuständen beziehungsweise auf der materiellen Ebene zustande kommen, da das *Prinzip der kausalen Geschlossenheit der physischen Welt* unterstellt wird, nach dem kausale Wechselwirkungen *nur* aufgrund physischer (konsequenterweise physikalischer) Eigenschaften möglich sind. Dies bedeutet nicht, dass für den Dogmatischen Naturalismus allgemein gilt, dass er *Beschreibungen oder Erklärungen* von Zuständen oder Ereignissen prinzipiell nur im Rahmen einer physischen Begrifflichkeit zulassen kann. Es heißt nur, dass davon

¹⁶ Wenn ich im Folgenden mitunter lediglich von „Naturalismus“ oder „naturalistischen Positionen“ rede, ist damit, wenn nicht anders vermerkt, DOGNA in Bezug auf Bedeutung und Geist gemeint.

¹⁷ Prominenter Vertreter dieser radikalisierten Form ‚naturalistischer Erkenntnistheorie‘ ist bekanntermaßen W. v. Quine (s. etwa: ders.: *Ontological Relativity and other Essays* (1969)).

¹⁸ Um nur ein Beispiel unter vielen zu nennen: A. Goldmann: *Epistemology and Cognition* (1986).

ausgegangen wird, dass *Kausalität* nur aufgrund der physischen Eigenschaften von Dingen, Ereignissen oder Zuständen gegeben ist.

Damit stellt sich natürlich die Frage nach der Art des Zusammenhangs zwischen physischen – bzw. am Ende physikalischen – und geistigen ‚Eigenschaften‘ (bzw. entsprechenden Typen von Dingen, Ereignissen oder Zuständen). Positionen innerhalb des DOGNA lassen sich durch die Art und Weise unterscheiden, in der sie diese Frage zu beantworten trachten. Dabei gibt es aber auch abweichende Auffassungen darüber, wie ernst man den ontologischen Status derjenigen Ereignisse oder Zustände nehmen sollte, die, nach verbreiteter Ansicht, in ‚intentionalen Kausalerklärungen‘ unserer *common-sensischen* Alltagspraxis zum Zuge kämen (bzw. wie ernst man entsprechende Formen kausaler Erklärungen nehmen sollte). Damit ist gemeint, dass unterschiedliche Varianten des DOGNA sich oft dadurch auszeichnen, dass sie hinsichtlich einer Antwort auf die Frage differieren, auf welcher Ebene der Beschreibung, also Individuierung und Systematisierung der Phänomene, man wirklich auf Fakten trifft.

Auch *sprachliche Bedeutungen* sind, nach naturalistischer Auffassung, am Ende nichts anderes, als – auf die ein oder andere Weise – *empirisch systematisierbare oder erklärbare* Phänomene, bestimmte kausale Eigenschaften ‚bedeutungsvoller Ereignisse‘ (typischerweise sprachliche Äußerungen), ‚Zustände‘ oder ‚Dinge‘ (Zeichen). Dies kann im Sinne der Ursachen und/oder Effekten von Sprachlichem (‚Sprachhandlungen‘) selbst und/oder den kausalen Eigenschaften derjenigen geistigen (‚psychologischen‘) Zustände oder Ereignisse, durch die Sprachliches überhaupt zu Bedeutungsvollem wird (wenn etwa ‚semantische Tatsachen‘ auf ‚psychologische Tatsachen‘ zurückgeführt werden), gemeint sein.

Schritte zu einer genaueren Bestimmung des Gegenstandes und der gemeinsamen Voraussetzungen naturalistischer Theorie

Nach diesen allgemein gehaltenen Charakterisierungen gilt es natürlich noch genauer zu bestimmen, was unter Versuchen einer Naturalisierung des Geistes und der Bedeutung im hier gemeinten Sinne zu verstehen ist. Dabei werde ich mich in der Folge schrittweise vom Abstrakten zum Konkreten vorarbeiten. – Zunächst zum Gegenstand: In den einschlägigen philosophischen Auseinandersetzungen (und nicht nur dort) werden üblicherweise auch sogenannte ‚subjektive‘ oder ‚qualitativ erfahrbare‘ Phänomene, wie im Sinne des ‚Erlebens-Wie‘ von Empfindungen (etwa Schmerzen) als „geistig“ oder „mental“ aufgefasst – meist unter der Rubrik ‚Bewusstsein‘. Mentale Phänomene in diesem Sinne werden mich hier jedoch – wie

gesagt – nicht beschäftigen. Sie fallen in einen deutlich anderen Bereich, auch wenn sie häufig in einem Atemzug mit jenen Phänomenen genannt werden, die ich hier als im engeren Sinne ‚geistig‘ verstehe. Auch zu sonstigen, eher alltagsüblichen Verständnissen von „psychisch“ oder „mental“, wie bestimmten Befindlichkeiten oder Charaktereigenschaften, habe ich hier nichts zu sagen. Um nun ein allgemeines und relativ unstrittiges Charakteristikum desjenigen anzuführen, was als der Gegenstand von Naturalisierungsbemühungen im hier in Frage stehenden engeren Sinne von „Geist“ gemeint ist, können wir sagen, dass dessen Beschreibung wesentlich eine Komponente enthalten muss, die in assertorischer Verwendungsweise zum Gegenstand einer Wahr-falsch-Beurteilung werden kann – also propositional strukturiert ist. Es handelt sich also um jenen Aspekt möglicher Verständnisse des Themas „Geist“, der unserer Praxis kognitiver (doxastischer oder volitiver) Zuschreibungen entspricht. Das genannte Charakteristikum gilt auch für ‚(sprachliche) Bedeutung‘, nämlich im Hinblick auf Beschreibungen ‚bedeutungsvoller Ereignisse‘ bzw. sprachlicher Handlungen.

Ich möchte schon hier anmerken, dass es mir ein interessantes Beispiel für die Sprachabhängigkeit bestimmter Sichtweisen zu sein scheint, wenn das Thema „Geist“ im angelsächsischen Sprachraum (vor allem jedoch von offenbar eher naturalistisch geneigten US-amerikanischen und australischen Philosophen) heute meist umstandslos in dieser Weise – ganz im Sinne des englischen „mind“ – auf individuenbezogene, propositionale Einstellungen eingeschränkt wird. In diesem Sinne gedeutet, scheint es eher naheliegend, sich auf Eigenschaften von Zuständen oder Ereignissen zu beschränken¹⁹. Formen überindividueller Verständnisse von Geist und Bedeutung von vornherein auszuschließen, kommt also naturalistischen Ambitionen offensichtlich entgegen. (Um den genannten Aspekt möglicher Verständnisse des Geistigen, also Geist im Sinne propositionaler Einstellungen, im Bedarfsfalle als solchen hervorheben zu können, werde ich ihn gegebenenfalls durch einen entsprechenden Index kennzeichnen und von „Geist^{pE}“ (pE = propositionale Einstellung) reden).

Der zur allgemeinen Bezeichnung der Eigentümlichkeit entsprechender Phänomene im Rahmen der neueren Debatten üblich gewordene, auf Brentanos Verwendungsweise zurückgehende, Oberbegriff ist der der ‚Intentionalität‘²⁰. Folgende

¹⁹ Dies gilt, wie sich noch zeigen wird, in ähnlicher Weise auch für das englische „meaning“, welches einerseits im Sinne von „Bedeutung“, andererseits jedoch zugleich im Sinne von „etwas meinen“ („to mean“) zu verstehen ist.

²⁰ Ich verwende den Begriff ‚Intentionalität‘ hier im Sinne der in der Philosophie des Geistes üblich gewordenen, an Brentano anschließenden, laxen Redeweise, mit der ursprünglich so etwas wie eine (Auf-Welt-) ‚Gerichtetheit‘ mentaler Zustände gemeint war, der inzwischen aber meist im Sinne von ‚Repräsentationalität‘, bzw. ‚Semantizität‘ verwendet wird. Da ich zum einen die ganze Rede von ‚Intentionalität‘ als *Eigenschaft* von Zuständen oder Ereignissen

Liste nennt wichtige Beispiele für das, was heute unter ‚intentionalen‘ Phänomenen verstanden werden soll:

Überzeugt sein, Glauben, Meinen, Intendieren (bzw. Beabsichtigen),
Erwarten, Sicher sein, Wissen, Erkennen, Wahrnehmen, Wünschen,
Hoffen, (Be-) Fürchten, Planen, Vorhaben.

Für alle diese Begriffe gilt, dass sie ‚intentionale‘ Einstellungsweisen zum Ausdruck bringen, die in der grammatischen ‚dass-Form‘ ergänzt bzw. durch einen propositionalen Gehalt spezifiziert werden können. Für einige dieser Begriffe gilt, dass sie auch durch andere grammatische Formen, die durch Wörter, wie „von“, „ob“ oder „zu“ und ein nachfolgendes Verb oder einen Objektausdruck ergänzt werden können – wobei eine solche Form der Ergänzung in einigen Fällen sogar die üblichere ist (etwa im Falle von „Vorhaben“ durch „zu“). Ähnliches gilt auch für Begriffe, die sprachliche Handlungen bezeichnen, wie etwa:

Mitteilen, Sagen, Auffordern, Befehlen, den Wunsch äußern, Versprechen,
Bitten, Ankündigen, Feststellen, Behaupten.

Ein übergreifendes Charakteristikum beider Listen ergibt sich nun insbesondere auch durch eine gemeinsame *semantische* Eigentümlichkeit derjenigen Formen von Sätzen, in denen solche Begriffe verwendet werden. Denn solche Sätze lassen sich nicht ohne weiteres *extensional*, also in einem ‚gegenstandstheoretischen‘ Sinne wahrheitsfunktional analysieren²¹. Eine extensionale Analyse geben zu können, bedeutet, dass sich der Wahrheitswert (die Wahr- oder Falschheit) von komplexen, also logisch nicht-elementaren Sätzen – gemäß bestimmten, wahrheitsfunktionalen Interpretationen der logischen Operatoren – rein *schematisch* aus den Wahrheitswerten logisch elementarer Sätze *herleiten* lassen können muss. Dabei sind Terme extensionsgleich, wenn sie in Sätzen untereinander austauschbar sind, ohne dass sich deren Wahrheitswert ändert. Logisch elementare Sätze, also solche, in denen keine

für zumindest missverständlich halte und zum anderen Verwechslungen mit der ursprünglichen Bedeutung des Terms „intentional“ im Sinne von absichtlich vermeiden will, bleibt mir nichts anderes übrig, als den Ausdruck nur im Zusammenhang mit Anführungszeichen zu verwenden.

²¹ Das ist gemeint, wenn man in diesem Zusammenhang auch von *intensionalen* oder opaken Kontexten oder Eigenschaften redet, die durch die entsprechenden Begriffe generiert werden. Unter ‚Intensionalität‘ versteht man meist den ‚kognitiven‘ Aspekt von Bedeutung, bzw. den ‚Sinn‘ oder die ‚Gegebenheitsweise der Referenz‘ für Sprachbenutzer oder Subjekte propositionaler Einstellungen (so haben, nach Freges berühmtem Beispiel, „Morgenstern“ und „Abendstern“ verschiedene Intensionen, aber die gleiche Extension, nämlich den Planeten Venus).

logischen Operatoren, wie beispielsweise „alle“, „und“, „wenn...dann“ vorkommen, sollen – nach den Standardverständnissen der ‚Wahrheitssemantik‘ – so zu verstehen sein, dass die Extension des Singularterms (des Satzsubjektes) demjenigen Gegenstand entspricht, auf den er referiert, während die Extension des generellen Terms (des Prädikates) der Gesamtheit oder der Klasse aller jener Gegenstände entsprechen soll, die durch ihre Instantiierung des entsprechenden Attributes den Satz wahr machen würden. Bekanntermaßen bereiten beispielsweise Sätze der Form „ – bedeutet –“, „ – glaubt, dass –“ oder „ – beabsichtigt, zu –“ für eine extensionale Analyse gewisse Schwierigkeiten. Ohne diese Schwierigkeiten hier im einzelnen diskutieren zu wollen, scheint für solche Satzformen jedenfalls zu gelten, dass es nicht ohne weiteres klar ist, wie der Wahrheitswert des ganzen Satzes (nur) von der Wahrheit seiner Teilsätze bzw. der Extension seiner Komponenten abhängen soll.

Nun mögen Naturalisten Nicht-Extensionales (bzw. ‚Intensionen‘) nicht besonders. Es passt nicht zu einer Weltsicht, nach der alles, was es gibt – was die Extension eines Ausdrucks bilden bzw. über das quantifiziert werden kann – klaren, empirisch operationalisierbaren Individuations- oder Gleichheitskriterien unterliegt. Um also eine unakzeptable ‚Ontologie von Bedeutungen‘²², die ohnehin keinen Beitrag zu einer ‚empirischen Theorie der Bedeutung‘ zu leisten vermag²³ zu vermeiden, versuchen Naturalisten die Satzform „ – bedeutet –“ so zu interpretieren, dass dabei entweder doch ein extensionales oder aber ein ‚naturalistisch akzeptables‘ intensionales (also z. B. einen Begriff ‚möglicher Welten‘ ins Spiel bringendes²⁴) Verständnis herauspringt. Die Alternativen bestehen demnach darin, entweder (wie etwa Quine) einen strikten Extensionalismus zu verfolgen und zu behaupten, dass die zusätzlichen Aspekte, die unser Verstehen solcher Sätze bedingen, nicht zu deren Semantik gehören (sondern zu deren ‚Pragmatik‘); oder zu versuchen, Intensionen eine ‚ontologisch unproblematische‘ Interpretation zu geben. Dies will ich an dieser Stelle nicht vertiefen. Ob man überhaupt von ‚Intensionen‘ reden sollte, bleibt unter Naturalisten letztlich strittig. Jedenfalls bevorzugt man grundsätzlich extensional interpretierte semantische Grundbegriffe, wie „Wahrheit“, „Referenz“ oder „Erfüllung“.

Auch die anderen beiden der genannten Satzformen sind für Extensionalisten problematisch, weil auch sie sich *prima facie* einer wahrheitsfunktionalen Analysierbarkeit im Sinne extensionaler Verfahren entziehen, deren Anwendbarkeit auf Sätze (Satzformen) eine Voraussetzung ihrer Kompatibilität mit dem Methodenkanon naturwissenschaftlicher Forschung ist. Damit würde, unter naturalistischen Voraussetzungen, aber auch die vermeintliche ‚Ontologie des Geistigen‘ fragwürdig.

²² W. v. Quine: *Wort und Gegenstand* (1960).

²³ D. Davidson: ‚Wahrheit und Bedeutung‘ (1967).

²⁴ Z. B. R. Stalnaker: *Inquiry* (1984).

Vertreter des DOGNA können dieser Unpässlichkeit auf zweierlei Weise begegnen: Entweder sie versuchen zu zeigen, dass solche ‚abweichenden‘ Sätze – *qua* einer Analyse, die ihre ‚eigentliche‘ logische Tiefenstruktur zu Tage fördern soll – als wahrheitsfunktional im extensionalen Sinne interpretierbar sind (in dem gezeigt wird, dass entsprechende Sätze über Dinge oder Ereignisse quantifizieren); oder sie disqualifizieren diese als eine Form von sprachlichen Residuen unwissenschaftlicher Alltagsmythologie – und verzichten damit auch auf die entsprechenden ontologischen Annahmen.

Nun lässt sich auch im Rahmen des NANÜCS *Kommunikation* grundsätzlich *auch* als eine *Verhaltensform* verstehen, die für Organismen, bzw. art-gleiche Gruppen von Organismen, von *Vorteil* ist. Für den *DOGNA* aber lässt sich nicht nur die *allgemeine Funktion* sprachlicher Kommunikation, sondern auch die *Signifikanz* sprachlicher Zeichen naturalistisch analysieren. Das naturalistische Grundmodell von Kommunikation geht davon aus, dass ein Sprecher (Sender) einen Hörer (Empfänger) mit sprachlichen (akustischen) Mitteln zu einer bestimmten Haltung oder zu bestimmten Reaktionen bringen (kausal veranlassen) kann (bzw. will). Dabei kommt, wie man auch sagt, eine Übertragung von *Information* zustande. Konsequente Formen einer naturalistischen Neubeschreibung sprachlicher Kommunikation versuchen diese lediglich auf der Basis des Austauschs von nicht-zufälligen, also strukturierten Geräusch- bzw. Lautfolgen zu analysieren, das heißt, nicht nur unter Absehung ihres Handlungscharakters, sondern auch ihrer semantischen Eigenschaften – ‚rein syntaktisch‘, wie man meint. Ziel ist es dabei, regelmäßige Zusammenhänge bezüglich der faktischen Verkettung zwischen hervorgebrachten Laut- oder Morphem- bzw. Äußerungstypen und zwischen diesen und nicht-sprachlichen Verhaltensformen in einem Sinne zu systematisieren, der einer Erfassung natürlicher Gesetzmäßigkeiten entspricht. Weniger restriktive Theorien scheuen sich dagegen nicht, bei der Beschreibung entsprechender Regelmäßigkeiten auch auf eine Begrifflichkeit zurückzugreifen, die uns aus semantischen Kontexten vertraut ist²⁵. Noch in einer anderen Hinsicht lassen sich Theorien, die man allgemein unter dem *Label* ‚Empirischer Sprachstrukturalismus‘ einsortieren kann, in eine restriktive, behavioristische (z. B. L. Bloomfield) und eine gemäßigte, ‚kognitivistische‘ (z. B. N. Chomsky) Variante unterscheiden. Letztere orientiert sich, anders als die behavioristische Variante, vor allem an ‚kognitiven Fähigkeiten‘ im Sinne von bestimmten Eigenschaften des Gehirns kompetenter Sprecher (und nicht an deren

²⁵ Dies scheint etwa dann der Fall zu sein, wenn man von ‚Wortverkettungen‘ redet, und dabei sowohl für die Beurteilung dessen, was ein Wort ist, als auch hinsichtlich der Beurteilung der Verkettungen als ‚erlaubt‘ oder nicht, offensichtlich auf semantische Vorverständnisse rekurren muss.

Verhaltensregelmäßigkeiten) und hält die jeweilige Befähigung zur Konsumtion und Produktion von Verkettungen sprachlicher Elemente für weitgehend angeboren.

Wissenschaftliche Theorien über empirisch vorfindbare sprachliche Kommunikationsformen, über die biologischen Voraussetzungen unserer Sprachfähigkeit, über die Funktionsweise von Gehirn und Nervensystem oder Theorien über analoge oder digitale Informationsverarbeitungssysteme müssen uns hier allerdings dann nicht weiter interessieren, wenn sie nicht den Anspruch erheben, reduktive oder eliminative Erklärungen (bzw. Bedingungen) für Bedeutung und Geist auf der Ebene des ‚Intentionalen‘, also von Begrifflichkeit und Propositionalität, zu liefern. Und dies scheint mir etwa in der *Neurologie* und den ihr nahestehenden ‚kognitiven‘ Theorien, wie etwa in jüngerer Zeit dem *Konnektionismus*, im Allgemeinen auch nicht (oder doch nicht notwendigerweise) der Fall zu sein. Deshalb gelten meine kritischen Bemühungen im allgemeinen nur solchen Theorien, die den *philosophisch motivierten* Anspruch erheben, die Themen „Geist“ und „Bedeutung“ *ausschließlich* in einer Weise angehen zu können, die alle entsprechenden Phänomene als Teil eines kausalen *Naturgeschehens* begreift – und sie gelten insbesondere für solche Positionen, die dabei auch noch unsere vorwissenschaftlichen Verständnisse von Geist und Bedeutung in *szientistischer* Weise *reformieren* wollen.

Bevor in der Folge auf wichtige Unterschiede innerhalb des von mir als „naturalistisch“ klassifizierten Lagers einzugehen sein wird, möchte ich dessen übergreifende, grundlegende Charakteristika zusammenfassend anhand von vier Punkten bestimmen. *Alle* Positionen, die ich zum DOGNA hinsichtlich Bedeutung und Geist rechne, teilen²⁶

- 1.) ein *exklusiv deskriptivistisches* Verständnis derjenigen sprachlichen Praktiken, für die die Begriffe in der oben genannten begrifflichen Liste konstitutiv sind: Die entsprechenden ‚intentionalen‘ Zuschreibungen oder Selbstzuschreibungen (von propositionalen Einstellungen oder Bedeutungen) werden *nur* als Formen der *Beschreibung* oder *Konstatierung* in Bezug auf *physische Dinge, bzw. Zustände oder Ereignisse* verstanden. [Ob es Dinge oder Zustände mit ‚intentionalen‘ Eigenschaften tatsächlich gibt, bzw. die entsprechenden Ereignisse tatsächlich vorkommen, ist allerdings strittig.]
- 2.) ein *kausalistisch-explanatives* Verständnis unserer Fähigkeit zur kompetenten Teilnahme an sprachlichen und nicht-sprachlichen Praktiken. Dabei wird besonders unser Verstehen von Handlungen als eine Fähigkeit gedeutet, ursächliche Erklärungen oder gar Vorhersagen von Verhalten anhand von wahren *empirischen*

²⁶ Die eckigen Klammern sollen jeweils eine Einschränkung des Allgemeinheitsanspruchs bedeuten.

Theorien über die ‚mentalen‘ bzw. ‚intentionalen Zustände‘ der jeweiligen Subjekte von Verhalten (von ‚Handlungen‘) geben zu können. [Als einer der Grundbegriffe entsprechender Theorien des ‚Verstehens‘ bzw. der Verhaltenserklärung wird üblicherweise ein ‚realistisch‘ gedeuteter Wahrheitsbegriff unterstellt, wobei die Spezifizierung/Individuierung ‚intentionaler Zustände‘ durch Angabe ihrer Wahrheitsbedingungen erfolgt.]

Die *meisten* Positionen, die ich zum DOGNA hinsichtlich Bedeutung und Geist rechne, teilen darüber hinaus

- 3.) ein *intentionalistisches* Sprachmodell: *Semantische Gehalte* von Äußerungen sind als derivativ zu den jeweiligen *individuellen intentionalen Einstellungen* von Sprechern zu verstehen. Das heißt, dass der Begriff der Wort- oder Satz oder Äußerungsbedeutung systematisch vom Begriff der (individuellen) Sprecherbedeutung und dieser wiederum systematisch vom ‚psychologischen‘ Begriff der *Sprechereinstellung* (z. B. Überzeugungen plus den Mitteilungsabsichten) abhängen soll. [Daraus wird häufig gefolgert, dass nur (individuelle) ‚mentale Zustände‘, nicht jedoch die Zeichen oder Zeichenhandlungen selbst intrinsisch ‚intentional‘, bzw. repräsentational oder bedeutungsvoll sein können. Für viele Naturalisten ist es zudem nicht wesentlich (konzeptuell notwendig) für sprachliche Verständigungspraktiken, dass Bedeutungen ‚geteilt‘ werden bzw. zwei Sprecher jemals in irgendeinem Sinne dieselbe Sprache sprechen.]
- 4.) ein *instrumentalistisches* Modell des ‚Handelns‘ bzw. Verhaltens: Die Rationalität von *Handlungen* beschränkt sich auf eine jeweilige Orientierung an individueller Nutzenmaximierung.

Diese Bestimmung des DOGNA ist so abstrakt gehalten, dass sie auch solche, in der gegenwärtigen *‚Meaning-and-Mind-Debatte‘* einflussreichen, zunächst aber eher unterschiedlich anmutenden, Positionen umfasst, wie etwa die Fodors und Davidsons, da beide zumindest die zwei ersten der genannten Voraussetzungen teilen. Ich hoffe zeigen zu können, dass es eine in den genannten Voraussetzungen zum Ausdruck kommende grundlegende Fehlkategorisierung von Bedeutung und Geist ist, aus der sich härtnäckige und vieldiskutierte Problemlagen dieser Debatte ergeben, wie sie sich etwa in Gegensätzen, wie *‚internalistisch vs. externalistisch‘* oder *‚realistisch vs. anti-realistisch‘* oder dem sogenannten ‚Problem des *Epiphänomenalismus*‘ (der ‚kausalen Überflüssigkeit‘ des Mentalen) widerspiegeln. Die Plausibilisierung dieser These muss natürlich mit einer kritischen Auseinandersetzung mit entsprechenden,

paradigmatischen Argumentationslinien des DOGNA einhergehen. Hier soll, zum Zwecke einer allgemeinen Übersicht, zunächst nur eine grobe Skizze der Grundgedanken unterschiedlicher, neuerer Theorieansätze gegeben werden, die ich diesem Projekt zuordne.

Grundvarianten des Naturalismus

Naturalistische Philosophien des *Geistes*^{PE} im hier gemeinten Sinne lassen sich zunächst in drei *Grundvarianten* unterscheiden, nämlich den *Reduktiven Materialismus*, den *Eliminativen Materialismus* und den *Interpretationismus*. Allerdings gibt es auch innerhalb dieser Varianten wichtige Unterschiede. So ist vor allem der Reduktive Materialismus noch einmal in *Typenidentitätstheorien* im *engeren* und im *weiteren Sinne* zu unterscheiden. Typenidentitätstheorien im engeren Sinne behaupten eine *unmittelbare* Identität von mentalen und physischen Eigenschaften, das heißt, ‚intentionale‘ Eigenschaften müssten in *eindeutiger* Weise auf physische reduzierbar sein. Dagegen postulieren Typenidentitätstheorien im weiteren Sinne lediglich eine Identifizierbarkeit von ‚intentionalen‘ mit *funktional bestimmten* physischen Eigenschaften. ‚Intentionale Eigenschaften‘, so behaupten solche Theorien, können in *multipler Weise* durch nicht-funktional bestimmte physische Eigenschaften realisiert sein – bzw. auf diese reduziert werden. Da Typenidentitätstheorien im engeren Sinne heute kaum noch eine Rolle spielen, werden sie mich nur am Rande beschäftigen.

Die heute einflussreichsten Varianten des Reduktiven Materialismus sind reduktive Theorien im weiteren Sinne und werden meist unter dem Oberbegriff „*Funktionalismus*“ subsumiert, wobei darunter allerdings, wie wir noch sehen werden, ziemlich Unterschiedliches verstanden wird. Zudem kommt es bei manchen der hier zu behandelnden Positionen nicht nur innerhalb der Grundvarianten, sondern auch zwischen diesen zu Überschneidungen, die die hier vorgenommenen – durchaus geläufigen – Unterscheidungen fragwürdig werden lassen können. Man sollte sie deshalb nicht *zu* ernst, sondern als ein Stück weit *ad hoc* nehmen, beziehungsweise als mehr oder weniger hilfreiche systematische Annäherungsmöglichkeit an Formen naturalistischer Theoriebildung.

Grundsätzlich geht der Reduktive Materialismus – im Gegensatz zum Eliminativen Materialismus – davon aus, dass es ‚*intentionale Tatsachen*‘ gibt und versucht, theoretische Strategien zu entwickeln, die zeigen sollen, wie diese als am Ende rein physischer Natur zu verstehen sein könnten. Er muss also Wege aufzeigen, physische

Bedingungen für Instantiierungen bestimmter ‚intentionaler Eigenschaften‘ der Form ‚...glaubt, dass p‘ oder ‚...kommt zu der Überzeugung, dass p‘ anzugeben. Der Interpretationismus lehnt dagegen jede Form der Identifizierbarkeit (vermeintlicher) ‚intentionaler‘ und physischer Eigenschaften ab – falls er überhaupt an erstere glaubt. Die materialistische Minimalbedingung beider Varianten ist dabei das sogenannte ‚*Supervenienzprinzip*‘, welches ungefähr besagt, dass jedem kausal relevanten Unterschied auf der Ebene des ‚Intentionalen‘ irgendeine Art des Unterschieds auf der Ebene des Physischen korrespondieren muss.

Der Reduktive Materialismus hat eine *behavioristische* Variante, die heute jedoch auch von den meisten Theoretikern innerhalb des naturalistischen Lagers als wenig erfolgversprechend beurteilt wird. Allerdings liegen die Dinge besonders hier nicht ganz so einfach, da es sehr verschiedene Formen von Behaviorismus gibt (was allerdings in den in Frage stehenden Debatten nicht immer Berücksichtigung findet). Die oben angesprochene Problematik möglicher Überschneidungen von Klassifikationen gilt dabei insbesondere auch für die unterschiedlichen Formen des Behaviorismus, vor allem, weil auch der Interpretationismus als eine – wenngleich schwache – Form von Behaviorismus gelten kann. Zudem ist eine Berücksichtigung entsprechender Unterschiedlichkeiten schon deshalb geboten, weil die Entwicklung jener naturalistischen Positionen, die den Hauptgegenstand meiner Kritik bilden werden, nicht zuletzt eine Folge der nachhaltigen Diskreditierung der bis *dato* vorherrschenden Varianten des Behaviorismus zu Anfang der 60‘er Jahre des letzten Jahrhunderts²⁷ war. Ich möchte deshalb hier kurz auf diese Unterschiede eingehen, und mich später nicht mehr um diejenigen naturalistischen Varianten des Behaviorismus kümmern, die unter Naturalisten selbst als erledigt gelten.

Der ursprüngliche, von Watson und Skinner vertretene ‚Psychologische Behaviorismus‘ ist eigentlich eine Variante (bzw. ein Vorläufer) des Eliminativen Materialismus, weshalb die Bezeichnung ‚*Eliminativer Behaviorismus*‘ angemessen erscheint. Eliminiert werden sollte vor allem die mit unserer ‚alltagspsychologischen‘ Zuschreibungspraxis (vermeintlich) implizierte ‚Ontologie‘ von Überzeugungen u.ä., vor allem, da solche ‚inneren psychologischen Zustände‘ und ‚Ereignisse‘, nach Ansicht der eliminativen Behaviouristen, nur einer subjektiv-introspektiven – und daher unwissenschaftlichen – Methode zugänglich und somit nicht objektiv existent seien. An ihre Stelle sollte eine objektiv gegebene ‚Verhaltensontologie‘, im Sinne von Körperbewegungen und akustischen Signalen treten²⁸. Auch die Position Quines lässt

²⁷ Als ‚*knock out*‘ gilt Chomskys berühmte Kritik an Skinner in ders.: ‚Review of B. F. Skinner’s *Verbal Behaviour*‘ (1959).

²⁸ Wie wir noch sehen werden, ist allerdings schon der dabei unterstellte Begriff von Verhalten problematisch, weil dabei Verhalten fälschlicherweise nicht als ein in bestimmter Weise kausal

sich dem Eliminativen Behaviorismus zurechnen, wenn auch seine Gründe gegen den Geist auf der Intensionalität (bzw. der ‚semantischen Undurchsichtigkeit‘) der entsprechenden Zuschreibungspraxis und deren *Nicht-Reduzierbarkeit* auf extensional Beschreibbares beruhen, weshalb er ‚intentionale‘ Zuschreibungen – wenngleich alltagspraktisch wohl nicht ersetzbar – als unwissenschaftlich ablehnt. So oder so ist jedenfalls das Hauptmotiv des Eliminativen Behaviorismus (wie eigentlich jeder Form des Behaviorismus) gegen den Geist, beziehungsweise für die Ablehnung ‚intentionaler Fakten‘, ein epistemisches: Das ‚intentionale Idiom‘ sei für objektive empirische Theorien der Verhaltensklärung oder -Prognose untauglich. Das Verhalten eines Organismus sollte – ohne unnötige Hypothesen über verborgene, obskure Phänomene – lediglich durch Annahmen über dessen natürliche Bedürfnisse, einige Prozesse oder Mechanismen behavioraler Modifikation (wie z.B. ‚Reizverstärkung‘) und durch die jeweils effektiven Reize erklärbar sein. Allerdings oszilliert sowohl Watsons, als auch Skinners Position zwischen der eliminativen Variante und einem *Reduktiv-Analytischen Behaviorismus*. Letzterer geht davon aus, dass sich alle Behauptungen, die ein ‚intentionales‘ Vokabular enthalten, als solche *analysieren* bzw. übersetzen lassen müssten, die nur Vokabeln einer physischen Verhaltensterminologie enthalten. Behavioristen dieses Schlages, wie etwa Hempel und Carnap, vertreten eine Kombination aus einer verifikationistischen Theorie der Bedeutung und dem Behaviorismus-typischen, evidentiellen Primat beobachtbaren Verhaltens.

Ryle und – mit Einschränkungen – vielleicht auch Wittgenstein könnte man dagegen als ‚*Nicht-Reduktive Logische Behavioristen*‘ einordnen. Obwohl beide die *kriteriale* Wichtigkeit des äußerlich Beobachtbaren von Geistigem betonen, propagieren sie weder eine Beschränkung auf nicht-intentional beschreibbares Verhalten (Körperbewegungen), noch sollen *einzelne* mentale bzw. ‚intentionale‘ Aussagen in Form von bestimmten Beschreibungen der sich in öffentlichem Verhalten, Gebaren oder Handeln potentiell manifestierenden entsprechenden Dispositionen *definierbar* sein. Besonders Wittgenstein, aber auch Ryle, ging es ja am Ende nicht darum, mentale Redeweisen in wissenschaftlich respektabler Weise behavioral zu reformulieren, sondern der Vorstellung des Geistes oder des Geistigen als eines verborgenen, inneren Mechanismus entgegenzutreten, eines Mechanismus, der mit den äußeren ‚Effekten‘ seines Wirkens lediglich in einem *kontingenten* (empirisch feststellbaren) Zusammenhang steht. Und beiden ging es, im Gegensatz zu den ‚empirischen‘ oder reduktiven Behavioristen, auch nicht darum, Geist im Sinne *empirischer* Theorien des Verhaltens zu analysieren – geschweige denn, Analysen von *spezifischen*

determinierter Prozess sondern als *Resultat* eines Prozesses aufgefasst wird; s. insbes. Dretske (1988).

propositionalen Einstellungen geben zu können, sondern um *begriffliche Erläuterungen*. Ohnehin können sie schon deshalb nicht als Naturalisten im hier gemeinten Sinne gelten, weil sie ja selbst Varianten des oben genannten Kategorienverstoß-Einwandes formulierten.

Die heute einflussreichste naturalistische Variante von Theorien, die wir in einem weiten Sinne noch als ‚behavioristisch‘ bezeichnen können, ist der *Interpretationismus*, wie er, wenngleich in unterschiedlicher Weise, vor allem durch die Positionen Dennetts und Davidsons repräsentiert wird. Nach interpretationistischer Auffassung wird der Sinn mentaler Zuschreibungen durch ein *holistisch* verstandenes ‚Interpretationsverfahren‘ bezüglich beobachtbaren Verhaltens bestimmt²⁹. Darauf werde ich gleich noch zurückkommen. An dieser Stelle gilt es zunächst nur festzuhalten, dass nicht nur der Eliminative Behaviorismus, sondern eigentlich alle reduktionistischen Varianten des Behaviorismus allgemein als gescheitert betrachtet werden, da sie mit ihren restriktiven Vorgaben in einen immer offensichtlicher werdenden Widerspruch zur tatsächlichen Praxis von Psychologie und Verhaltensforschung gerieten, vor allem wenn es um die angemessene Beschreibung oder Erklärung von Lernprozessen ging. Doch auch nicht-reduktive, aber gleichwohl *nicht-holistische* Varianten des Behaviorismus sahen sich mit einem schlagenden Einwand gegen die Möglichkeit *atomistischer Analysen mentaler Zuschreibungen* konfrontiert, ein Einwand, der zunächst von Geach³⁰ vorgebracht wurde.

Geach erinnerte einfach daran, dass das, was ein Subjekt propositionaler Einstellungen tut oder sagt, nicht nur von bestimmten Überzeugungen abhängt, sondern auch von seinen Wünschen oder Absichten und von anderen Überzeugungen. Üblicherweise hegen Personen unterschiedliche Wünsche mit variierender Stärke, die zudem untereinander konfliktieren können. Ob ein Wunsch oder eine Absicht handelnd umgesetzt wird, hängt dann nicht nur von bestimmten Überzeugungen, sondern auch von komplexen Präferenzordnungen ab. Das würde bedeuten, dass sich einzelne Überzeugungen, aber auch einzelne Wünsche nicht an bestimmte Handlungsweisen koppeln lassen. Dies aber bringt ein grundlegendes, methodologisches Problem mit sich: Die Möglichkeiten unterschiedlicher Anpassungen eines entsprechenden

²⁹ Die von mir bereits eingeräumte Unzulänglichkeit der hier vorgenommenen Einteilung zeigt sich besonders in der Möglichkeit, einerseits Quine in gewisser Hinsicht auch zu den Interpretationisten und andererseits Dennett in gewisser Hinsicht auch zu den Eliminativen Behavioristen rechnen zu können. Die hier favorisierte Einteilung hoffe ich aber durch die spätere Diskussion der jeweiligen Positionen plausibel machen zu können (sie macht sich besonders an Dennetts weniger restriktiven, nur noch in einem eingeschränkten Sinne ‚behavioristischen‘ Vorgaben im Sinne der Einnahme eines ‚Intentional Stance‘ fest).

³⁰ P. Geach: *Mental Acts* (1957).

personalen ‚Überzeugungs-Wunsch-Systems‘³¹ an gegebene ‚Handlungsereignisse‘, bzw. behaviorale Daten.

Diese Problemlage lässt sich gut an einem eindringlichen Beispiel von Gareth Evans³² veranschaulichen: Meine Überzeugung, dass eine bestimmte Substanz giftig ist, könnte sich in einer praktisch unendlichen Variation von Verhaltensweisen manifestieren. Ich könnte die Substanz vermeiden. Aber ich könnte auch meine Verwandten dazu bringen, sie zu vermeiden – oder, sie nicht zu vermeiden! Ich könnte eine sich täglich steigernde, geringe Menge des Mittels zu mir nehmen, in der Überzeugung, dass ich dadurch gegen das Gift immun werde, was unter entsprechenden Bedingungen durchaus vorteilhaft sein könnte (und ich das Eintreten dieser Bedingungen für wahrscheinlich oder doch zumindest für möglich halte). Ich könnte eine große Menge zu mir nehmen, um Selbstmord zu begehen – oder eine kleine, um mich vor einer peinlichen Verpflichtung zu drücken. Meine Überzeugung, dass es sich um Gift handelt, könnte praktisch zu jedem möglichen Verhalten führen – gegeben, ich variere genügend Komponenten im Rest meines ‚Überzeugungs-Wunsch-Systems‘. Genauso kann, umgekehrt, ein bestimmtes Verhalten durch erstaunlich viele, völlig unterschiedliche propositionale Zuschreibungen ‚rationalisiert‘ werden. Damit erscheint aber jeder Versuch, *einzelne* propositionale Einstellungen an *spezifische* Verhaltensweisen zu koppeln, von vornherein als aussichtslos. Alle diejenigen Varianten des DOGNA, mit denen ich mich in der Folge auseinandersetzen werde, berücksichtigen diesen Einwand, indem sie entweder holistisch oder nicht behavioristisch sind.

Nicht-behavioristische Theorien des DOGNA verstehen mentale Aussagen grundsätzlich als Aussagen über *körperinnere* Organisationsformen des Geistes im Sinne eines *Organs*. Auch hier können wir uns auf die heute favorisierten Versionen des Materialismus beschränken. So haben Typenidentitätstheorien im engeren Sinne kaum noch ernstzunehmende Fürsprecher. Die die Debatten in der ‚*Philosophy of Mind*‘ seit den siebziger Jahren dominierenden Formen des Reduktiven Materialismus sind unterschiedliche Varianten funktionalistischer Identitätstheorien, die jedoch, wie bereits angemerkt wurde, nur noch in einem eingeschränkten Sinne als ‚reduktionistisch‘ gelten können. In den letzten Jahren scheint in den Auseinandersetzungen um funktionalistische Theorien allerdings eine gewisse argumentative Erschöpfung eingetreten zu sein.

Der Niedergang sogenannter ‚psycho-physischer‘ Typenidentitätstheorien – also von Identitätstheorien im engeren Sinne – und der Aufstieg des Funktionalismus wurde

³¹ Was H.-D. Heckmann ein „individuelles, intentionales Profil“ nennt (in: ders.: *Mentales Leben und Materielle Welt* (1994)).

³² G. Evans: ‚Reply: Semantic Theory and tacit Knowledge‘ (1981).

interessanterweise praktisch zeitgleich mit der anti-behavioristischen, ‚kognitiven Wende‘ eingeläutet: Hilary Putnam hatte in ‚Minds and Machines‘ (1960) eine Art von Argument vorgebracht, wie es Husserl bereits um die Jahrhundertwende in ganz ähnlicher Weise gegen den *Psychologismus* geltend gemacht hatte. Husserl hatte in seinen *Logischen Untersuchungen* (1900) nachdrücklich auf einer Trennung von *Urteilsakt* und *Urteilsinhalt* bestanden. Seine Einwände gegen die zum Ende des 19. Jahrhunderts gängigen Versuche, die Logik als Gegenstand einer empirisch-psychologischen Forschung verständlich zu machen (bzw. zu naturalisieren), kann man, in gewisser Weise, als Vorläufer heute üblicher Argumentationslinien gegen Typenidentitätstheorien ansehen. Jedenfalls ist es mittlerweile unstrittig, dass es sehr unterschiedliche Möglichkeiten der materiellen Realisierung von geistigen Akktypen, wie etwa von bestimmten Rechenoperationen, von geistigen Widerfahrnistypen, wie bestimmten Wahrnehmungen oder von mentalen Zustands- oder Ereignistypen, wie Empfindungen, geben kann. So ist beispielsweise die *Hardware* von Rechenmaschinen, mit deren Hilfe wir Additionen ausführen können, in rein physischer Hinsicht natürlich ganz anders aufgebaut als das menschliche Gehirn, welches uns zu genau gleichen Operationen befähigt³³. Ebenso spricht nichts dafür, dass alle Wesen, denen wir Schmerzempfindungen zuschreiben, sich in, rein physisch gesehen, identischen Zuständen befinden (bzw. befinden müssen).

Typenidentitätstheorien im engeren Sinne spielen deshalb für mich hier nur noch die Rolle einer Kontrastfolie, vor allem für *funktionalistische* Theorien. Der Funktionalismus galt bis in die jüngste Zeit in weiten Kreisen der *cognitive-science-Community* als die aussichtsreichste Möglichkeit, unter der grundsätzlichen Prämisse des monistischen Materialismus mit dessen üblichen Schwierigkeiten fertig zu werden, indem er den behaupteten Zusammenhang zwischen dem Mentalen und dem Physischen sozusagen lockert und eher indirekt, über verhaltensrelevante *kausale Rollen* von propositionalen Einstellungen, herstellen will. Entsprechende funktionale Eigenschaften (bzw. Rollen) sollen durch ihre, in bestimmter Weise zu beschreibenden, kausalen Eigenschaften definierbar sein. Wie diese dann jeweils physisch realisiert sind (sein können), ist dann eine nachgeordnete Frage. Sich hier aufdrängende Anschlussfragen, etwa danach, auf welche Weise die Einführung funktionaler Rollen eine naturalistische ‚Reduktion‘ oder – vielleicht besser – *Integration* von ‚Intentionalität‘ überhaupt gewährleisten können soll, bzw. wie diese Rollen unter naturalistischen Prämissen zu beschreiben sein könnten, bergen jedoch grundlegende Schwierigkeiten, die zu diskutieren sein werden.

³³ Allerdings wird dieses Argument von modernen Naturalisten nicht, wie von Husserl, als Einwand im Sinne eines Kategorienverstoßes verstanden.

Der *Eliminative Materialismus*, wie er explizit vor allem von Paul Churchland vertreten wird³⁴, scheint wiederum von allgemeinen Schwierigkeiten naturalistischer Theorien der ‚Intentionalität‘ zu profitieren, die auch den Funktionalismus betreffen. Diese Schwierigkeiten werden darauf zurückgeführt, dass die sogenannte ‚Alltagspsychologie‘ („*folk psychology*“), also unsere Alltagspraxis propositionaler Zuschreibungen und rationaler Handlungserklärungen, deren Verwissenschaftlichung ja das Ziel funktionalistischer Theorien ist, in weiten Bereichen schlichtweg falsch oder hoffnungslos unpräzise sei und deswegen auch als Ausgangsbasis für wissenschaftliche Erklärungsweisen ungeeignet. ‚Alltagspsychologie‘ müsse deshalb, im Rahmen eines wissenschaftlichen Weltbildes, nicht reformuliert, präzisiert oder korrigiert, sondern eliminiert beziehungsweise ersetzt werden. Wichtig für meine systematische Kritik am Naturalismus ist im Hinblick auf diese Position eigentlich nur, dass auch sie das ‚Problem der Intentionalität‘ von vornherein als Frage nach der Angemessenheit und dem Wahrheitsgehalt von propositionalen Zuschreibungen versteht, die als empirische Aussagen über subjektinterne Zustände oder Ereignisse gedeutet werden. Da der (explizite) Eliminative Materialismus jedoch innerhalb des naturalistischen Lagers eher eine Außenseiterposition darstellt und auch gar nicht mehr den Anspruch erhebt, ein naturalistisches Verständnis von Bedeutung und Geist im Sinne von Propositionalität und Begrifflichkeit zu ermöglichen, wird er mich nur am Rande beschäftigen.

Interpretationistische Positionen, wie Davidsons oder Dennetts, gehen davon aus, dass eine Erklärung propositionaler Einstellungen in einer Erklärung unserer *Zuschreibungspraxis* solcher Einstellungen besteht. Was wir über propositionale Einstellungen sagen können, betrifft lediglich die Rolle, die sie im Rahmen einer ‚Rationalen Psychologie intentionaler Kausalerklärungen‘ von Verhalten nach dem Überzeugungs-Wunsch-Schema einnehmen. Nach Ansicht der Interpretationisten bestimmt und erklärt die ‚Psychologie‘ ihre Gegenstände auf eine ihr eigentümliche Weise, da ‚intentionale Eigenschaften‘ weder auf physische Eigenschaften reduziert, noch als unabhängig von einer ‚Interpretationspraxis‘ existierend verstanden werden können. Diese Eigentümlichkeit besteht in der notwendigen Inanspruchnahme methodologisch begründeter, *normativer Optimierungsannahmen*, wie vor allem einem *Holismus-Prinzip* und einem *Rationalitäts-Prinzip*. Für Davidson muss dabei, anders als für Dennett, die Zuschreibung eines ‚intentionalen Profils‘ Hand in Hand gehen mit einer Theorie der Bedeutung („Interpretation“) von Sprecheräußerungen.

Sowohl die Unterschiede, als auch die Gemeinsamkeiten der hier vorgestellten Varianten des Naturalismus lassen sich am besten über ihr jeweiliges Verhältnis zu jener schon erwähnten, grundsätzlichen Unterstellung einer ‚Alltagspsychologie‘ der

³⁴ P. A. Churchland: *A Neurocomputational Perspective* (1989).

Verhaltensklärung oder -prognose verdeutlichen. Die Form jenes Überzeugungs-Wunsch-Schemas ‚alltagspsychologischer‘ oder ‚intentionaler‘ Erklärungen, mit dem wir uns nach – nicht nur unter naturalistischen Philosophen – weit verbreiteter Auffassung in alltäglichen Situationen im Sinne einer *empirischen Theorie* der Verhaltensklärung und -prognose durchschlagen, entspricht in etwa dem Schema des *praktischen Syllogismus*:

X wünscht, dass q .

X glaubt, dass wenn p , dann führt X 's Handlung H zu q .

X glaubt, dass p .

: X wird H tun.

‚Alltagspsychologie‘ soll, neben der Kenntnis formal- und material-inferentieller Zusammenhänge, besonders auch in einem Verallgemeinerungswissen dieser Form im Sinne einer *Kenntnis von Naturgesetzen* bestehen. Die zugrundeliegende unstrittige Intuition besteht darin, dass wir aus der Tatsache, dass eine Person X eine Handlung H ausführt, bei einer Kenntnis von X 's handlungsleitenden *Wünschen* zumindest begründete Vermutungen über bestimmte seiner *Überzeugungen* anstellen können und umgekehrt. Kennen wir andererseits bestimmte Überzeugungen und Wünsche einer Person, so gibt uns dies brauchbare Hinweise auf mögliches oder gar wahrscheinliches Handeln in zukünftigen Situationen. Eine dem obigen Schema entsprechende nomologische Generalisierung hat dann die Form:

Jeder (jede, jedes) X , der (die, das) glaubt, dass p und der (die, das) wünscht, dass q und der (die, das) glaubt, dass wenn p wahr ist, H zu q führt, wird - *ceteris paribus* - H tun.

also etwa:

Jeder Mensch, der glaubt, dass Arsen giftig ist und der zu sterben wünscht und der glaubt, dass, wenn Arsen giftig ist, die rasche Einnahme einer größeren Menge Arsens dazu führen wird, dass er stirbt, wird – *ceteris paribus* – rasch eine größere Menge Arsen zu sich nehmen.

Wie gesagt, verstehen Naturalisten solche Schemata von Generalisierungen über propositionale Einstellungen als Schemata von Formen *kausaler Gesetzesaussagen*. Da Eliminative Materialisten der Ansicht sind, dass sogenannte ‚intentionale Prädikate‘

(wie etwa „---glaubt, dass p “) nicht wirklich kausalwirksame physische Eigenschaften individuieren, könnte man sie dementsprechend, gemäß einem zeitgenössischen philosophischen *Jargon*, auch als „Anti-Realisten“ (bezüglich des Geistes^{PE}) bezeichnen. Komplizierter liegen die Dinge im Hinblick auf den Interpretationismus. Dessen Hauptprotagonisten, Dennett und Davidson, vertreten nämlich explizit abweichende Meinungen hinsichtlich des ‚ontologischen Status‘ von Geist^{PE}. Während Davidson sich selbst als einen ‚Realisten‘ bezüglich ‚intentionaler Ereignisse‘ begreift³⁵, gilt Dennett als ‚Instrumentalist‘, für den die Zuschreibung propositionaler Einstellungen in Verhaltenserklärungen lediglich heuristisch-praktischen Wert besitzt, nicht aber ‚harten Fakten‘ korrespondiert³⁶. Für beide aber kann es kein ‚normales‘ gesetzesartiges (Reduktions-)Verhältnis zwischen ‚alltagspsychologischen‘ kausalen Generalisierungen und basaleren (allgemeineren) physischen Kausalgesetzen (bzw. den entsprechenden Vokabularen) geben, ein Verhältnis also, wie es etwa zwischen der Geologie und der Physik besteht. Allerdings teilen sowohl reduktive Realisten, die eine solche Kompatibilität, bzw. Reduzierbarkeit oder Integrierbarkeit der ‚Psychologie‘ unterstellen (wie etwa Fodor), als auch nicht-reduktive Realisten (wie Davidson) zumindest die Auffassung, dass semantisch spezifizierte propositionale Einstellungen tatsächlich kausalwirksame Ereignisse individuieren³⁷.

Ein weiteres wichtiges Unterscheidungsmerkmal naturalistischer Positionen liegt in ihrer jeweiligen Bestimmung des *Verhältnisses von Sprache (sprachlicher Bedeutung) und Geist^{PE}*. Dass ein solcher Zusammenhang überhaupt besteht, wird allgemein zunächst damit begründet, dass sprachliche Bedeutung (bzw. eine Theorie der Bedeutung) nicht als unabhängig von propositionalen Einstellungen (bzw. einer Theorie propositionaler Einstellungen) möglich verstanden werden könne. Strittig ist jedoch nicht nur die Art dieser Abhängigkeit, sondern auch, ob umgekehrt Geist^{PE} ohne sprachliche Kommunikation – bzw., ob eine Theorie über die propositionalen Einstellungen eines Subjektes ohne eine Theorie der Bedeutung seiner Sprache – überhaupt möglich wäre. Es bestehen grundsätzlich zwei Optionen, die der DOGNA hier anbietet. Diejenige Option, die glaubt, dass letzteres nicht der Fall ist, steht in einer systematischen Nähe zum Interpretationismus; wir können sie „*DAVIDSON-Programm*“ nennen. Die andere Option wird dagegen vom Reduktiven Materialismus

³⁵ Insbes. in: D. Davidson: ‚Unbestimmtheit und Anti-Realismus‘ (1998).

³⁶ Allerdings hat selbst der prototypische ‚Anti-Realist‘ Dennett in jüngerer Zeit seine ablehnende Haltung gegenüber einem realistischen Verständnis von Geist^{PE} in gewisser Weise relativiert (s. ders.: ‚Real Patterns‘ (1990)).

³⁷ Wobei sich Davidson auf konzeptuelle Überlegungen zu Fragen der Ontologie und Kausalität beruft (Ders.: *Handlung und Ereignis* (1980/1985)), Fodor hingegen vor allem auf den ‚prognostischen Erfolg‘ der ‚Alltagspsychologie‘ (s. etwa ders.: *Psychosemantics* (1987)). Mehr dazu später.

favorisiert; ich werde sie „*IBS-Programm*“ nennen (IBS = *Intention based Semantics*³⁸).

Das IBS-Programm unterstellt, dass es semantische Tatsachen (etwa der Form: Ein Ausdruck *A*, verwendet durch einen Sprecher *S* oder eine Gruppe von Sprechern *GS*, hat die Bedeutung *B*) und dass sich diese ‚Tatsachen‘ auf ‚psychologische Tatsachen‘ im Sinne der propositionalen Einstellungen und Mitteilungsabsichten eines Sprechers (Sprecher *S* verwendet einen Ausdruck *A* mit einer Intention *I* und einer entsprechenden Überzeugung *Ü*) zurückführen lassen. ‚Psychologische Tatsachen‘ im Sinne propositionaler Einstellungen (etwa der Form ‚*S*^x hat *Ü*^p‘) sind, nach dem IBS-Programm, prinzipiell unabhängig von ‚semantischen Tatsachen‘ möglich. Wenn sich nun ‚psychologische Tatsachen‘ ohne Rückgriff auf ihre semantischen Eigenschaften (den begrifflichen, bzw. propositionalen Gehalt) beschreiben ließen, hätte man schließlich *eine naturalistische Reduktion (Integration) von Bedeutung(en)* (‚Bedeutungstatsachen‘) *qua einer naturalistischen Reduktion (Integration) von Geist*^{pe}. Den entsprechenden sprachlichen und geistigen Phänomenen ließe sich ein Platz innerhalb der ‚normalen‘ materiellen Ding-, Ereignis- oder Zustandsontologie der natürlichen Welt zuordnen, womit sie ‚quantitativen‘ Verfahren einer wissenschaftlichen Erforschung offen stünden.

Dagegen geht das DAVIDSON-Programm – wie bereits angedeutet wurde – davon aus, dass sich sprachliche Bedeutungen und propositionale Einstellungen nicht unabhängig voneinander erfassen lassen und überdies, aus methodologischen Gründen, von semantischen ‚Fakten‘, wenn überhaupt, nur in einem eingeschränkten Sinne die Rede sein kann. Die ‚harten‘ behavioralen Fakten seien nämlich aus einer methodischen Perspektive ‚*Radikaler Interpretation*‘, bei der weder die propositionalen Einstellungen, noch die Bedeutung der Sprache der zu ‚interpretierenden‘ Sprecher als bekannt vorausgesetzt werden dürfen, für eindeutige Festlegungen auf bestimmte Äußerungsbedeutungen (Referenzen) unzureichend³⁹. Sowohl die Bedeutung der Äußerungen von Sprechern, als auch ihre Einstellungen seien nur mittelbar und in wechselseitiger Abhängigkeit zugänglich. Der Begriff der Bedeutung lässt sich nach Davidson also nicht einseitig auf psychologische Begriffe wie ‚Überzeugung‘, ‚Wunsch‘ u.ä. zurückführen, weil vor allem komplexe Überzeugungen nur über die Kenntnis eines Mindestmaßes an Äußerungsbedeutungen zugänglich seien (was

³⁸ Diesen Ausdruck übernehme ich von S. Schiffer (in ders.: *Remnants of Meaning* (1987)). Für eine entsprechende Unterscheidung der beiden Optionen vergl. auch B. Loar: *Mind and Meaning* (1981). Der moderne *Locus Classicus* einer intentionalistischen Reduzierbarkeit von Semantik ist natürlich P. Grice‘ Aufsatz ‚*Meaning*‘ (1957).

³⁹ S. etwa Davidson: ‚*Radikale Interpretation*‘ (1973) und ders.: ‚*Die Unerforschlichkeit der Bezugnahme*‘ (1979).

natürlich zur Folge hat, dass wir Tieren – wenn überhaupt – zumindest solche Überzeugungen nicht begründet zuschreiben könnten).

Allerdings sind die Verhältnisse hier insofern nicht ganz eindeutig, als sich nicht alle Theoretiker, die ich (in einem weiten Sinne) zu den Reduktiven Materialisten zähle, als IBS'ler verstehen. So wendet sich etwa Ruth Millikan sogar explizit dagegen. Andererseits ist Davidson zwar kein Reduktionist, er vertritt aber ein bedeutungsintentionalistisches Grundmodell der Verständigung. Bevor wir auf das IBS-Programm und die Positionen des Reduktiven Materialismus näher eingehen, sollen noch einige systematische Überlegungen angestellt werden, die die Grundlagen für meine kritische Auseinandersetzung mit naturalistischen Positionen liefern sollen.

2. Zur kategorialen Differenz von Natur und Sinn

Kategorienfehler – Die Ausgangslage

Argumente des Kategorienfehlertyps, wie ich sie gegen den Naturalismus ins Feld führen will, sind explizit bekanntlich zunächst durch Ryle und – von Wittgenstein inspiriert – im Rahmen der sogenannten ‚*ordinary language philosophy*‘ hervorgebracht worden. Sie können ihrerseits als sprachphilosophisch-deontologisierende Variante des Kantischen Aspekt dualismus verstanden werden, der uns als Wesen zweier Welten, einer (*phänomenalen*) Welt der Erscheinungen und einer (*noumenalen*) Welt des Intelligiblen, sieht. Intuitive Grundlage dieser Zweiteilung ist die Unterscheidung von Freiheit (des Willens) und (kausaler Natur-) Notwendigkeit, der Möglichkeit der *Anerkennung* des bindenden Charakters einer Norm oder Regel gegenüber dem ‚blinden Zwang‘ der Naturgesetze. Auch Frege und Husserl lassen sich zweifellos als klassische Vertreter einer modernen Szientismus- bzw. Psychologismuskritik anführen, die sich, in der ein oder anderen Weise, auf eine

kategoriale – und damit begrifflich nicht vermittelbare – Differenz zwischen ‚logischem‘ und ‚kausalem Müssen‘ beruft.

In einem sowohl thematisch, als auch zeitlich engen Zusammenhang mit der Diskussion um Kategorienfehler steht die – zwischen Fragen zur Methodologie der Geistes- und Sozialwissenschaften und Fragen der Handlungstheorie angesiedelte – Debatte um *Ursachen versus Gründe* innerhalb der analytischen Philosophie⁴⁰. Im deutschsprachigen Raum lief eine in mancher Hinsicht ähnlich gelagerte Diskussion zunächst als ‚Positivismusstreit‘, später eher unter dem *Label* ‚Erklären-Verstehen-Kontroverse‘. Vorläufer hier ist bekanntermaßen jene Ende des 19. Jahrhunderts besonders durch Dilthey initiierte Auseinandersetzung um die Eigenständigkeit einer spezifischen, hermeneutischen Methode zur Erfassung der ‚sinnstrukturierten Teile unserer Wirklichkeit‘.

Argumente des Kategorienfehler-Typs werden häufig mit einer *sprachdualistischen* Position gleichgesetzt. Der Begriff ‚sprachdualistisch‘ scheint so etwas, wie eine klare Unterscheidbarkeit von ‚Diskursuniversen‘, etwa über Mentales und über Physisches, anhand eindeutiger *lexikalischer* Kriterien nahezulegen. Ryle etwa versucht seinen Vorschlag durch einen Substitutionstest für Begriffe unterschiedlicher Kategorien zu operationalisieren. So kann man beispielsweise in Sätze der Form: „– löste die Lawine aus“ oder „– kommt morgen früh hierher“ als Subjekt „der Sturm“ oder Personennamen einsetzen, nicht aber „der Schmerz“ oder „seine Überzeugung“. Nicht-Substituierbarkeit als allgemein verstandenes syntaktisches Kriterium führt aber, wie sich schnell zeigt, zu unzähligen Fällen von Unterscheidungen (‚Kategorien‘), die mit den von Ryle intendierten gar nichts zu tun haben. Zudem entstehen Abgrenzungsprobleme zwischen – *qua* Kategorienverstoß – *unsinnigen* Sätzen und solchen, die schlicht *empirisch falsch* sind (eine Differenz, die viele Naturalisten ohnehin nivellieren wollen). Vor allem jedoch scheint es häufig vom Kontext einer Äußerung abzuhängen, ob sie ‚durchgeht‘. Für Kategorienfehler-Argumente gilt offenbar, was auch für Versuche der Formulierung von *Sinnkriterien* gilt: Wenn rein formale Kriterien nicht zur Verfügung stehen, bleibt nur der Weg des *argumentativen Aufweises* eines in einem bestimmten Kontext unsinnigen

⁴⁰ S. etwa: G. E. M. Anscombe: *Intention* (1957/1963); P. Winch: *Die Idee der Sozialwissenschaft und ihr Verhältnis zur Philosophie* (1958); A. J. P. Kenny: *Action, Emotion and Will* (1963); D. Davidson: ‚Handlungen, Gründe und Ursachen‘ (1963); C. G. Hempel: *Aspekte wissenschaftlicher Erklärung* (1965/1977); P. M. Churchland: ‚Der logische Status von Handlungserklärungen‘ (1970/1977), A. Goldman: *A Theory of Human Action* (1970), G. H. von Wright: *Erklären und Verstehen* (1971/1974).

Die Bezeichnung „Ursachen *versus* Gründe“ ist allerdings insofern irreführend, als die Vertreter einer *kausalen* Handlungstheorie ja gerade darauf bestanden haben, dass Gründe Ursachen sein könnten.

Sprachgebrauchs durch möglicherweise komplizierte Präsuppositions- und Implikationsanalysen.

Die Unangemessenheit lexikalischer Kriterien zeigt sich insbesondere auch an der durchdringenden *Anthromorphizität* unserer *umgangssprachlich-metaphorischen Redeweisen über die Natur* einerseits („Das Wetter ist freundlich“, „Der Berg war ohne Gnade“, „Der Fluss sucht sich seinen Weg zum Meer“) und einer nicht minder tief verwurzelten *Physiomorphizität* entsprechender *Redeweisen über Personen und Mentales* andererseits („Sie überkam eine dunkle Vorahnung“, „Er hat ein dickes Fell“), bzw. Geistiges („Sie zerbrach sich den Kopf“, „Der Gedanke kreiste ihm im Kopf herum“)⁴¹. Es dürfte kaum weiterhelfen, auf den bloß metaphorischen Charakter solcher Redensarten zu verweisen. Denn erstens scheint es mir fraglich, ob sich überhaupt eine für sprachkritische Zwecke brauchbare Trennungslinie zwischen metaphorischem und nicht-metaphorischem Sprachgebrauch ausmachen lässt. Und zweitens ist auch nicht ganz klar, ob und wieso man metaphorische Sprachgebräuche *per se* als in irgendeiner Weise sekundär oder ‚uneigentlich‘ qualifizieren sollte. Dies gilt keineswegs nur im Hinblick auf künstlerisch-literarische Kreativität. So könnten sich Naturalisten wohl mit Recht gerade auch auf die Geschichte der modernen Naturwissenschaften berufen, um intuitiv vorliegende Kategorienverstöße durch die Behauptung *einer möglichen heuristischen Fruchtbarkeit* von spezifischen, Kategoriengrenzen in produktiver Weise überschreitenden Metapherngebräuchen zu legitimieren. Auch hier gilt, dass über die Angemessenheit oder Unangemessenheit einer Ausdrucksverwendung wohl nur argumentativ, unter Berücksichtigung unterschiedlicher Aspekte von Angemessenheit und Sinnhaftigkeit, entschieden werden kann.

Eine prominenter Spezialfall ‚sprachdualistischer‘, beziehungsweise sich auf Kategoriendifferenzen berufenden Typs von Argumenten besteht in dem gegen eine Interpretation von Handlungserklärungen (im Sinne einer Antwort auf die Frage: „Warum hat X H getan?“) als Form der Kausalerklärungen gerichteten Einwand, Handlungen könnten nicht im üblichen Sinne ursächlich erklärt werden, weil Handlungsabsichten (Intentionen) oder Gründe einfach keine *Ursachen* von Handlungen abgäben. Ursachen müssten, so ein der Argument, als von ihrer behaupteten Wirkung (hier: der ‚äußeren‘ Handlung) logisch unabhängig beschreibbar sein, was aber bei Handlungsabsichten oder Gründen gerade nicht der Fall sei. Der Zusammenhang zwischen Handlungsabsichten oder Gründen für eine Handlung und dem Ausführen der Handlung sei daher kein kausaler, gesetzesartig zu beschreibender, sondern ein logischer (bzw. begrifflicher). Und Handlungen als (durch ‚innere‘

⁴¹ Hierzu ausführlich: Keil (1993), Kap. III.

Ereignisse) *verursachte Ereignisse* zu denken, liefe zudem der Möglichkeit zuwider, sie auch zu *unterlassen*, was eine ziemlich unstrittige *begriffliche Voraussetzung* unseres Handlungsverständnisses im Sinne *freier*, absichtlicher Handlungen ist.

Offenbar gelten jedoch diese oder ähnliche Argumente des genannten Typs innerhalb des gegenwärtigen *Mainstreams* analytisch orientierter Philosophen als weitgehend entkräftet, wenn nicht gar widerlegt⁴² – zu Unrecht, wie ich denke. Wesentliche Grundlage dieser Einschätzung ist zum einen die Ansicht, dass Quine mit seiner berühmten Kritik an der Unterscheidung zwischen analytischen und synthetischen Sätzen die Möglichkeit einer konsequenten Trennbarkeit konzeptueller und empirischer Fragen desavouiert und damit der Methode begrifflicher Analyse, auf die sich Kategorienfehlerargumente stützen, die Grundlage entzogen hat⁴³. Zum anderen habe Davidson überzeugend für ein Verständnis von Gründen als Ursachen argumentiert, dass die vermeintliche dichotomische Entgegensetzung von Kausalität und Rationalität aufhebt⁴⁴. Quines Argument beruht auf gewissen methodologischen und bedeutungstheoretischen Annahmen, auf die später genauer einzugehen sein wird. Ich möchte dazu hier nur soviel sagen, dass selbst dann, wenn es nicht gelingen sollte, Kriterien für eine eindeutige Abgrenzbarkeit analytischer und synthetischer Sätze anzugeben, daraus noch nicht folgt, dass diese Unterscheidung sinnlos ist, bzw. keinen argumentativen Stellenwert hat. Bezüglich des zweiten Punktes scheinen Davidson und andere Vertreter der *kausaler Handlungstheorien* vor allem mit zwei einleuchtenden Hinweisen punkten zu können: Einerseits stehen Handlungen ja auf jeden Fall in kausalen Zusammenhängen mit physischen Ereignissen, insofern sie unzweifelhaft Wirkungen in der physischen Welt haben können. Und andererseits ist es eine Sache, einen noch so guten Grund für eine Handlung zu haben, eine andere aber, die Handlung *aus* einem Grund auch tatsächlich auszuführen. Theorien, die die Ursächlichkeit von Gründen bestreiten, scheinen diese Differenz nicht erklären, bzw. nichts dazu sagen zu können, wie ein guter Grund nicht nur ein möglicher, sondern auch ein tatsächlich handlungsleitender sein kann.

Zeitgenössische Vertreter kausaler Handlungstheorien sehen demnach rationale Handlungserklärungen nach dem Schema des praktischen Syllogismus selbst als eine Form kausaler Erklärung an. Vor allem Davidson hat darüber hinaus allerdings die (einflussreiche) These ins Spiel gebracht, dass sich solche Erklärungen nicht auf strikte Gesetze stützen können, was ihrem kausalen Charakter allerdings keinen Abbruch

⁴² Stellvertretend für den deutschen Sprachraum, s. etwa: A. Beckermann: ‚Handeln und Handlungserklärungen‘ (1977).

⁴³ W. v. Quine: ‚Zwei Dogmen des Empirismus‘ (1951/1979)

⁴⁴ Zuerst in Davidson (1963). Ähnlich aber auch Goldman (1970).

täte⁴⁵. Damit aber hat sich der Charakter der Auseinandersetzung zwischen Vertretern eines ‚Monopolanspruchs‘ kausaler Erklärungsweisen und ihren Gegnern eindeutig verschoben: Argumente für die Inkompatibilität bestimmter begrifflicher Sphären, also der Unvereinbarkeit von ‚Personen-Handlungs-Begründungs-Sprachspielen‘ mit ‚Ding-Ereignis-Ursache-Sprachspielen‘⁴⁶ lassen sich nicht mehr allein dadurch gewinnen, dass man auf die Unzulänglichkeit des ‚subsumtionstheoretischen‘ Hempel-Oppenheim-Schemas gesetzestützter Kausalerklärungen in Bezug auf Handlungen, bzw. sozialwissenschaftliche Kontexte verweist. Ohnehin scheint die beabsichtigte Differenzierung durch neuere Debatten um ein angemessenes Verständnis der ‚Logik‘ wissenschaftlichen Erklärens, beziehungsweise des Bestehens oder Nicht-Bestehens einer grundsätzlichen methodischen Differenz zwischen naturwissenschaftlichen und – wie man heute sagt – ‚kulturwissenschaftlichen‘ Disziplinen, an Prägnanz einzubüßen. So wird etwa von einigen Autoren der Zusammenhang von Kausalität und (striker) Gesetzesartigkeit in Frage gestellt⁴⁷ oder gar die Ansicht vertreten, der Kausalbegriff sei überhaupt nur *durch* den Handlungsbegriff zu analysieren⁴⁸.

Vergegenwärtigen wir uns also zunächst noch einmal den intuitiven Ausgangspunkt, der die Grundlage des von mir behaupteten, tiefgreifenden Unterschieds zwischen Begriffen oder Kategorien, wie „Bedeutung“, „Begriff“, „Aussage“ (bzw. „Proposition“ – im Sinne von Aussage- oder Urteilsgehalt), „Begründung“, „Geist“, „Erkennen“, „Überzeugung“, „Wissen“, „Wahrheit“, „Person“ und „Handlung“ auf der einen Seite und Kategorien oder Begriffen, wie „Ding“, „Kausalität“, „Naturgesetz“, „Zustand“, „Ereignis“, „System“ oder „Mikroeigenschaft“ auf der anderen bildet. Ich hatte oben bereits – etwas vage – vorgeschlagen, die Differenz daran festzumachen, dass mögliche Redeweisen, die sich im Umkreis von Begriffen oder Kategorien der erstgenannten Liste bewegen, irgendwie in Zusammenhang mit Formen einer Richtig-Falsch-Beurteilbarkeit stehen, was nicht für Redeweisen gilt, die sich im Umkreis von Begriffen oder Kategorien bewegen, die der anderen Liste angehören. Etwas genauer, können wir dieses Kriterium so verstehen, dass der primäre Sinn von Redeweisen, die durch Begriffe oder Kategorien der ersten Liste bestimmt sind, an Kontexte möglicher *Wahr-Falsch-*, bzw. (unterschiedlichen Formen) der *Vernünftig-Unvernünftig-* oder *Gut-Böse-Beurteilungen* rückgebunden ist. Diese Bestimmung als für den Moment akzeptabel unterstellend, scheint jedenfalls klar, dass der Anspruch naturalistischer Theorien darin bestehen muss, eine *kategoriale Inkompatibilität* der genannten Listen –

⁴⁵ D. Davidson: ‚Mentale Ereignisse‘ (1970).

⁴⁶ Diese Redeweise geht wohl auf Habermas (1971) zurück. Habermas redet allerdings nur von „Personen-Handlungs-Sprachspielen“ und von „Ding-Ereignis-Sprachspielen“.

⁴⁷ Hierzu jüngst etwa G. Keil: *Handeln und Verursachen* (2000)

⁴⁸ G. H. von Wright (1971 /1974), Kap. II.

zumindest für bestimmte Begriffspaare – zu *negieren* und Vorschläge für eine (Neu-) Beschreibung der mit der ersten Liste in Zusammenhang stehenden Phänomenen zu liefern, die eine Form von ‚Nichts-anderes-als-These‘ plausibel erscheinen lässt.

Ich ziehe es vor, von unterschiedlichen kategorialen *Rahmen* zu sprechen, denen man die beiden Listen zuordnen muss, weil man natürlich auch unterhalb der Ebene jener dualen Unterscheidung, die ich im Auge habe, von Kategorien bzw. Kategorienunterschieden reden kann. Ich bin also nicht – wie der etwas unglückliche Begriff „Sprachdualismus“ nahelegt – darauf festgenagelt, von lediglich *zwei*, in irgendeinem Sinne miteinander unverträglichen, begrifflichen *Kategorien* auszugehen. Dies liefe auf eine eher willkürliche Einschränkung des Kategorienbegriffs hinaus.

Es dürfte auch klar sein, dass wir uns zur Begründung einer kategorialen Unvereinbarkeit zwischen bereits ‚vorphilosophisch‘ mehr oder minder etablierten Begriffen, wie „Bedeutung“ und „Überzeugung“ mit Begriffen, wie „Ereignis“ und „Zustand“ nicht mit der schlichten Erinnerung an pragmatistische Platitüden, wie „Reden ist Handeln“ oder der Behauptung, dass sich Unterschiede ‚im Geist‘ (bzw. im Denken) letztendlich in Unterschieden im Handeln zeigen müssen, begnügen können. Solche Feststellungen werden ja von den meisten zeitgenössischen Vertretern kausaler Handlungstheorien, die ich dem naturalistischen Lager zurechne, gar nicht bestritten. Dann aber muss die Frage lauten, ob der Handlungsbegriff selbst in naturalistischen Ambitionen zu Passe kommenden Kategorien analysiert werden kann, also etwa, indem Handlungen als Ereignisse in Raum und Zeit gedeutet werden – was ja auch ohne naturalistisches Selbstverständnis als plausibel erscheinen kann. Auch von propositionalen Einstellungen, wie Überzeugungen und dem Bilden eben solcher, scheinen wir auch vortheoretisch als Zustände bzw. Ereignisse zu reden – obgleich nur im Hinblick auf Zeitlichkeit. Es müsste also zumindest plausibel gemacht werden *warum*, bzw. erläutert werden, *in welchem Sinne* Handlungen oder propositionale Einstellungen nicht schmerzlos in den kategorial deskriptiven Rahmen von ‚Ding-Ereignis-Ursache-Sprachspielen‘ integriert werden können – jedenfalls nicht, wenn diese *in einer Weise verstanden* werden, die den zu kritisierenden Theorien überhaupt erst ihren naturalistischen Biss verleiht. Da der Schwerpunkt meiner kritischen Auseinandersetzung mit dem Naturalismus aber nicht im komplexen Feld handlungstheoretischer Fragen im engeren Sinne liegen soll⁴⁹, sondern bei Fragen im

⁴⁹ Etwa mit der Absicht, dadurch so etwas wie einen ‚Freiheitsbeweis‘ zu erbringen. Ich neige ohnehin zu der Auffassung, dass die Möglichkeit von Freiheit nicht wirklich zu *beweisen*, sondern – auf die ein oder andere Weise – so sehr mit den Fundamenten menschlicher Selbstverständnisse verwoben ist, dass eigentlich gar nicht so klar ist, was es heißen könnte, ihre Möglichkeit umfassend zu beweisen oder zu widerlegen (in ähnlichem Sinne Strawson (1962)). Am ehesten noch scheinen mir transzendente Argumente im Sinne K.-O. Apels so etwas, wie einen ‚indirekten Freiheitsbeweis‘ liefern zu können, indem man möglicherweise

Zusammenhang mit semantisch evaluierbaren Phänomene, möchte ich an dieser Stelle dazu nur soviel sagen, dass mir die richtige Pointe des Kategorienfehler-Einwandes in Bezug auf die sogenannte ‚Gründe-Ursachen-Debatte‘ darin zu bestehen scheint, dass *Gründe* und deren Relata, nämlich *mögliche* Handlungen (oder Urteile), nicht als in *irgendeiner* Form mit den möglichen Relata ursächlicher Zusammenhänge, nämlich (raum-zeitlichen) Ereignissen, identisch zu deuten sind: Ereignisse sind genauso wenig Kandidaten für eine Beurteilung als wahr oder falsch, bzw. richtig oder unrichtig, wie sie (etwa als Körperbewegungen) durch Gründe ‚rationalisiert‘ werden.

Wichtig für die mit dieser Arbeit verfolgten Ziele ist es allerdings, genauer zu klären, *in welchem Sinne* ‚Bedeutung‘, ‚Überzeugung‘ und andere der oben genannten Begriffe als *kategorial normative* verstanden werden sollten. So sind bedeutungsvolle sprachliche *Äußerungen* zwar fraglos Handlungen, aber damit sind jene, durch die Einordnung in einen gemeinsamen kategorialen Rahmen unterstellten, internen Zusammenhänge zwischen Bedeutungen, propositionalen Einstellungen und sprachlichem oder nicht-sprachlichem Handeln (bzw. Handlungstypen) ja noch nicht ohne weiteres geklärt. Die erforderliche Klärung dieses Zusammenhangs muss allerdings nicht darauf hinauslaufen, so etwas wie eine systematische Rekonstruktion von ‚intentionalen‘ und semantischen Begriffen aus pragmatistischer Perspektive zu versuchen. Ich bin gegenüber solchen rekonstruktiven Großvorhaben eher skeptisch eingestellt. Mir geht es hier in systematischer Hinsicht zunächst nur darum, *Vorschläge* zur Ordnung unserer begrifflichen Verhältnisse auf einer sehr grundlegenden Ebene zu machen, um auf dieser Grundlage die Einordnung der hier zur Diskussion stehenden Begriffe oder Kategorien des Handelns, der Bedeutung und des Geistes in einen gemeinsamen, von sinnvollen naturalistischen Kategorien klar unterschiedenen, kategorialen *Rahmen* zu begründen.

In der Folge muss dann allerdings, wofür ich bereits argumentiert hatte, das *Vorliegen* eines Kategorienverstoßes jeweils im einzelnen *argumentativ* aufgewiesen werden, was offensichtlich nur anhand eines Durchgangs durch prominente, zeitgenössische Naturalisierungsstrategien möglich ist. Die grundsätzliche Strategie wird dabei darin bestehen, zu zeigen, dass und wie solche Theorien bei dem Versuch scheitern, propositionale Einstellungen als zugleich genuin semantisch individuiert *und* im Sinne physischer Ereignisse charakterisierbar plausibel zu machen. Dieses Scheitern einer naturalistischen Beschreibung von propositionalen Einstellungen liefert dabei allerdings, wie ich meine, zugleich Gründe dafür, sie als für die Rolle *mentaler Ursachen* von Handlungen ungeeignet aufzufassen.

demjenigen, der die Möglichkeit der Freiheit (mit Argumenten) *bestreitet*, einen ‚performativen Selbstwiderspruch‘ nachweisen kann.

Wie bereits in der Einleitung dargelegt wurde, gilt es wichtige Unterschiede hinsichtlich der Form und des Anspruchs naturalistischer Ansätze zu berücksichtigen. Stichhaltige Argumente gegen *reduktionistische* Theorien, also solche, die Wege aufzeigen wollen, *wie bestimmte* begriffliche oder propositionale Gehalte als kausalwirksame, physische Eigenschaften verständlich gemacht werden können, mögen wirkungslos sein gegen jene Theorien, die unsere ‚intentionalen‘ Redeweisen lediglich als *Beschreibungsformen* von Ereignistoken deuten, deren Kausalität zwar vollständig durch ihre physischen Eigenschaften bestimmt, aber dennoch nicht auf diese reduzierbar sei. (Jene eliminativistische Position, die unsere ganze Redeweise über ‚intentionale‘ Phänomene als illusionär und ohne reale Grundlage einstuft, hat, wie schon gesagt, nur wenige Anhänger, erscheint zudem als unmittelbar selbstwidersprüchlich und wird mich deshalb nur am Rande interessieren.)

Geist: Eigenschaft oder Sinnpotential?

Besonders die eher vorsichtigen Varianten moderner naturalistischer Theorien der Bedeutung und des Geistes operieren auch mit Vokabularen, die traditionell als zu üblichen Formen naturalistisch zulässiger Redeweisen konträr verstanden wurden. Dies betrifft insbesondere die (mehr oder minder austauschbare) Verwendung der Ausdrücke ‚Verstehen‘, ‚Interpretieren‘ und ‚Sinn machen‘, deren Inanspruchnahme durch naturalistische Theorien eine thematische Nähe zu klassischen hermeneutischen Methodenfragen naheulegen scheint. Solche Redeweisen sind allerdings insofern irreführend, als ‚Verstehen‘ oder ‚Interpretieren‘ dabei als Fähigkeit zur Erstellung einer wahren empirischen Theorie über ein konkretes sprachliches Verhalten gedeutet wird. Ohne an dieser Stelle in die Details zu gehen, sollte klar sein, dass eine solche, auf die Zuordnung von Wahrheitswerten aus einer Beobachterperspektive beschränkte Verwendungsweise entsprechender Begriffe nicht den üblichen Selbstverständnissen hermeneutischer Methodik entspricht, die ja – vereinfachend ausgedrückt – insbesondere den Stellenwert des ‚Geschichten-Erzählens‘⁵⁰, bzw. die Notwendigkeit einer virtuellen Teilnahme⁵¹ betont.

⁵⁰ Vergl. etwa C. Geertz: *Dichte Beschreibung* (1987). *Locus classicus* ist natürlich: H. G. Gadamer: *Wahrheit und Methode* (1960).

⁵¹ Dieser Begriff ist natürlich unklar und problematisch. Eine hilfreiche Konkretisierung liefert m. E. Habermas‘ Gedanke der Fähigkeit zur Beurteilung von mit sprachlichen Handlungen erhobenen Geltungsansprüchen (J. Habermas: *Theorie des kommunikativen Handelns* (1981)).

Die Motive, die hinter solchen verbalen Annäherungsversuchen stecken, liegen auf der Hand: So erheben Naturalisten ja einerseits den Anspruch, grundsätzlich die gleichen Felder beackern zu können, wie ihre hermeneutischen Kollegen. Andererseits soll dabei aber empirische Theoriebildung, im Sinne der beobachtungsgestützten Überprüfbarkeit von Gesetzesaussagen, der Axiomatisierbarkeit der Theorie und einer gewissen Präzision in der Operationalisierung methodologischer Annahmen und theoretischer Begriffen, die einzig verfügbarer Methode des Erklärens darstellen (wobei mitunter gegen die der Hermeneutik generell unterstellte Metaphorik des ‚Einfühlens‘ polemisiert wird). Wenn also überhaupt die Chance bestehen soll, eine Art naturalistischer Integration von sprachlicher Bedeutung und Geist zu erreichen, die sie zu ganz ‚normalen Dingen‘ innerhalb einer empirisch erfahrbaren Welt machte, dann müssen die entsprechenden, ontologisch zunächst eher ‚sperrig‘ erscheinenden Phänomene von vornherein konzeptuell so ‚zurechtgebogen‘ werden, dass sie sich in den kategorialen Rahmen von ‚Ding-Ereignis-Ursache-Sprachspielen‘ einfügen, ohne dass der Verdacht aufkommt, schlicht das Thema gewechselt zu haben.

Damit allerdings wird die ganze Fragestellung von Beginn an aufs ontologische Gleis gesetzt. Und obwohl sich Naturalisten gerade auch durch die Ablehnung eines *Substanzdualismus* definieren, orientieren sie sich doch in einigen wesentlichen methodologischen, ontologischen und bedeutungstheoretischen Vorentscheidungen noch immer an Vorgaben nicht nur des klassischen *Empirismus*, sondern auch des *Kartesianismus*. Dies gilt insbesondere im Hinblick darauf, Geist (bzw. ‚Intentionalität‘ /Semantizität /Repräsentationalität /Symbolizität) generell im Sinne *psychologischer Eigenschaften* von Individuen und die Bedeutung sprachlicher Äußerungen (fast immer) im Sinne eines *Zum-Ausdruck-Bringens subjektintern konstituierter*, mentaler Repräsentationalität zu deuten. ‚Sprachliche Intentionalität‘ wird damit zu einer derivativen Form ‚mentaler Intentionalität‘. Der moderne Naturalismus hängt also nicht nur dem cartesianisch-repräsentationalen Modell eines als subjektiv verstandenen Geistes an, sondern auch einem mindestens auf Locke zurückgehenden Sprach- bzw. Bedeutungsmodell, nach dem die Bedeutung von sprachlichen Ausdrücken in einer öffentlichen Mediatisierung *privat konstituierter Repräsentationen* von Welt besteht⁵².

Das Konzept eines weltrepräsentierenden, *subjektinhärenten* Geistes zum Ausgangspunkt für ein Verständnis von Geist und Bedeutung überhaupt zu machen gehört, so denke ich, zu den hartnäckigsten Dogmen der neuzeitlichen Philosophiegeschichte. Und es ist eines der wesentlichen konstruktiven Ziel dieser Arbeit, im Rahmen der Kritik naturalistischer Positionen auch für eine *sprachlich-*

⁵² Für eine aufschlussreiche, kritische Analyse des lockeschen Sprach- bzw. Bedeutungsmodells und seiner modernen Varianten, siehe C. Gauker: *Thinking Out Loud* (1994).

intersubjektive Konstitution von Sinn und Bedeutung zu argumentieren, dessen subjektive Instantiierungen dagegen als explanativ nachgeordnet zu betrachten. Dabei soll ‚subjektiver Geist‘ (bzw. die Zuschreibbarkeit propositionaler Einstellungen) als eine *originär auf verantwortbare Subjekte bezogene normative Lokalisierbarkeit* bezüglich *diskursiver und praktischer Verhältnisse* im Rahmen eines *sozial konstituierten Raums möglichen Sinns* verstanden werden. Dieser Raum entspricht einer vorläufigen, also dynamischen Bandbreite des Verstehbaren und möglicher Verständigung; dem *Rahmen möglicher Unterscheidungen und Individuierungen*; vor allem der *Einordenbarkeit individueller Handlungstoken in ein allgemeines Muster*. ‚Subjektiver Geist‘ sollte demnach nicht so verstanden werden, als ob wir mit entsprechenden Zuschreibungen *individuenintern* konstituierte *Zustände* oder *Ereignisse* (geistiger Art) *beschreiben* oder auf sie *Bezug nehmen*. Vielmehr werden mit propositionalen Zuschreibungen – wie auch mit Handlungsbeschreibungen – Subjekten *Positionen* innerhalb bestimmter *normativer Verhältnisse von Gültigkeit und Geltungsanforderungen* im Rahmen einer Handlungs- und Verständigungspraxis *zugewiesen*.

Nach der hier vertretenen Auffassung sollte also „Geist“ weder als ein je individuell kognitive Leistungen ermöglichendes Organ („*mind-brain*“) oder dessen Eigenschaften, noch als eine sich aus einem Interpretationsschema ergebende Beschreibungsform besonderer Arten von Zuständen oder Ereignissen verstanden werden, sondern primär als ein sprachlich verkörpertes Sinn- und Vernunftspotential sozialer Praktiken. Dabei entsprechen die *sprachlichen Unterscheidungs- und (Re-) Identifizierungsmöglichkeiten* weitgehend dem, was die phänomenologisch orientierte, sogenannte ‚Wissenssoziologie‘ einen gesellschaftlichen „Wissensvorrat“ bezüglich der Regelung sozialer Beziehungen, der Lösung weltbezogener Probleme und der Stiftung von Zwecken, subjektiver Identität und Lebenssinn nennt⁵³. Für die einer sozialen Praxis zugehörigen Subjekte sind mit der Verwendung bestimmter Kombinationen von sprachlichen Elementen bestimmte wechselseitige Anschlussmöglichkeiten innerhalb eines gemeinsamen ‚Sinnhorizontes‘ festgelegt. Propositionale Einstellungen, also auch die Gehalte entsprechender Zuschreibungen, sind, nach dieser Sichtweise, als derivativ zu den durch einen Sprachgebrauch *generierten gemeinsamen Sinnpotentiale* sprachlicher Symbole zu verstehen – und nicht umgekehrt.

⁵³ S. hierzu – wenngleich unter eher subjektivistisch-phänomenologischen Voraussetzungen: A. Schütz/T. Luckmann: *Strukturen der Lebenswelt* – Bd. 1 (1979) und Bd. 2 (1984). Ähnlich – allerdings ohne subjektivistische Fundierung – auch Habermas‘ Konzept von *Lebenswelt* in: ders.(1981).

Schon diese ziemlich grobe Skizzierung eines in der Tradition von Hermeneutik und Pragmatismus stehenden Gegenbildes zum Naturalismus legt auch ein eher starkes Konzept *geteilter* Bedeutung nahe. Nach meinem, in wichtigen Punkten an den Überlegungen des späteren Wittgenstein orientierten, Verständnis entspricht die primäre Bedeutung sprachlicher Zeichen ihrer durch institutionalisierten Gebrauch bestimmten, öffentlichen *Funktion* im Rahmen einer sozialen Praxis. Die Bedeutung eines Zeichens zu verstehen, heißt demnach zunächst einmal, zu wissen, wie man es relativ zu einem Äußerungskontext (meist in Kombination mit anderen Zeichen) verwenden muss, um dieser Funktion gerecht zu werden. Da sprachliches Handeln aber keine hinsichtlich einzelner Zeichen oder Sätze isolierbare Fähigkeit ist, beinhaltet sprachliche Kompetenz die Kenntnis inferenzieller Zusammenhänge zwischen möglichen sprachlichen und nicht-sprachlichen Handlungen. Diese Zusammenhänge werden thematisch, wenn Handlungsvollzüge im Hinblick auf ihre Richtigkeit reflexiv problematisiert werden.

Ich werde also, über das Konzept *geteilten Sinnes* hinaus, auch eine *soziale Konzeption von Bedeutung*, bzw. einer *geteilten Sprache* verteidigen, indem ich dafür argumentiere, dass begrifflicher, bzw. propositionaler Gehalt nur im Rahmen einer Verständigungspraxis möglich ist, deren Teilnehmer sich im Standardfalle wechselseitig unterstellen können, sich an *gemeinsamen Standards* der richtigen Verwendung sprachlicher Ausdrücke zu *orientieren*. Auch die Möglichkeit abweichender, partikularer Praktiken sprachlicher *Verständigung* ist nämlich auf eine intersubjektiv *kontrollierbare Beurteilbarkeit der Richtigkeit* von Äußerungen, bzw. *der Gültigkeit* der mit ihnen erhobenen Ansprüche angewiesen.

Ein solcher, relativ starker Begriff von geteilter Bedeutung ist allerdings, wie schon bemerkt wurde, für viele (und nicht nur naturalistisch gesinnte) Philosophen alles andere als selbstverständlich⁵⁴. Ich denke, die weitverbreitete Skepsis gegenüber einem solchen Bedeutungsverständnis hat – jedenfalls auf Seiten der Naturalisten – vor allem auch mit deren oben genannten, methodischen Selbstbeschränkungen zu tun. Denn diese Beschränkungen haben zur Folge, dass sprachliche Bedeutung nicht primär als (mehr oder minder kontextsensitive) Eigenschaft von *Symbolen*, sondern als sprecherrelative Eigenschaft von *Äußerungsereignissen* begriffen wird. Anstatt in sprachlicher Bedeutung dasjenige zu sehen, was bestimmt, *wie durch situationsrelative Verwendungen von Zeichen* sprachliche Handlungstypen aktualisiert werden, bzw. auf etwas in bestimmter Weise Bezug genommen wird, *identifizieren* Naturalisten (und nicht nur diese) Bedeutung meist unmittelbar mit dem Bezug oder dem Sinn. Diese

⁵⁴ Abgelehnt wird er unter anderem von so unterschiedlichen zeitgenössischen Philosophen, wie D. Davidson, J. Searle, R. Brandom, C. McGinn, J. Derrida, M. Frank, J. Bennett, J. P. Baker und P. M. S. Hacker.

Identifikation von sprachlicher Bedeutung mit dem Sinn oder der ‚Referenz‘ von Äußerungsereignissen kann aber, wie ich zeigen will, überhaupt nur dann als zwingend erscheinen, wenn sprachliche Verständigung als, in einem öffentlichem Medium kodierte, zwischensubjektive Übermittlung von zunächst *subjektintern generierten* Repräsentationen von Welt gedeutet wird.

Innerhalb der analytischen Philosophie gibt es eine fortdauernde Debatte über die Frage, ob die Gehalte (*content*) propositionaler Zuschreibungen als ‚weit‘ oder als ‚eng‘ verstanden werden sollten⁵⁵ – eine Unterscheidung, die auf die einschlägigen Fregeschen Überlegungen zur Differenz von ‚Bedeutung‘ (Bezug (‚Referenz‘), Extension) und ‚Sinn‘ (Gegebenheitsweise des Bezugs, Intension) zurückgeht. ‚Weiter Gehalt‘ soll dabei den Wahrheitsbedingungen einer Äußerung oder Überzeugung entsprechen, ‚enger Gehalt‘ dagegen demjenigen, was das Subjekt entsprechender Zuschreibungen glaubt, bzw. für wahr *hält* (ihm als wahr erscheint). ‚Weiter Gehalt‘ lässt sich in Form von sogenannten ‚*De-re*-Beschreibungen‘ (‚*x* glaubt von *y*, dass es *F* ist“) angeben, ‚enger Gehalt‘ nur in Form von ‚*De-dicto*-Beschreibungen‘ (‚*x* glaubt, dass *y F* ist“). Die vermeintliche Schwierigkeit, um die es dabei geht, besteht in einer unbefriedigenden Alternative: Entweder könnten wir das Subjekt der Zuschreibungen nicht *verstehen*, wenn wir keine ‚weite‘, *unseren* Überzeugungen entsprechende Beschreibung vornehmen; oder wir verlieren den *verhaltensklärenden* Aspekt solcher ‚Zustände‘ oder ‚Ereignisse‘ aus den Augen, wenn wir ihnen keine ‚enge‘, *seinen* Überzeugungen entsprechende Beschreibung geben. Dies ist natürlich eine starke Vereinfachung des ‚Problems‘. Tatsächlich wirken die, in unterschiedlichen Varianten geführten, Diskussionen zwischen sogenannten ‚Internalisten‘ und ‚Externalisten‘ meist eher sophistisch. Sie weisen ohnehin, wie ich meine, bereits in der Art ihrer Fragestellung in die falsche Richtung: Da sich vor allem die an dieser Debatte beteiligten Vertreter naturalistischer Positionen grundsätzlich an einem individualistischen Repräsentationsmodell des Geistes und der Frage nach der richtigen Form der Beschreibung entsprechender Zustände oder Ereignisse orientieren, können die tatsächlichen kognitiven und kommunikativen Funktionen der zugrundeliegenden Unterscheidung gar nicht in den Blick geraten. Denn diese bestehen, so jedenfalls werde ich argumentieren, zum einen in der – mit der Differenz von Sinn und Bezug gegebenen – *Ermöglichung des Lernens durch Neubeschreibung*, zum anderen in der – mit der Differenz von *De-re*- und *De-dicto*-Beschreibungen gegebenen – *Möglichkeit*

⁵⁵ Für ‚*Narrow-Content*-Theorien‘, siehe etwa J. Fodor: ‚Methodological solipsism considered as a research strategy in cognitive science‘ (1980); für eine Diskussion ihrer Schwierigkeiten: H. Putnam: *Repräsentation und Realität* (1988/1991); für einen Versuch der Synthese: C. McGinn: ‚The Structure of Content‘ (1982).

der *Übernahme oder Enthaltung von Verantwortung hinsichtlich bestimmter Beschreibungen*⁵⁶.

Diese eigentlichen Funktionen der Differenz von Sinn und Bezug werden allerdings erst vor dem Hintergrund zweier für das Verständnis von Bedeutung und Propositionalität wesentlicher, durch am naturalistischen Modell orientierte Theorien aber entweder fehlgedeuteter oder ganz unterschlagener, Faktoren deutlich: Erstens, der spezifischen *Struktur*, bzw. der *Wahrheitsfähigkeit* propositionaler Gehalte, deren wesentliche Funktion erst durch Strawson (und in der Folge insbesondere durch Tugendhat⁵⁷) herausgearbeitet wurde; der entscheidende Gedanke besteht hier darin, dass, zum einen, eine *situationsunabhängige Verständigung* zwischen Sprechern mit je *subjektiver Situationsbindung* und, zum anderen, individuelles und (wie eigentlich schon Frege gesehen hatte⁵⁸) gesellschaftliches *Lernen* erst unter Berücksichtigung der *spezifischen Funktionsweise unseres Repertoires singulärer Termini* verständlich gemacht werden kann. Und zweitens, der *diskursiven Relevanz* sprachlicher Handlungen (bzw. des inferenziellen Potentials ihrer Gehalte), dessen *normativer* Charakter, in der ein oder anderen Weise, von Wittgenstein, Sellars, Dummett, den ‚Erlanger Konstruktivisten‘, Habermas und in jüngster Zeit besonders von R. Brandom betont wurde; der entscheidende Gedanke hierbei ist, dass wir propositionale Gehalte – im Rahmen von Berechtigungsbedingungen für die Instantiierung von sprachlichen Handlungstypen – wesentlich über die Kenntnis ihres (*normativ* verstanden) *logischen Ortes im ‚Raum der Gründe‘* (Sellars) verstehen.

Beiden dieser Faktoren können naturalistische Theorien, die unserer Fähigkeit des Verstehens von sprachlichen Äußerungen in kategorial deskriptiven Begriffen zu reformulieren versuchen, wie ich zeigen will, nicht wirklich gerecht werden. Insbesondere die Analyse der spezifischen Rolle unseres Repertoires singulärer Termini für die perspektiven- und situationstranszendente *Beurteilbarkeit* einer Klassifikation oder Charakterisierung (im Hinblick auf Wahr- oder Falschheit) wird deshalb, in den konstruktiven Passagen meiner Kritik, einen zentralen Bestandteil des Aufweises der Unangemessenheit naturalistischer Erklärungs Bemühungen um das Verhältnis von ‚weitem‘ und ‚engem‘ Gehalt bilden.

⁵⁶ Das hier vertretene Verständnis des wesentlichen Unterschieds von *De-re-* und *De-dicto-*Beschreibungen stimmt weitgehend mit den entsprechenden Überlegungen von R. Brandom (1994) überein, die ja darauf hinauslaufen, diesen Unterschied im Hinblick auf Weisen möglicher *Stellungnahme*, bzw. der Übernahme oder Nicht-Übernahme von Verantwortung im Sinne inferentieller, bzw. argumentativer und praktischer Handlungsverpflichtungen oder – Berechtigungen zu deuten.

⁵⁷ P. F. Strawson: *Einzelding und logisches Subjekt* (engl.: *Individuals*) (1959/1972); E. Tugendhat: *Vorlesungen zur Einführung in die sprachanalytische Philosophie* (1976).

⁵⁸ G. Frege: ‚Über Sinn und Bedeutung‘ (1892/1986).

Fassen wir kurz zusammen: Wenngleich sich die hier verteidigte Version einer normativ-pragmatistischen Alternative zu naturalistischen Erklärungsversuchen von Bedeutung und Geist nicht auf eine pragmatistische Rekonstruierbarkeit im starken Sinne festlegt, behauptete sie doch eine *kategorienspezifische*, interne Beziehung der Begriffe „Bedeutung“, „Verständigung“ und „Handlung“ – jedenfalls dann, wenn Bedeutung und Verständigung auf dem Niveau propositionaler Gehalte gemeint ist. Dabei ist es sinnvoll, den Unterschied zwischen sprachlichen Handlungen (bzw. Handlungstypen) und der Bedeutung sprachlicher Symbolen im Auge zu behalten. Letztere bestimmt, im Sinne der Bedingungen der korrekten Verwendung von Ausdrücken (Ausdruckstypen), mögliche Handlungstypen, bzw. die Möglichkeit ihrer Aktualisierung in konkreten Handlungen. Es soll dann in der Folge dafür argumentiert werden, dass die Zuschreibbarkeit propositionaler Einstellungen, also Geist im Sinne von *mind*, auf die Vorraussetzung eines intersubjektiv bereits etablierten Sinnpotentials angewiesen ist, welches seinerseits nur im Rahmen einer – zumindest partiell – als geteilt unterstellten sprachlichen Praxis verständlich gemacht werden kann. Eine solche ‚Rekonstruktionslogik‘ widerspricht natürlich der Explikationsrichtung der meisten naturalistischen Ansätze, die Bedeutung mit Sinn oder Bezug *gleichsetzen* und *qua* propositionaler Einstellungen ‚erklären‘ wollen.

Nach dem bisher Gesagten dürfte übrigens auch deutlich geworden sein, dass ich assertorische Äußerungen – etwa gegenüber imperativischen Äußerungen – als für das Verständnis propositionaler Gehalte grundlegend betrachte. Der Grund hierfür ist einfach: Um zu verstehen, was es etwa heißt, einen Zustand *herbeiführen zu sollen*, muss man verstehen, was es heißt, dass ein Zustand *besteht oder nicht*, bzw. wann man sein Bestehen berechtigterweise behaupten kann. – Plakativ gesprochen: Man muss schon eine Welt haben, bevor man sie manipulieren kann.

Die Dichotomie von Verantwortung und Geschehen

In den bisherigen Ausführungen bin ich eine grundlegende Bestimmung dessen, was hier mit einer *normativ-pragmatistischen* Perspektive auf Bedeutung und Geist gemeint sein soll, noch schuldig geblieben. Diese sollte ja von der Art sein, dass eine klare Abgrenzung auch gegenüber nicht unüblichen, naturalistischen Verwendungen des

Begriffs „Pragmatismus“⁵⁹ möglich wird. Insbesondere bedarf die oben postulierte kategoriale Unterscheidbarkeit jener Formen von ‚Sprachspielen‘, die sich durch die Verwendung von Begriffen auszeichnet, die ich oben ‚kategorial normative‘ genannt habe, noch einer überzeugenden *kriterialen Fundierung*. Eine der geläufigsten, und oben auch bereits in Anspruch genommenen, Formen der Unterscheidung zwischen einer naturalistischen und einer nicht-naturalistischen Perspektive besteht in der Gegenüberstellung von Regelmäßigkeit und Regelhaftigkeit. Diese Gegenüberstellung ist natürlich überhaupt nur dann sinnvoll, wenn Regelhaftigkeit *normativ*, also im Sinne der *Beurteilbarkeit als richtig oder falsch und der Möglichkeit von Fehlern oder freiwilligem Nichtbefolgen* gedeutet wird.

Nun stellen allerdings funktionale Beschreibungen oder Erklärungen der Natur, wie sie etwa in der Biologie gang und gäbe sind, für diese Gegenüberstellung ein gewisses Problem dar, da auch hier so etwas wie die Möglichkeit von Fehlern im Sinne von *Dysfunktionalitäten* unterstellt wird. Dies gilt jedenfalls dann, wenn man, wie auch der Verfasser, eine Form normativen Verständnisses des Funktionsbegriffs als notwendig erachtet⁶⁰. Damit aber scheinen funktionale Erklärungsweisen paradoxerweise zugleich so etwas wie Absichten und Zwecke – bzw. ein in Hinblick auf diese *entworfen*es Design – innerhalb des Bereichs des Natürlichen zu unterstellen (was übrigens nicht nur für naturalistische *Hardliner* einen Grund darstellt, entsprechende Deutungen funktionaler Erklärungsweisen als ‚unwissenschaftlich‘ abzulehnen). Ich schlage vor, solche – offenbar nützlichen – teleofunktionalen Vokabulare, bzw. die entsprechenden Erklärungsweisen, dennoch vorbehaltlos der Seite der Naturerklärung zuzuschlagen. Was teleofunktionale Erklärungen zu Naturerklärungen macht, ist nicht nur, dass entsprechende Eigenschaften – anders als Bedeutung und Geist – auf der Grundlage rein physisch-kausaler Mechanismen *erklärbar* zu sein scheinen. Obschon nicht ohne explanativen Verlust *eliminierbar*, sind sie zudem als gegenüber *originären Fällen kategorial-normativer* Redeweise lediglich *derivative Form* zu verstehen – Redeweisen, die sich durch einen, für die hier behauptete kategoriale Differenz wesentlichen, zusätzlichen Aspekt der Möglichkeit von Richtig-falsch-Beurteilbarkeiten von jener über Natürliches unterscheiden, nämlich durch die *Verantwortbarkeit von Subjekten*.

Verantwortbarkeit verstehe ich als den *Grundbegriff des kategorialen Rahmens unserer Diskurse über Handeln, Geist und Bedeutung*. Dem steht, als *dichotomisches* Gegenstück, der Begriff des *Geschehnishaften*, als Grundbegriff des kategorialen Rahmens unserer Diskurse über *Natürliches*, gegenüber. Meine *These* ist also, *dass die*

⁵⁹ So wird beispielsweise Quine mitunter als ‚Pragmatist‘ eingestuft.

⁶⁰ Für ein solches Verständnis argumentiere ich unten mit Dennett und Millikan.

Dichotomie zwischen Geschehnis und Verantwortung den eigentlichen Kern jenes intuitiven Gegensatzes bildet, den wir oben vorläufig als einen zwischen ‚Ding-Ereignis-Ursache-Sprachspielen‘ und ‚Personen-Handlungs-Begründungs-Sprachspielen‘ gekennzeichnet hatten; sie ist der Dreh- und Angelpunkt meiner Naturalismuskritik. Im Lichte dieser kriterialen Dichotomie zwischen Verantwortbarkeit und Geschehnishaftigkeit werden, wie ich meine, gewisse Missverständnisse oder Schwächen der herkömmlichen, den Vorwurf der Kategorienfehlerhaftigkeit erhebenden, sprachdualistischen Argumentationsstrategie vermieden. Es ist dann nämlich für eine Abgrenzung des Natürlichen vom Nicht-Natürlichen unerheblich, dass wir häufig, einerseits, in anthromorphen Vokabularen über die Natur und, andererseits, in physisomorphen Vokabularen über geistige Phänomene, über uns selbst und andere Personen reden. Worauf es ankommt, ist – so behaupte ich zumindest – , dass Begriffe, wie „Handlung“, „Überzeugung“ oder „Bedeutung“ einen internen Bezug auf den Begriff der Verantwortbarkeit (von Urteilen oder Handlungen) haben und dies für die Begriffe, wie „Ding“, „(natürliches) Ereignis“ oder „(natürlicher) Zustand“ nicht gilt.

Auf der Grundlage der genannten Gegenüberstellung verstehe ich Handlungen demnach grundsätzlich als dasjenige Tun, für das Personen verantwortbar sind⁶¹. Lexikalisch betrachtet, verwenden wir zwar auch „Geschehen“ häufig in Bezug auf Handlungen. Wir meinen dies dann aber gerade *nicht* im Sinne eines *Gegensatzes zum Verantwortbaren* – den wir, in Fällen möglichen Missverständnisses, auch als „*bloßes Geschehen*“ bezeichnen können. Nicht die verwendeten *Ausdruckstypen* sind maßgeblich, sondern deren *Verwendung*, bzw. ihr, im Zweifelfalle klärbares, begriffliches Verständnis. Was ich behaupte ist, dass bestimmte Begriffe oder Kategorien den Begriff der Verantwortbarkeit von Subjekten voraussetzen und andere nicht, *sich beide Perspektiven aber wechselseitig ausschließen*. Zwar können wohl auch Handlungen in gewissem Sinne als Ereignisse verstanden, Ereignisse in solchen Kontexten aber nicht als (bloße) Geschehnisse gedeutet werden. Dies gilt insbesondere in Fällen, in denen wir Ursachen im Rahmen von ‚Handlungsdiskursen‘ ins Spiel bringen, indem wir von Handlungen als Ursachen von Ereignissen in der Welt sprechen (wenn etwa gesagt wird, dass es der Selbstmord des Kopiloten war, der den Absturz des Flugzeugs verursacht hat). Wir verwenden hier einen Ursachenbegriff, der offensichtlich nicht auf die Möglichkeit beschränkt ist, eine Relation zwischen Geschehnissen zu bilden. Vor allem aber scheint es mir *keinen* verständlichen Sinn zu geben, in dem man sagen könnte, Handlungen seien ihrerseits *durch Ereignisse* – im

⁶¹ Spezifische *Handlungsbeschreibungen* entsprechen dann spezifizierten *Zuschreibungen* von Verantwortbarkeit. Dieser Gedanke findet sich meines Wissens nach zuerst in: H. L. A. Hart: ‚The Ascription of Responsibility and Rights‘ (1948/1949).

Sinne von *bloßen Geschehnissen* – *verursacht*, da sie dem ‚Handelnden‘ sonst *widerführen* und somit nicht auf die Seite des Verantwortbaren, sondern des Geschehnishaften fielen⁶².

Wie der Ursachenbegriff überhaupt angemessen zu explizieren ist, ist ein notorisch schwieriges und nach wie vor strittiges Thema (auf das wir noch zurückkommen werden). Für meine hiesigen Zwecke reicht es aber aus, sich darauf festzulegen, dass in Fällen, wo von Verursachungsrelationen zwischen Ereignissen im Sinne von physisch beschreibbaren Naturgeschehnissen die Rede ist, für Verantwortbarkeit kein Raum bleibt. Und auch wenn es Sinn machen kann, von Handlungen als Ereignissen zu reden, so scheint es doch kriterial nicht wesentlich zu sein, dass wir immer auf ein physisches Ereignis (eine Körperbewegung) Bezug nehmen können müssen, wenn wir Handlungen zuschreiben. Ein einschlägiges, aber gern vernachlässigtes, Gegenbeispiel stellt die Kategorie der *Unterlassungen* dar, die sicher keine Ereignisse im fraglichen Sinne sind, für die wir aber gleichwohl verantwortlich gemacht werden können. Weil die kausale Handlungstheorie durch die Unterlassungen in ernsthafte Schwierigkeiten gerät, kann sie diese nicht als Handlungen deuten. Mir scheint es aber sinnvoller, sich an unserer bewährten *juristischen Praxis* zu orientieren, wo Unterlassungen mit ‚aktivem‘ Handeln praktisch (wenn auch nicht ganz) gleichgesetzt werden.

In den kategorialen Rahmen des *Geschehnishaften* fällt zunächst alles, was Gegenstand gesetzesartiger Kausal- oder Eigenschaftserklärungen⁶³ ist oder sein kann, also entsprechende Ereignisse, Eigenschaften oder Zustände. Aber auch alle auf Personen bezogene Erlebnisse und Eigenschaften im Sinne von Widerfahrnissen, soziale Verhaltensmuster, sowie – mit gewissen Einschränkungen – auch Empfindungen und Wahrnehmungen (im Sinne des Widerfahrens von Empfindungs- oder Wahrnehmungserlebnissen) gehören hierher, auch wenn wir in diesen Bereichen nicht unbedingt mit der Möglichkeit von kausalen- oder Eigenschaftserklärungen rechnen können. Geschehnishaft ist damit all dasjenige, was aus einer Perspektive der *Ausklammerung der Zuschreibbarkeit* von (Formen der) *Verantwortung* verstanden wird. Dies gilt – jedenfalls gemäß NANÜCS – unmittelbar für alle Eigenschaften und Vorgänge anorganischer oder pflanzlicher Natur, für das *Verhalten* von nicht-menschlichen, unter Umständen aber auch (zumindest nach gängigen juristischen Maßstäben) für das von menschlichen Tieren – solange letzteres vernünftigerweise nicht im Sinne von Handlungen von Personen verstanden werden kann. Es gilt nämlich

⁶² Dies wäre nur möglich, wenn wir (etwa mit Davidson) Handlungen als verursachte Körperbewegungen unter einer bestimmten Beschreibung verstehen könnten, was ich aber unten zurückweisen werde.

⁶³ Auf den Unterschied zwischen Eigenschafts- und Kausalerklärungen gehe ich im 2. Teil der Arbeit näher ein.

grundsätzlich auch für all dasjenige, dem *Personen* in unverfügter oder unverfügbarer Weise unterliegen oder was ihnen in dieser Weise zustößt oder widerfährt (etwa, wenn sie psychisch krank sind).

Ich stufe den Verantwortungsbegriff als für die Bestimmung des dichotomischen Gegensatzes zum Geschehnishaften grundlegend ein, weil er mir als der gegenüber dem Handlungsbegriff *umfassendere* zu sein scheint. So kann Verantwortbarkeit zunächst ganz allgemein als die *Möglichkeit* bestimmt werden, *Subjekt einer Richtig-falsch-Beurteilung* (wie etwa korrekt/inkorrekt, wahr/falsch oder gut/böse) zu sein, die in Kontexten unterschiedlicher Formen *normativer Konsequenzen* steht. Verantwortbarkeit kann es nur da geben, wo es Freiheit gibt, und zwar in dem Sinne, dass eine Alternativenwahl- oder Unterlassungsmöglichkeit besteht (oder bestand). Dies gilt aber auch für Überzeugungen, bzw. für sich nicht in direkter Weise sprachlich äußernde Urteile. Der auch hier vertretene *pragmatistische ‚Paradigmenwechsel‘ von der Repräsentation zur Verständigung* sollte uns – wie bereits angemahnt – nicht dazu verleiten, den systematisch wichtigen Unterschied zwischen sprachlichen und nicht-sprachlichen Handlungstypen und deren propositionalen bzw. semantischen Gehalten zu vernachlässigen. Die Weise, in der beides Gegenstand einer Richtig-falsch-Beurteilung und Subjekte dementsprechend, wie ich behaupte, verantwortlich sein können, ist durchaus eine unterschiedliche. Dies gilt nicht nur für den Unterschied zwischen Denken und Handeln sondern auch für sprachliche Handlungen selbst. Man kann etwas Wahres sagen und dabei dennoch in gewisser Weise falsch Handeln (etwa dann, wenn man jemanden dadurch verletzt). Und es ist im Hinblick auf die *Art der ins Spiel kommenden Normativität* sicher nicht dasselbe, einerseits möglichst angemessene begriffliche Unterscheidungen treffen oder richtige Schlüsse ziehen, andererseits nicht betrügen oder stehlen zu sollen. Während es in letzteren Fällen einen klaren Sinn gibt, in dem entsprechende Handlungen als nicht-richtig (‚verboten‘), bzw. als unter Sanktionsandrohung stehend verstanden werden können, ist dies in den zuerst genannten Fällen nicht ohne weiteres der Fall.

Zur Frage der Normativität von Rationalität und Bedeutung

Es stellt sich die grundsätzliche die Frage, in welchem Sinne Bedeutungsregeln, besonders aber logische oder Rationalitätsregeln, beziehungsweise Vernunftregeln ganz allgemein, überhaupt als *normativ* zu verstehen sein können oder sollten. So hat jüngst Kathrin Glüer den unreflektierten ‚Normativismus‘ gegenwärtiger Kritiker des

Naturalismus (wie etwa Kripke, Dummett und R. Brandom) kritisiert und zu zeigen versucht, dass sich dem Begriff der Normativität nicht nur im Hinblick auf Regeln des richtigen Schließens oder Urteilens, sondern auch im Hinblick auf ‚Bedeutungsregeln‘, gar kein verständlicher Sinn geben lässt⁶⁴. Sie weist, unter Berufung auf Davidson, zunächst mit Recht darauf hin, dass gewisse zentrale Vernunft- oder Rationalitätsstandards eher den Charakter *konstitutiver* Regeln haben. Dann aber bleibe kein Raum mehr zwischen der Zuschreibung bestimmter intentionaler Einstellungen und dem Nicht-Befolgen ihrer Konsequenzen, der einen vorschreibenden Sinn von Normativität zuließe:

„Besteht ein interner Zusammenhang zwischen zwei Überzeugungen, dann kann ich die eine gar nicht haben, ohne auch die andere zu besitzen. Genau das aber wäre nicht der Fall, wenn im Falle des Habens der einen die andere vorgeschrieben wäre; dafür müsste es vielmehr möglich sein, die eine zu haben, ohne die fragliche Folgerung zu ziehen. In den einfachsten und grundlegendsten Fällen aber scheint es zweifelhaft, dass diese Möglichkeit überhaupt besteht; hier scheint vielmehr zu gelten, dass ein bestimmter intentionaler Zustand überhaupt nur dann vorliegt, wenn bestimmte konstitutive Standards tatsächlich erfüllt sind.“ (Glüer 2000, S. 453)

Wenn ich beispielsweise (a.) glaube, dass es regnet, (b.) wünsche, nicht durch den Regen nass zu werden und (c.) glaube, dass ich trocken bleibe, wenn ich mich unter den Baum stelle, *sollte* ich mich dann unter den Baum stellen? Es scheint hier angemessener, zu sagen, dass ich dies alles nur dann wirklich glaube und wünsche, wenn ich mich schließlich unter den Baum stelle (*ceteris paribus*). Wenn propositionale Einstellungen vor allem durch ihre Beziehungen zu anderen Einstellungen individuiert werden, macht es, nach dieser Überlegung, keinen Sinn, zu fordern, dass wer die eine hat, auch die andere haben sollte, da man, in gewisser Weise, die eine gar nicht haben kann, ohne die andere zu haben. Allerdings will Glüer nicht das für intentionale Zustände und sprachliche Handlungen spezifische Charakteristikum der Richtig-falsch-Unterscheidbarkeit in Frage stellen, bzw. Regeln auf Regelmäßigkeiten reduzieren. Sie will lediglich bestreiten, dass dieses Charakteristikum *normativ* zu deuten ist. So bestimmten Rationalitätsstandards „nicht nur, wie wir denken, wenn wir ‚richtig‘ oder ‚wahr‘ denken, sondern zugleich, was Denken überhaupt ist“ (ebd.). – Sie bestimmten jedoch nicht, wie wir denken *sollten*.

Wie Glüer allerdings selbst bemerkt, gilt diese Überlegung, falls sie zutrifft, zunächst nur für einen zentralen Teil („minimale Rationalität“) dessen, was unter Rationalität oder ‚Intentionalität‘ verstanden werden kann. Sie tut jedoch in der Folge so, als gälte ihr Argument generell. Dabei unterscheidet sie hinsichtlich der Reichweite

⁶⁴ K. Glüer: ‚Bedeutung zwischen Norm und Naturgesetz‘ (2000).

ihres Argumentes auch nicht zwischen formalen und materialen inferentiellen Beziehungen. So scheint ihr Argument eher noch für die *Regeln* des logischen und praktischen Schließens plausibel. Es ist offenbar nicht möglich, sich einen Fall vorzustellen, in dem wir von jemanden fordern könnten, er/sie solle sich, in einem generellen Sinne, an die Regeln der Logik halten, denn wenn dies generell nicht der Fall wäre, würden wir gar nicht davon ausgehen, es mit einem Ansprechpartner für Sollens-Forderungen in irgendeinem Sinne zu tun zu haben. Dies ist jedoch im Hinblick auf die Zusammenhänge, durch die propositionale Einstellungen individuiert werden, nur in einem sehr eingeschränkten Sinne der Fall. Schon die von Glüer selbst in dem Beispiel angefügte *ceteris-paribus*-Klausel liefert einen Hinweis darauf, dass es mit den Regeln ‚minimaler Rationalität‘ oder ‚Intentionalität‘ beim Verstehen von sprachlichen und nicht-sprachlichen Handlungen nicht getan ist⁶⁵. Es gibt ja kein gleichsam zeitloses und unstrittiges Set an material-inferentiellen Beziehungen, durch die bestimmte propositionale Einstellungen individuiert werden. Außerdem erstreckt sich der Bereich der Regeln – oder, wie ich doch sagen würde, Normen – des Rationalen oder Vernünftigen ja weit über die des logischen und praktischen Schließens im von Glüer berücksichtigten Sinne hinaus. Dies gilt nicht nur im Hinblick auf Maßstäbe der Objektivität oder Angemessenheit, sondern insbesondere auch im Hinblick auf die Bewertung der Werte oder Zwecke als rational oder vernünftig. Unterschiedliche Erwägungen können ins Spiel kommen. Hier gibt es, wie ich denke, ein Verständnis interner Zusammenhänge, das sehr wohl einen Spielraum für die Möglichkeit des Befolgens oder Nicht-Befolgens gemäß entsprechender Maßstäbe oder Kriterien lässt. Wenn wir etwa sagen, jemand *sollte* eine Sache mal objektiver betrachten, dann geben wir damit eine *wertende Stellungnahme* hinsichtlich der *Art* von Maßstäben oder deren *richtigem Verständnis* oder *angemessener Gewichtung* ab, die für sein Urteil oder seine Handlung ausschlaggebend sein *sollte(n)*. Die jeweils möglichen *Interpretationen* von und *Abwägungen* zwischen verschiedenen Maßstäben

⁶⁵ In diesem Sinne unterscheidet auch Schnädelbach zwischen Rationalität als *formalem* und als *materialelem* Kriterium. Nur als materiales Kriterium sei Rationalität normativ zu verstehen und entspreche weitgehend der *evaluativen Beurteilung* von Personen oder ihren Verhaltensweisen nach Maßstäben der Rationalität. Formale Kriterien der Rationalität seien dagegen deskriptiv zu verstehen, und sollen den *Bedingungen der Verständlichkeit* von Wesen oder Verhaltensweisen entsprechen, also eine Präsupposition im Hinblick auf die Beurteilbarkeit von Rationalität im materialen Sinne bilden (H. Schnädelbach: ‚Rationalität und Normativität‘ (1990)). Schnädelbachs Konzept der formalen Kriterien der Rationalität als Präsuppositionen des Verstehens erinnert an Davidsons ‚transzendental-konstitutives‘ Verständnis seiner einschlägigen, sogenannten ‚Prinzipien der Nachsichtigkeit‘ (auf die wir noch zu sprechen kommen werden). Allerdings glaube ich, anders als Schnädelbach, dass sich aus dem evaluativen Sinn von Kriterien der Rationalität/Vernunft durchaus ein *präskriptiver* Sinn ableiten lässt, denn wenn eine Urteil oder eine Verhaltens- oder Handlungsweise im evaluativen Sinne von Vernunft gut oder richtig ist, dann *sollte* so geurteilt bzw. gehandelt werden.

und Standards rationalen bzw. vernünftigen Denkens und Handelns können nicht konstitutiv sein, weder fürs Denken, noch fürs Handeln. Es ist nicht nur üblich, sondern macht auch einen guten Sinn, zu sagen „Wenn du vernünftig wärest, würdest Du das nicht tun“, ohne dass wir damit schon meinen, jemand hätte bestimmte Überzeugungen nicht (bzw. ohne das wir damit meinen *können*, jemand denke gar nicht). Was jeweils vernünftig(er), also etwa ausgewogen(er), objektiv(er), explanativ(er), effektiv(er), einfach(er), sicher(er) usw. ist, ist eine nicht endgültig ausgemachte und zudem ein Mehr oder Weniger zulassende Frage (wie ja etwa auch die anhaltenden Auseinandersetzungen in der Wissenschaftstheorie verdeutlichen). Dies zeigt sich auch an Redeweisen wie „Du solltest daraus den Schluss ziehen, dass...“, die nur verständlich sind, wenn das Ziehen eines entsprechenden Schlusses nicht konstitutiv ist für dessen Prämissen bildende propositionale Einstellungen oder das Denken überhaupt. – Nur dann ist es ja auch möglich durch Neubewertungen auf der Grundlage bekannter Tatsachen zu neuen Einsichten oder Einschätzungen zu kommen.

Glüer stimmt Frege zu, dass sich Regeln des Für-wahr-Haltens, richtigen Urteilens und Schließens in dem Sinne als Vorschriften (im Sinne hypothetischer Imperative) verstehen ließen, dass sie sagen, wie zu denken ist, *damit* das Gedachte *wahr* ist (S. 451). Das scheint mir nicht ganz korrekt. Eher sollten wir sagen, dass diese Vorschriften sagen, wann ein Schluss *wahrheitserhaltend* ist, *wenn* die Prämissen wahr sind. Gerade um deren Wahrheit kann es aber rationalen ‚Streit‘ geben. Auch hier haben wir es wieder mit der Möglichkeit des Berufens auf Rationalität oder Vernunft in einem Sinne zu tun, der sich von einem den Naturgesetzmäßigkeiten Unterliegen gerade dadurch unterscheidet, dass sie sich in einer Weise an Subjekte richtet, welche die Möglichkeit offen lässt, dass ihnen *nicht* genügt wird – ohne dass diese Subjekte deshalb bereits, wie im Falle der konstitutiven Regeln der Rationalität, nicht mehr als Adressaten der Beurteilung nach Maßstäben der Vernunft in Frage kämen. Diesen Unterschied will ja, wie gesagt, auch Glüer nicht bestreiten. So bemerkt sie etwa im Hinblick auf die Rationale Entscheidungstheorie, diese beschreibe nicht, wie wir *tatsächlich* entscheiden. Sie hält jedoch daran fest, dass es auch nicht darum gehen könne, wie wir entscheiden *sollen*, sondern darum, *was* rationale Entscheidungen *sind*. Einen richtigen Punkt scheinen mir Glüers Überlegungen dabei insofern zu treffen, als wir Regeln des richtigen Urteilens oder Handelns nicht, wie etwa Normen positiven Rechts, willkürlich setzen können. Der entscheidende Unterschied zu Naturgesetzen liegt aber nicht nur in der nicht auf Regelmäßigkeiten reduzierbaren Regelmäßigkeit zentraler Regeln der Rationalität, sondern eben auch darin, dass es oberhalb ihrer einen nicht-konstitutiven Spielraum für die *Möglichkeit des Nicht-Befolgens* gibt. Und in diesem Bereich hat die Verwendung von Ausdrücken, wie „objektiv“, „klar“,

„explanativ fruchtbar“ oder „wahr“ immer auch den Charakter *wertender Stellungnahmen* hinsichtlich dessen, was der so Stellungnehmende als im Hinblick auf Urteile oder Handlungen *richtig* und *deshalb gesollt* erachtet. (Dies gilt übrigens ebenso für reflexive Urteile über die Vernunft selbst. Es kann, so denke ich, keine vollständige – gleichsam algorithmische – Explikation des Vernünftigen geben, weil jede Explikation ihrerseits auf ihre Vernünftigkeit hin zu Beurteilen wäre, und zwar anhand von Maßstäben (bzw. Intuitionen), die nicht zugleich (vollständig) *Gegenstand* der Beurteilung sein können. Urteile über Vernünftigkeit haben nicht nur einen Aspekt des Entdeckens, sondern auch einen zuletzt nicht tilgbaren Aspekt des Stellungnehmens. Diese Überlegungen werden auch unten, in der Kritik am Funktionalismus, zum tragen kommen)

Dass es selbst bei Einvernehmen über die (etwa empirischen) Prämissen auch in Fällen konstitutiver Rationalitätsregeln noch strittig sein kann, was zu folgern ist, liegt offenbar vor allem an der Möglichkeit von *Uneinigkeit über das richtige Verständnis der involvierten Ausdrücke*. Glüer ist nun allerdings der Ansicht, dass ihre Überlegungen zur Konstitutivität bzw. zum *sui-generis*-Charakter bestimmter Regeln auch für sprachliche Bedeutung gilt. Ihre Argumente sind dabei im Wesentlichen von der gleichen Art, wie jene gegen ein normatives Verständnis der Regeln des richtigen Schließens oder Urteilens. Wenn man davon ausginge, so Glüer, dass eine Bedeutungsregel *präskriptiv* zu verstehen sei, in dem sie uns sagt, *wie* ein Ausdruck verwendet werden *soll*, *damit* er eine bestimmte Bedeutung hat, würde jeder *inkorrekte* „Gebrauch“ (gemeint ist Verwendung) zu einer *Bedeutungsänderung* führen und es wäre nicht möglich, beispielsweise zwischen der *Korrektheit der Ausführung* einer bestimmte Behauptung und ihrer *empirischen Berechtigung* zu unterscheiden. Könnte man nicht mit „grün“ *grün meinen*, ohne es auch auf Nicht-Grünes anwenden zu können, wären weder empirische Irrtümer, noch Lügen, indirekte Rede, Ironie, Fragen u. ä. m. möglich. Doch selbst wenn wir dies berücksichtigen, lässt sich, so Glüer, kein normativer Zusammenhang verständlich machen:

„Wenn ‚grün‘ für einen Sprecher *grün* bedeutet, er also beabsichtigt, den Begriff *grün* auszudrücken, er sich aufrichtig und direkt äußert und dabei keinerlei empirischem Irrtum unterliegt, dann wird er ‚grün‘ antworten, wenn wir ihn nach der Farbe eines grünen Gegenstandes fragen. Tut er das nicht, dann hat er eine der angegebenen Bedingungen nicht erfüllt. Für ein Sollen ist hier kein Platz.“
(S. 461)

Hier wird allerdings auch deutlich, dass Glüers Argumentation ein individualistisches Konzept von Begriffen und Bedeutung unterstellt. So zeigt zunächst die Redeweise von der Bedeutung eines Ausdrucks für einen Sprecher, dass ihr Verständnis des Ausdrucks

„Gebrauch“ insofern voreingenommen ist, als sie erst gar nicht die Möglichkeit der Unterscheidung zwischen einer überindividuellen, also sozial üblichen oder institutionalisierten Verwendungsweise und einer je individuellen Verwendung in Betracht zieht. Ein solches überindividuelles Verständnis könnte ja immerhin die Grundlage für jenes in Frage stehende normative Verhältnis liefern, nämlich im Sinne einer Spannung zwischen dem, wie man sprachliche Ausdruckstypen üblicherweise verwendet (um bestimmte Dinge zu tun) und dem, wie sie jeweils – durch ein Individuum oder in einer bestimmten Situation – verwendet werden: Gilt in einer Sprachgemeinschaft eine bestimmte Verwendungsweise als richtig, kann ein Sprecher korrigiert werden, wenn er eigentlich beabsichtigt, den gemeinsamen ‚Regeln‘⁶⁶ bzw. dem üblichen Gebrauch der Sprachgemeinschaft zu entsprechen, dabei jedoch einen Fehler begeht, bzw. von diesem Gebrauch abweicht.

Wie wir schon gesehen haben, ist es jedoch strittig, ob ein normatives Verständnis von Bedeutung konzeptuell auf ein soziales Konzept von „Gebrauch“, bzw. auf so etwas, wie eine gemeinsame Sprache angewiesen ist (obschon ich unten genau dafür argumentieren werde). Ich will deshalb an dieser Stelle noch eine Überlegung pro ein normatives Verständnis von Bedeutung und Begrifflichkeit anstellen, bei der ich versuche, ohne eine solche Unterstellung auszukommen. Zu diesem Zweck müssen wir nochmal daran anknüpfen, wie Glüer, unter Berufung auf Überlegungen Kripkes und Boghossians⁶⁷, den Ausgangspunkt normativer Verständnisse von Bedeutung im Sinne von Präskriptionen bestimmt:

(G) Wenn „grün“ grün bedeutet, dann ist es korrekt, „grün“ nur auf einen Gegenstand x anzuwenden genau dann, wenn x grün ist.

Die deontische Interpretation von (G) – die ja von Glüer in der oben genannten Weise kritisiert wurde – wäre dann, dass „grün“ nur auf grüne Gegenstände angewendet werden soll (oder darf). Ich denke aber nicht, dass (G), bzw. seine deontische Interpretation, ein Beispiel für eine angemessene Explikation dessen darstellt, was wir unter der Normativität von Begrifflichkeit und Bedeutung verstehen sollten⁶⁸. Ein

⁶⁶ Hinsichtlich des Begriffs ‚gemeinsamer sprachlicher Regeln‘ sollte man es allerdings nicht zu genau nehmen. Ich verstehe dieses Konzept auch eher im Sinne einer idealisierenden Unterstellung, als der Beschreibung einer Tatsache. Dies werde ich vor allem im letzten Kapitel – im Zusammenhang mit Davidsons Bedeutungstheorie – erörtern.

⁶⁷ S. Kripke: *Wittgenstein über Regeln und Privatsprache* (1982/); P. Boghossian: ‚The Rule-Following Considerations‘ (1989).

⁶⁸ Die folgende Kritik betrifft in mancher Hinsicht auch Kripkes (1982/ 1987) Interpretation des sogenannten ‚Regelfolgeproblems‘ und seine vergebliche Suche nach ‚Bedeutungstatsachen‘. Denn Bedeutungen (bzw. begriffliche Unterscheidungen) konstituieren ja erst mögliche Tatsachen, weshalb Tatsachen, jedenfalls in ‚rekonstruktionslogischen‘

Bikonditional, wie (G), ist eigentlich trivial und verschiebt die Angelegenheit, um die es eigentlich gehen muss, lediglich auf die Ebene der Metasprache. Was aber eine normorientierte Perspektive auf sprachliche Bedeutung, meiner Ansicht nach, thematisieren muss, ist der *begriffskonstitutive Charakter des Gebrauchs sprachlicher Ausdrücke*. Dass man überhaupt über einen Begriff des Grünen verfügen und der Ausdruck „grün“ eine Bedeutung haben kann, ist ja zunächst mal alles andere als selbstverständlich. Die eigentlich interessante Frage ist doch, was wir überhaupt meinen, wenn wir von Begriffen reden und was es heißt, über einen Begriff zu verfügen oder sich über Begriffe zu streiten. Was hier ins Spiel kommt, sind doch offensichtlich Fragen von der Art, wie man etwas (richtig) machen soll oder sollte, wenn man, *in einer bestimmten Hinsicht*, zum Zeitpunkt $t_{(1)}$ (sprachlich-prädikativ) *das gleiche tun will*, wie zum Zeitpunkt $t_{(0)}$. Dabei scheint es mir angebracht, den Begriff „Begriff“ als objektivierende Abstraktion des normativen Musters anzusehen, das sich im Gebrauch sprachlicher Ausdrücke spiegelt, bzw. ihre Verwendung leitet – und nicht als eine platonische Entität, auf die wir irgendwie Zugriff haben.

Eine Person, die sich in diesem Sinne an Paradigmen der richtigen Verwendung von Ausdrücken orientiert, kann dies offensichtlich nicht auf der Grundlage einer Regel wie (G) tun. Für einen Begriff, wie den des Grünen, lässt sich eine solche Regel rein sprachlich auch gar nicht angeben⁶⁹. Dies wird besonders deutlich im Falle des Erwerbs begrifflicher Kompetenz im Rahmen des Spracherwerbs. Wer etwa über die Fähigkeit der regelgeleiteten Verwendung eines Ausdrucks, die wir mit dem Beherrschen eines Begriffs verknüpfen, noch nicht verfügt, dem nützen ja Erläuterungen nichts, die den Begriff als bereits verstanden voraussetzen. Und gerade im Hinblick auf die Lernphase gibt es genügend Spielraum fürs Fehlermachen, da die normativen Muster, bzw. stimmige Projektionen aus Beispielfällen, erst gebildet werden müssen. Die explizite Form einer entsprechenden kontextbezogenen ‚Regel‘ könnte, im Sinne einer Orientierungsbemühung aus der Perspektive der ersten Person, für den diskutierten Fall etwa in folgender Weise formuliert werden:

(G_N) Wenn ich „grün“ gemäß den Beispielen verwenden will, die mir als typisch für eine bestimmte Unterscheidungsmöglichkeit beigebracht wurden (bzw. bei

Bemühungen, nicht die normative Grundlage für die Interpretation von Begriffen abgeben können. Auch halte ich es nicht für sehr glücklich, mathematische Beispiele zu diskutieren und die Folgerungen dann umstandslos auf alle Arten begrifflicher Verständnisse zu verallgemeinern, da mathematische Begriffe oder Regeln von ganz anderer Art sind und eher platonistische Intuitionen nähren, als etwa Farbprädikate.

⁶⁹ Dass dies *am Ende* für jede Art von begrifflichem Verständnis gilt, da der Versuch metasprachlicher Formulierungen von Regeln des richtigen Verständnisses von Regeln zwangsläufig in einen vitiösen Regress führt, will Wittgenstein mit seiner Diskussion des sogenannten ‚Problems des Regelbefolgens‘ ja vor allem zeigen.

denen ich diesen Ausdruck bisher typischerweise angewendet habe), dann sollte ich auch dies↓ „grün“ nennen.

Wenn wir den normativen Aspekt von Bedeutung im Sinne von (G_N) verstehen, gibt es einen klaren ‚Spielraum‘ für ein Sollen (bzw. Dürfen). (G_N) lässt sich zudem, insbesondere im Hinblick auf das, was unter „typisch“ zu verstehen sein kann, in vielfältiger Weise kommentierend anreichern: So gehört zur Fähigkeit einer typischen Verwendung des deutschen Wortes „grün“ (oder eines anderen gleich verwendeten Ausdrucks), bzw. zu einer entsprechenden begrifflichen Grün-Kompetenz, die Fähigkeit zu einer extrapolierenden Orientierung an natürlicherweise relativ stabilen paradigmatischen Beispielen, wie etwa der (regelmäßig gewässerte) Rasen von Gärten oder auf Fußballplätzen; aber wohl auch das – zumindest implizite – Wissen, dass Grünes unter bestimmten, nicht-standardmäßigen Beleuchtungsverhältnissen als andersfarbig erscheinen kann und es fließende Übergänge zum Blauen, nicht aber zum Roten gibt (u. ä. m). Es spricht allerdings manches dafür, dass sich unser jeweiliges begriffliches Orientierungswissen nicht vollständig explizit machen lässt. Ohnehin ist eine *umfassende*, kontextfrei formulierbare sprachliche Explikation unserer Bedeutungskompetenz schon deshalb nicht denkbar, weil es, auch hier – ähnlich wie im Falle der Vernunft – einen Zirkularitätseinwand gibt: Erklärungen von Bedeutung (oder Begriffen) können nicht von Grund auf rein sprachliche sein, ohne vorauszusetzen, was erklärt werden soll. (Dieser Zirkularitätseinwand und die Feststellung, dass der ‚Sprung‘ von der Regel zur Anwendung am Ende nicht seinerseits durch eine Regel formuliert werden kann, ohne in einen vitiösen Regress zu verfallen – worauf ja Wittgenstein eindringlich hingewiesen hat – scheinen mir zwei Seiten desselben Phänomens zu sein.)

Im Falle höherstufiger Begriffe kommt noch hinzu, dass es hier aus Gründen der Komplexität der Zusammenhänge mit anderen Begriffen weit mehr Spielräume für intersubjektive Uneinigkeit geben kann. Dass Formulierungen des Regelcharakters von Bedeutung im Sinne von (G) unangemessen im Hinblick auf eine Erläuterung ihres normativen Orientierungscharakters sind, zeigt sich ja in ganz ähnlicher Weise, wie im Fall des Spracherwerbs, besonders auch dann, wenn begriffliche Verständnisse – etwa des Ausdrucks „Gerechtigkeit“ – *strittig* sind. Was wir in Fällen des Spracherwerbs oder der Uneinigkeit zunächst sozusagen ‚objektiv‘ haben, sind sprachliche Ausdruckstypen, für deren, in einem *vernünftigen Sinne, als gleich oder typisch zu verstehende Verwendung* (möglicherweise auch durch nur ein Individuum) dann bestimmte Kriterien, Umstände oder Zusammenhänge *als maßgeblich gefordert* werden.

Die genannten Fälle – der Spracherwerb und ‚Meta-Diskurse‘ über Sprachliches in Fällen der Uneinigkeit – sind allerdings nur explizite Fälle dessen, was, in gewisser Weise, ganz allgemein gilt. Denn jede deskriptiv-assertorische Äußerung ist nicht nur eine *Stellungnahme* hinsichtlich dessen, wie es sich in der Welt verhält, sondern *zugleich* auch eine implizite Stellungnahme hinsichtlich der Frage, *wie oder wann man einen Ausdruck*, gegebenenfalls unter Maßgabe verschiedenerlei Erwägungen, *verwenden sollte, wenn man, in gewisser Hinsicht, etwas Gleiches tun will*⁷⁰. Diese Aspekte lassen sich nicht sauber trennen. In Fällen komplexer, wahrnehmungsfernerer Begriffe, wie etwa „Krankheit“ oder „Gerechtigkeit“, ist dieser, für die *Möglichkeit* von Tatsachen ‚konstitutive‘ Charakter unserer (impliziten, expliziten oder begründenden) Stellungnahmen hinsichtlich einer richtigen (vernünftigen⁷¹) Verwendung entsprechender Ausdrücke nur besonders augenfällig.

Worauf hingegen Glüers Überlegungen zum Regelcharakter von Bedeutung letztlich hinauslaufen, zeigt sich an ihrer Kritik des (zum präskriptiven Verständnis alternativen) Versuchs, *konstitutive* Regeln der Bedeutung als normativ zu deuten. Zwar entgehe eine Regel, wie

(KG) Es ist korrekt, „grün“ auf einen Gegenstand x anzuwenden genau dann, wenn x grün ist

zunächst den von ihr genannten Schwierigkeiten von (G), verschöbe das Problem jedoch lediglich auf die Gedankeninhalte, da sich ein Sprecher an ihnen lediglich bei der Wahl seiner *Ausdrucksmittel* orientieren könne (S. 466). Während also (KG) normativ zu verstehen sei, könne sie doch keine bedeutungskonstitutive Regel sein, da sie die Korrektheitsregeln für *Gedanken* schon voraussetze, welche nun ihrerseits *zwar bedeutungskonstitutiven, aber* wiederum *nicht-normativen*, lediglich eine *binäre Klassifikation* (wahr/falsch) *ermöglichenden* Regeln entsprächen. Damit scheint sich Glüer allerdings endgültig auf einen weitgehenden Platonismus im Sinne Freges festgelegt zu haben. Dies wird durch ihre Rede vom *sui-generis*-Charakter von

⁷⁰ So gibt es etwa Türkis-Töne, die manche eher noch als in einem weiten Sinne grünlich, andere jedoch eher als in einem weiten Sinne bläulich bezeichnen würden. Auf dieses Thema werde ich im Zusammenhang mit der Kritik an Millikans Korrespondenztheorie nochmal ausführlicher zu sprechen kommen.

⁷¹ Die hier vorgeschlagene Interpretation des normativen Charakters von Bedeutung und Begriffen impliziert natürlich, dass unsere begrifflichen Orientierungen auch andere sein könnten. Aber sie würden dann unter Umständen keine sinnvollen oder angemessenen Unterscheidungen und damit keine aufschlussreichen Erkenntnisse oder keine erfolgreiche Praxis ermöglichen. Auch sollte das „typisch“ im jeweils *einfachsten, projezierbaren* Sinne gedeutet werden, ohne temporäre oder andere Spezifizierungen, um Fälle wie Kripkes berühmte ‚Quus-Funktion‘ (Kripke (1982/1987)) oder Goodmans ‚Grot-Prädikat‘ (N. Goodman: *Tatsache, Fiktion, Voraussage* (1954/1988)) auszuschließen.

Rationalitäts- und Bedeutungsregeln lediglich terminologisch abgemildert und erscheint mir – zumindest im Hinblick auf Bedeutung und Begrifflichkeit (bzw. nicht rein formale Zusammenhänge) – aus den genannten Gründen wenig attraktiv. Solche Regeln können ja nur so verstanden werden, dass sie zeitlos gelten und alle möglichen ‚Bedeutungen‘ bzw. Begriffe und damit alle möglichen Tatsachen bestimmen, egal ob es je eine Praxis gibt, in der sprachliche Ausdrücke ihnen gemäß verwendet werden.

Dennoch können wir festhalten, dass Glüers Einwände, besonders im Hinblick auf die Regeln des richtigen Schließens oder Urteilens, eine gewisse Berechtigung haben. Denn zum einen kann der Sinn von Normativität hier offenbar nicht der sein, den wir mit dem präskriptiven Charakter gängiger regulativer Handlungsnormen (wie etwa: „An der roten Ampel muss man halten“) verbinden. Und zum anderen kann die Gültigkeit von Schlüssen oder die Wahrheit von Urteilen nicht (nur) von unseren kognitiven Vermögen oder von sozialen Konventionen und normativen ‚Einstellungen‘ abhängen, ohne dass – ebenso unattraktive – relativistische Konsequenzen drohten. Deshalb halte ich es für wichtig, den grundlegenden, aber weiten Begriff von Verantwortbarkeit idealtypisch in eine *intersubjektive* und in eine *objektive* Form zu unterscheiden. Diese Unterscheidung drängt sich auf, wenn man die Möglichkeiten normativer Beurteilung von verantwortbaren Subjekten bzw. ihrem Tun, Urteilen oder Denken, in drei Richtungen hin differenziert: hinsichtlich der *Reichweite ihres Geltungsanspruchs* (1.), hinsichtlich des *Gegenübers der Verantwortung* (2.), und schließlich hinsichtlich der möglichen *Konsequenzen im Falle des Nicht-Befolgens* (3.).

ad (1.): Hier lassen sich, meist *intersubjektive Beziehungen* regelnde, Normen oder Regeln mit *lokalem Geltungsanspruch* bzw. alle als bloß *konventionell gültig mögliche* Regeln (auch die konstitutiven) einerseits von *nicht-konventionellen*, häufig auch *solitäre Handlungen oder Urteile* orientierende Regeln, Normen, Ideale, Paradigmen, Maßstäbe oder Standards mit *universellem Geltungsanspruch* andererseits unterscheiden, also solchen, deren Gültigkeit sich nicht am faktisch Geltenden bzw. am mehrheitlich, konsensuell oder durch Autoritäten als gültig Erachtetem bemisst.

In letztere Kategorie fallen jene schon diskutierten ‚Vernunftgebote‘, also nicht-konstitutive Normen (u. ä.) des richtigen Urteilens oder Handelns im weitesten Sinne, wie etwa moralische Normen (oder Ideale oder Intuitionen), epistemische bzw. methodische Normen, Ideale oder Maximen, Heuristiken und sogenannte ‚Klugheitsregeln‘. In erstere Kategorie fallen dagegen Normen des positiven Rechts, der Sitte oder von Riten, aber auch Spielregeln. Klare Beispiele hierfür sind etwa, dass in Deutschland auf der rechten Straßenseite zu fahren ist (in Großbritannien dagegen auf der linken) oder bestimmte Heiratszeremonien. (Mit einem nicht intersubjektiv

angelegten Fall konventionell geregelter, lokaler Geltung haben wir es bei Regeln für Spiele zu tun, die man alleine spielt.)

Natürlich können oder sollten auch viele der zunächst lokal geltende Normen oder Regeln mit intersubjektivem Verpflichtungscharakter, wie etwa Rechtsnormen oder Riten, im Hinblick auf ihre objektive Vernünftigkeit beurteilt werden, doch dies ist ein zusätzlicher, die Geltungsgründe betreffender Schritt, der für ihren normativen Charakter nicht ausschlaggebend ist. ‚Bedeutungsregeln‘ (bzw. Regeln, Standards oder Paradigmen der richtigen Verwendung sprachlicher Ausdrücke) fallen interessanterweise zwischen intersubjektive und objektive gültige Normen, da sie zwar eine starke intersubjektive, lokal-konventionelle, aber auch eine wichtige objektive Komponente haben, indem die Bedeutung von Ausdrücken, ist ein mehr oder minder willkürlicher Gebrauch erst etabliert, auch Richtigkeitsmaßstäben unterworfen ist, die über ihre *faktischen* Verwendungen und Beurteilungen hinausweisen.

ad (2.): Intersubjektiv geltende Normen, wie „wenn man jemand, den man gut kennt, am Tag zum ersten mal trifft, sollte man grüßen“, „Innerhalb geschlossener Ortschaften gilt, wenn nicht anders angezeigt, Tempo 50“ oder „Wer sagen will, dass der Schnee weiß ist, muss (im deutschen Sprachraum) sagen ‚Der Schnee ist weiß‘“, legen Verantwortbarkeiten *nur* gegenüber anderen Subjekten oder gegenüber Institutionen fest. Die soziale Gruppe, bzw. bestimmte Autoritäten innerhalb ihrer oder einfach ich und mein jeweiliges Gegenüber, bestimmen, welche Regeln gelten und ob ihnen gemäß gehandelt wurde. Dies gilt nicht im Falle objektive Geltung beanspruchender Normen. Moralische Normen (etwa als Hintergrund der Erfüllungsbedingungen von Sprechakten, wie „Versprechen muss man halten“), begriffliche Richtigkeitsstandards, Objektivitätsmaximen oder –ideale machen uns, wie wir es ausdrücken können, *auch* gegenüber dem ‚Gerichtshof der Vernunft‘ (mit seinen ‚Unterabteilungen‘, wie etwa dem ‚Tribunal der Erfahrung‘) verantwortbar.

ad (3.): Die Nicht-Beachtung, bzw. -Befolgung sozialer, lokal geltender Normen ist, je nach Art und Inhalt von Regel oder Norm (bzw. Schwere des Verstoßes) mit dem ganzen Spektrum ‚externer‘ (sozialer) Sanktionsdrohungen belegt – vom Ignoriertwerden bis zur Haft- oder gar Todesstrafe. Die Normbefolgung kann dabei aus reinen Klugheitserwägungen, also wegen der Vermeidung möglicher Sanktionen, erfolgen. Anders ist die Lage im Fall von Normen oder Richtigkeitsstandards, die nicht nur lokale, sondern auch universelle Gültigkeit, bzw. objektive Verantwortbarkeit begründen. Sie zu befolgen heißt zugleich, Stellung zu nehmen, in dem ihre Gültigkeit anerkannt wird. Die Konsequenzen einer entsprechenden Nicht-Beachtung oder -Befolgung bestehen hier zunächst nur in so etwas, wie dem Begehen von Irrtümern

oder Fehlschlüssen, dem Verpassen von Einsichten und möglichen besseren Verständnissen oder dem Verfehlen der richtigen Einstellung. Unsere Handlungen, Einstellungen, Schlüsse oder Bedeutungsverständnisse sind dann falsch, nicht-objektiv, inkonsequent, unmoralisch, unangemessen, voreilig oder geschmacklos (was übrigens bei Gelegenheit des Bemerkens entsprechender Unzulänglichkeiten im eigenen Fall durchaus ein *fühlbares* persönliches *Unbehagen* – und insofern so auch etwas wie eine ‚innerliche Sanktion‘ – hervorrufen kann). Zwar führt eine entsprechende Nichtbeachtung häufig auch zu sozialen oder instrumentell-praktischen Schwierigkeiten, wie nicht verstanden zu werden oder dem Scheitern von Handlungsplänen. Ein Mangel an Vernunft, bzw. der Befolgung ihrer ‚Regeln‘ kann auch durch eine Minderung sozialer Wertschätzung bis hin zum nicht mehr ernstgenommen zu werden sanktioniert sein. Doch es besteht kein jeweils *spezifischer* Zusammenhang zwischen bestimmten Verstößen gegen Gebote der Vernunft und bestimmten externen Folgen dieser Art.

Subjekte (bzw. Personen) sind also nicht nur intersubjektiv, gegenüber anderen Personen oder Institutionen, verantwortbar, sondern auch – wenngleich in einer anderen Weise – objektiv, gegenüber den ‚Forderungen der Vernunft‘⁷². Diese Unterscheidung ist insofern idealtypisch (im Sinne einer Idealisierung zu analytische Zwecken), als wir es ‚in der freien Wildbahn‘ realer Verhältnisse natürlich häufig mit *Mischverhältnissen* zu tun haben. Intersubjektiv geltende und entsprechend sanktionsgestützte Normen treten dann mit dem Anspruch auf objektive Gültigkeit auf und werden von Subjekten häufig auch als solche anerkannt. Andererseits impliziert die Berufung auf die objektive Gültigkeit einer Norm durch Subjekte grundsätzlich auch die Forderung nach deren intersubjektiver Anerkennung, bzw., nach faktischer sozialer Geltung. Auf jeden Fall aber ist es für Subjekte möglich, Verantwortlichkeit entsprechend zu differenzieren. Wenn ich etwa an einer gut übersichtlichen und momentan völlig verkehrsfreien Stelle eine rote Fußgängerampel ignoriere, unterliege ich zwar einerseits der (lokal) uneingeschränkt geltenden Rechtsnorm, kann ich mich aber fragen, ob es wirklich sinnvoll bzw. vernünftig ist, solche Gebote ausnahmslos zu befolgen (und zu dem Schluss kommen, dass dies nicht der Fall ist).

Jedenfalls stehen Verstöße im Sinne der Verantwortlichkeit gegenüber der Vernunft (idealtypisch) nicht in gleicher Weise unter Sanktionsandrohung, wie in Fällen der Verantwortlichkeit gegenüber Personen oder Institutionen. Die Vernunft ‚droht‘ uns nicht, sondern ‚appelliert‘ gleichsam an einen guten *Willen zu intellektueller Redlichkeit* und kritischer Selbstdistanz, an unseren Sinn fürs Machbare oder zur

⁷² Und für Gläubige auch gegenüber den Forderungen eines allmächtigen Gottes.

Rücksichtnahme gegenüber anderen. Und wenn wir uns anderen gegenüber mit Argumenten oder Einwänden zu ‚Anwälten der Vernunft‘ machen, berufen wir uns auf diese nicht im Sinne von Normen oder Regeln, die andere oder ‚wir selbst‘ uns gegeben haben.

Verantwortung zu *übernehmen* oder *anzuerkennen*, heißt, sich willentlich innerhalb bestimmter, ‚normativer Räume‘⁷³ zu *positionieren*, bzw. *Stellung zu beziehen* und sich zugleich deren Geltungsanforderungen und eventuellen Sanktionsmöglichkeiten zu unterwerfen. Sie *zuschreiben* heißt dann, *andere so zu positionieren* und damit als bestimmten Geltungsanforderungen oder Sanktionsmöglichkeiten unterworfen zu behandeln bzw. einzustufen. Normative ‚Räume‘, bzw. Verhältnisse sind durch *normative* Konsequenzen für übernommene oder zugeschriebene Verantwortung bestimmt. Solche Konsequenzen können, wie gesagt, etwa rechtlich geregelten Folgen von bestimmten Handlungen oder (nicht-konstitutiven) diskursiven Folgerungen aus bestimmten Überzeugungen entsprechen.

Exkurs: Zu Robert Brandoms normativ-pragmatischer Grundlegung der Semantik

Robert Brandom hat in *Making it Explicit* (1994) den großangelegten Versuch unternommen, die Bedingungen zu beschreiben, die es (für einen Interpreten) gerechtfertigt erscheinen zu lassen, eine Praxis als *diskursive*, als eine des Gebens und Forderns (bzw. Akzeptierens) von Gründen, zu *explizieren*. – Welche Struktur eine Praxis haben muss, um (implizit) jene normativen Ressourcen zu generieren, die für eine Explikation auf der logisch-semantischen Ebene begrifflicher (propositionaler) Gehalte und ihrer inferentiellen Zusammenhänge hinreichen. Brandom bestimmt eine solche Praxis als notwendig sprachliche und soziale Praxis der Übernahme (Anerkennung) und Zuschreibung von *commitments* und *entitlements*⁷⁴. Der *deontische*

⁷³ Diese Wortwahl spielt natürlich auf Sellars ‚*logical space of reason*‘ an (in ders.: ‚*Empiricism and the Philosophy of Mind*‘ (1956)). Der von mir verwendete Plural soll anzeigen, dass innerhalb des Raums der Vernunft Differenzierungen möglich sind, wobei sich die unterschiedlichen normativen Räume jeweils durch die für sie spezifischen Formen von Gründen konstituieren.

⁷⁴ Wenn man „entitlement“ als *Berechtigung* übersetzt, scheint mir *Verpflichtung*, als komplementärer Begriff, die angemessene Übersetzung von „commitment“ zu sein. Die in der deutschen Ausgabe von *Making it Explicit* mit dem Titel ‚*Expressive Vernunft*‘ (2000) gewählte Übersetzung als „Festlegung“ scheint dagegen jene Problematik einer normativen Deutung von Rationalität und Bedeutung zu entschärfen, die wir oben schon angesprochen hatten. Da aber Brandom ausdrücklich ein *normatives* – und nicht ein lediglich auf einen *sui generis* *Regelbegriff* zurückgreifendes – Programm verfolgt, muss auch „Festlegung“ präskriptiv, bzw. im Sinne einer Verpflichtung gedeutet werden, womit am Ende also nichts gewonnen ist. Es ist

Status von sprachlichen Handlungen, also etwa die sich aus ihren jeweiligen Aktualisierungen ergebenden inferentiellen Verpflichtungen, soll grundsätzlich lediglich als *Form eines sozialen Status* verstanden werden, der, in zunächst ‚implizit-praktischer‘ Form, über die *normativen Einstellungen* derer, die *commitments* und *entitlements* zuschreiben und übernehmen, *gestiftet* wird. „Implizit-praktisch“ meint hier, dass sich die Teilnehmer als entsprechend *committed* oder *entitled behandeln*, ohne sich dabei die normativen Zusammenhänge explizit zu vergegenwärtigen zu müssen. Brandom glaubt nun, die Möglichkeit der *Objektivität* begrifflichen und propositionalen Gehaltes in einem Sinne verständlich machen zu können, der jener *hier* vorgenommenen Unterscheidung intersubjektiver und objektiver Verantwortbarkeit zu entsprechen scheint:

„Being a reason is to be understood in the first instance in terms of what it is for a community to *treat* something in practice as such a reason, on the practical side of reasons for actions just as on the doxastic side of reasons for claims. In neither case is this approach to normative status (what one is *really* entitled or committed to) through normative attitude (what one is *taken* to be entitled or committed to) incompatible with making eventual sense *objective* norms, which underwrite the possibility that everyone’s attitudes toward them are wrong“ (Brandom 1994, S. 253)

Deontische bzw. normative Status werden also zunächst durch die praktisch-impliziten *Einstellungen* von Teilnehmern einer sozialen Praxis instituiert, die sich wechselseitig als zu unterschiedlichen weiteren Handlungen *committed* oder *entitled behandeln*. Dabei kommt es – schon aufgrund jeweiliger biographischer Besonderheiten – zu Unterschieden in den entsprechenden normativen Einstellungen, das heißt, es gibt keine (vollständige) Übereinstimmung darüber, worauf sich ein Subjekt mit einer sprachlichen Handlung (inferentiell) *committed* hat oder zu was es sich *entitled* hat, da keine (vollständige) Übereinstimmung darüber besteht, welche sprachlichen Handlungen berechtigt (z. B. wahr) sind, bzw. was als Grund für was gilt. Um sich dennoch wechselseitig richtig interpretieren (und Handlungen ‚prognostizieren‘) zu können, müssten Teilnehmer einer solchen norm-stiftenden Praxis – zumindest in einem implizit-praktischen Sinne – in der Lage sein, den ‚Spiel-‘ oder ‚Kontostand‘ hinsichtlich *eigener* (soll heißen: tatsächlicher oder *objektiver*) und *fremder* (soll heißen: scheinbarer oder *subjektiver*) *commitments* und *entitlements* im Auge zu behalten, bzw. darüber ‚Buch zu führen‘, und, je nach den sprachlichen Handlungen, welche die Teilnehmer dieser Praxis ausführen, in systematischer Weise anzupassen. Für das Gelingen sprachlicher Kommunikation (und das Verstehen von Handlungen

wohl am besten, die Ausdrücke im Englischen zu übernehmen, da sie bei Brandom einen mehr oder minder technischen Gebrauch erfahren.

überhaupt) bedarf es also, nach Brandom, zweierlei Art von normstiftenden ‚Listen‘: einmal derjenigen, mit der der Interpret darüber Buch führt, zu was sich der Sprecher – nach Maßgabe des Interpreteten – festgelegt zu haben *glaubt* und zum anderen derjenigen, mit der der Interpret darüber Buch führt, zu was sich der Sprecher – nach Maßgabe des Interpreteten – *tatsächlich* festgelegt *hat*⁷⁵.

Die Ausführungen Brandoms, denen die in dieser Arbeit vertretene Position in vielem nahesteht, sind ebenso erhellend wie komplex, und ich kann seinem Explikationsniveau hier nicht wirklich gerecht werden. Es soll aber immerhin kurz auf einige grundlegende Punkte eingegangen werden, in denen mir die philosophisch-konzeptuelle Basis seines Programms einer normativ-pragmatischen⁷⁶ Fundierung von Bedeutung und begrifflichem Gehalt als unzureichend erscheint. So halte ich es zunächst für eine wesentliche systematische Schwäche seines Ansatzes, dass er eine tendenzielle Angleichung von sozialen Handlungsnormen und Normen des richtigen Urteilens vornimmt und deshalb nicht in überzeugender Weise zwischen intersubjektiver und objektiver Verantwortbarkeit unterscheiden kann⁷⁷. Dies wird etwa dort deutlich, wo, im Rahmen der Begründung der normativ-pragmatischen Ausgangsbasis, bzw. des begrifflichen ‚Rohmaterials‘ von *Making it Explicit*, der durch die praktischen Bewertungen gestiftete normative Charakter von *commitments* durch ihre ‚interne Sanktioniertheit‘ – im Gegensatz zu einer rein externen, auch nicht-normativ (etwa durch positive oder negative Verstärkung im behavioristischen Sinne) beschreibbaren Sanktioniertheit – erläutert werden soll (Kap 1, bes. Abschn. IV/5). Interne Sanktionen sind für Brandom solche, bei denen die Konsequenzen einer Nicht-Befolgung lediglich in einer Veränderung des deontischen Status nachfolgender Ausführungen bestehen (also jemand den Status des Berechtig-Seins verliert):

⁷⁵ Dies entspricht in etwa der schon angesprochenen Unterscheidung zwischen ‚engem‘ und ‚weitem‘ Gehalt durch naturalistische Ansätze.

⁷⁶ Als die Kernthese des *Pragmatismus* verstehe ich die Behauptung, dass jeder kognitiv signifikante Unterschied im Hinblick auf Bedeutung oder propositionale Einstellung *irgendeinen* öffentlich erkennbaren Unterschied im (sprachlichen oder nicht-sprachlichen) Handeln machen muss – auch wenn wir hier, wegen des holistischen Charakters der genannten Phänomene, keine ‚Eins-zu-eins-Zuordnungen‘ erwarten sollten. Dass semantisch charakterisierte Phänomene, also Instantiierungen von Bedeutung und Geist_{pe}, innerhalb des begrifflich-kategorialen Rahmens von Verhältnissen intersubjektiver und objektiver Verantwortbarkeit verstanden werden müssen, sehe ich dagegen als die – anti-naturalistische – Kernthese eines *Normativismus* an. Zumindest im Hinblick auf den Begriff des Pragmatismus scheint Brandom ein abweichendes Verständnis zu vertreten. Er bezeichnet mit ‚Pragmatismus‘ mitunter lediglich diejenige Ebene der Rekonstruktion, auf der wir nur davon reden können, dass Äußerungen (implizit) als berechtigt oder unberechtigt *behandelt* werden.

⁷⁷ Sowohl in diesem Punkt, als auch mit anderen der folgenden kritischen Bemerkungen stimme ich in vielem mit den von Habermas gegen Brandom vorgebrachten Einwänden überein (s. J. Habermas: ‚Von Kant zu Hegel. Zu Robert Brandoms Sprachpragmatik‘ (1999)).

„The consequences of an assessment of a performance as correct or incorrect with respect to one norm may extend no further than other assessments of correctness, with respect to other norms. It is possible to interpret a community as instituting normative statuses by their attitudes of assessment, even though each such status that is discerned is responded to by sanctions that involve only other normative statuses“ (S. 44)

Mir erscheint es allerdings fraglich, ob der Begriff der *internen Sanktion* tatsächlich einen verständlichen Sinn hat und es bleibt zudem unklar, welchen systematischen Stellenwert er einnehmen soll. So ist es bereits zumindest zweifelhaft, ob man überhaupt einen rein *normativen* Begriff von *Sanktion* im Sinne Brandoms verständlich machen kann, also einen, bei dem man am Ende nicht doch auf irgendeine Form nicht-normativ zu beschreibender Bestrafung (,negative Reizverstärkung‘) angewiesen ist. Vor allem aber kann dasjenige, *woraufhin* die Teilnehmer ihre normativen Bewertungen vornehmen, in Fällen von unter ,internen Sanktionsandrohungen‘ stehenden Verbindlichkeiten nicht dasselbe sein, wie in Fällen externer Sanktionsandrohung – ob normativ beschrieben oder nicht. In Fällen der Beurteilung eines Handelns unter dem Aspekt seiner *intersubjektiven* Verantwortbarkeit besteht die Grundlage der Beurteilung, wie gesagt, normalerweise in sittlichen, rechtlichen oder sonstigen Formen *faktisch geltender* bzw. institutionalisierter Handlungsnormen, die verbindliche *Richtig-falsch-* (Gut-böse-; Erlaubt-verboten-) *Unterscheidungen* etablieren und deren Nicht-Befolgung irgendeine Art von unangenehmen Konsequenzen *seitens anderer Subjekte* nach sich ziehen kann. Nur vor diesem Hintergrund hat die Rede von Sanktionen oder Sanktionsandrohung einen klaren Sinn (und dies gilt auch dann, wenn die entsprechenden Normen durch die praktischen Einstellungen/Bewertungen der an einer Praxis teilnehmenden Subjekte erst instituiert werden). In Fällen des Beurteilens jedoch, in denen der Aspekt *objektiver* Verantwortbarkeit im Vordergrund steht, wie etwa im Hinblick auf inferentielle *commitments*, sind entsprechende Voraussetzungen, die der Rede von Sanktionen einen Sinn geben könnten, nicht ohne weiteres zu erkennen. So wird mit der Rede von ,internen Sanktionen‘ die qualitative Andersartigkeit von intersubjektiver und objektiver Verantwortbarkeit mehr verschleiert als erklärt und es bleibt am Ende unklar, wieso überhaupt die Wahrheit zu sagen etwas als richtig oder gut zu Beurteilendes, hingegen einen falschen Schluss zu ziehen etwas als unrichtig oder schlecht zu Beurteilendes ist. Anders ausgedrückt: Warum sollten wir überhaupt Widersprüche vermeiden oder uns der Kraft des besseren Argumentes beugen? Was ist das *Sanktionierende* an einem Widerspruch, *was* ,zwingt‘ uns *intern*?

Diese Unklarheit hat, wie mir scheint, auch mit Schwierigkeiten zu tun, die auf Brandoms Anspruch gründen, begrifflichen Gehalt im Sinne der Möglichkeit *objektiver*

Richtigkeit oder *Berechtigung* auf der Basis von normativ-sozialen *Einstellungen* (bzw. denen aus ihnen folgenden normativen *Bewertungen*) begründen zu können (insbes. Kap. 8). Eine wesentliche Bedingung für die Angemessenheit einer normativ-pragmatischen Erklärung von begrifflichen- bzw. Bedeutungsnormen soll ja sein, dass sie der Unterscheidbarkeit zwischen einer ihnen gemäßen und einer ihnen nur scheinbar gemäßen Ausführung gerecht werden kann. Die Idee, objektiven begrifflichen Gehalt und inferentielle Richtigkeit als aus sozialen Verhältnissen emergierend verständlich zu machen, in denen sich Subjekte wechselseitig als (zu bestimmten Ausführungen) *committed* oder *entitled behandeln*, will nun Brandom durch die Inanspruchnahme des Sanktionsbegriffs verdeutlichen. Dabei soll, wie wir gesehen haben, die *normative* Binnenstruktur inferentieller Verhältnisse nicht auf die Voraussetzung *externer* Sanktionen angewiesen sein, sondern allein über einen Begriff *interner* normativer Sanktionen etabliert werden.

Offenbar ist nun aber – selbst wenn wir von den genannten Zweifeln hinsichtlich des Begriffs der ‚internen Sanktion‘ absehen – die von Brandom proklamierte *Wechselseitigkeit* der normativen Einstellungen nicht wirklich eine *intersubjektive*. Vielmehr spielt ein genuin intersubjektives Verständnis für seine Konzeption der Möglichkeit ‚interner Sanktionen‘, bzw. für Fragen nach der Konstitution der internen normativen Struktur inferentieller Verhältnisse, gar keine systematische Rolle. Da jeder Teilnehmer einer kommunikativen Praxis seine Beurteilungen anderer auf der Grundlage je eigener normativer ‚Punktekontoführung‘ vornimmt, handelt es sich offensichtlich gerade *nicht* um echte Wechselseitigkeit der Beurteilung, Begründung und Kritik im Rahmen *gemeinsamer Orientierungen*, sondern eher um eine *gegenseitige* Beurteilung von aus jeweils unterschiedlichen Perspektiven *individuell normsetzenden* Subjekten. „Richtig sein“ soll zwar nicht auf „für richtig halten“ reduzierbar sein; aber es sind schließlich doch individuelle Haltungen, bzw. Einstellungen, aus denen heraus dieser Unterschied verständlich gemacht werden soll („It’s attitudes all the way down“). Objektivität wird von Brandom dabei rein formal, nämlich im Sinne *struktureller Bedingungen der Möglichkeit von Objektivität*, als der *Unterscheidbarkeit von Sein und Schein*, analysiert: „(...) the distinction between claims or applications of concepts that are objectively correct and those that are merely taken to be correct is a structural feature of each scorekeeping perspective“ (S. 595). Auf einer solch schmalen individual-perspektivischen Basis scheint mir aber weder der quasi konstitutive Charakter insbesondere formal-logischer oder formal-praktischer Regeln der Vernunft, noch unser modernes Verständnis möglicher Facetten der Differenz von Schein und Sein *qua* eines objektiven ‚Zwangs‘ von ‚Geboten der Vernunft‘, noch ein mögliches Scheitern von Handlungsplänen und das dadurch

mögliche Lernen (etwa im Hinblick auf angemessene begriffliche Differenzierungen) in Konfrontation mit einer intrinsisch strukturierten natürlichen Welt in plausibler Weise einholbar zu sein⁷⁸.

Brandoms perspektivische Konzeption von Objektivität wird, wie ich denke, dem reflexiven Selbstverständnis vernunftorientierter Subjekte insofern nicht wirklich gerecht, als sie zwar *notwendige Bedingungen* jeder Form von Objektivitätsverständnissen anführt, aber sicherlich keine *hinreichenden* – und zwar nicht mal in dem schwachen Sinne einer *angemessenen Explikation* unterschiedlicher, diskursvarianter Kriterien. Wir verstehen unsere entsprechenden kognitiven Bestrebungen eben nicht im Sinne brandomscher Interpreten, die sich, im Falle diskursiver Differenzen, notwendigerweise selbst als Maßgabe hinsichtlich dessen nehmen, wozu jemand *committed* oder *entitled*, bzw. was richtig oder wahr ist, sondern im Sinne des Bemühens um eine Orientierung an *substantiellen* Maßstäben des Objektiven, die *subjektunabhängig* sind. Sowohl individuelle, als auch gesellschaftliche Lernprozesse müssen wir, retrospektiv, immer auch als einen Annäherungsprozess – eher im Sinne des Entdeckens, als des Hervorbringens – im Hinblick auf die *Aneignung* solcher Maßstäbe bestimmen, etwa im Sinne der fortschreitenden Dezentrierung je eigener Perspektive oder des zunehmenden Gelingens einer kohärenten Explikation bestimmter Konsequenzen oder Zusammenhänge partieller Überzeugungs- oder Wertbereiche. Die Aufgabe einer besonderen Beherrschung von Maßstäben objektiver Urteilsbildung wird dabei häufig auch, diskursspezifisch, an bestimmte ‚Expertengruppen‘ delegiert, deren Urteile (bzw. ihre Grundlage) allerdings ihrerseits einer externen Kritik unterliegen können. Insofern ist das Bemühen um richtiges Urteilen und Objektivität grundsätzlich eine fortlaufende, *gemeinsame* Anstrengung, für die es wesentlich ist, dass individuelle (oder gruppenspezifische) Beurteilungen *durch andere korrigiert* werden können.

Was richtig oder wahr ist – und nicht nur so scheint – wird zwar, wie Brandom richtig feststellt, nicht einfach durch die Gemeinschaft oder eine Mehrheit bestimmt,

⁷⁸ Die hier behauptete moderne Intuition, dass die Befolgung von Maßstäben der Objektivität zu einer Annäherung an Beschreibungen intrinsischer Strukturen der Welt führt, ist offenbar stärker, als eine Präsupposition formaler Konzepte geteilter Welt(en) im Sinne Habermas‘ (1981, S. 82 f.). Ich denke, die Unterstellung einer gemeinsamen, subjektunabhängigen Welt kann nicht rein formal oder abstrakt („von allen bestimmten Inhalten losgelöst“) gedacht werden, sondern immer nur substantiell – selbst wenn wir glauben, (noch) nicht über eine angemessene Beschreibung zu verfügen oder gar je verfügen zu können. Auch ein ‚struktureller Realismus‘ – eine Konzeption, die ich unten wenigstens skizzieren und soweit auch befürworten werde – versteht eine entsprechende Annäherung nicht in diesem Sinne rein formal. (Zu einer aufschlussreichen Gegenüberstellung der Konzeptionen von Objektivität bei Brandom und Habermas, s. auch C. Lafont: ‚Ist Objektivität perspektivisch? Ein Vergleich zwischen Brandoms und Habermas‘ Konzeption von Objektivität‘ (2001).)

denn einerseits ist die Gemeinschaft kein Subjekt von Einstellungen oder Bewertungen, andererseits können, versteht man die Gemeinschaft lediglich als die Summe aller Individuen, alle irren. Zu allen gehört aber auch der Interpret einer sprachlichen Praxis. Diese einfache Logik wird auch nicht durch eine (potentiell unendliche) Iterierung von Perspektivendifferenzen aus der jeweils nächst höheren Metaperspektive urteilender Interpreten aufgehoben. Die Unterscheidung zwischen Sein und Schein lässt sich wohl besser als eine in unserem begrifflich-propositionalen Verständnis strukturell angelegte, sich wechselseitig bedingende, Möglichkeit der Einsicht in kognitive Unzulänglichkeiten deuten, mit der wir grundsätzlich immer zu rechnen haben. Sie resultiert nicht aus einer perspektivisch wahrgenommenen Differenz eigener und fremder *commitments*, sondern aus dem Wissen um die immer gegebene Möglichkeit der Konfrontation mit der Kraft des besseren Argumentes, die Individuen durch andere erfahren können.

Ein brandonscher Minimalismus kann auch nicht die Frage beantworten, warum wir unsere Beschreibungen der Welt überhaupt in eine Richtung weitertreiben sollten, die es erlaubt, Widersprüche, die durch falsifizierende Erfahrungen entstehen, auszuräumen. Was treibt uns zu immer allgemeineren theoretischen Erklärungen der für uns erfahrbaren Phänomene. Es gibt (oder gab doch zumindest) sprachliche Gemeinschaften, die sich im Rahmen von ‚Weltbildern‘ orientieren, nach deren ‚internen‘ Verständnissen etwas, was uns als Widerspruch erscheinen mag, gar nicht als solcher wahrgenommen wird⁷⁹. *Innerhalb* solcher Gemeinschaften kann es zwar wohl eine Differenz von Sein und Schein im Sinne von Brandons rein formalen Bedingungen für Objektivität geben. *Wir* würden einem entsprechenden Überzeugungssystem allerdings nicht aufgrund abweichender Buchführungen hinsichtlich ihrer und unserer Überzeugungen gemäß den von *ihnen* instituierten begrifflichen Normen, sondern aufgrund der *Unangemessenheit* bestimmter Teile ihres (wenngleich möglicherweise konsistenten) *Begriffssysteme* und der – substantiell verstandenen – Nicht-Objektivität bzw. Unvernünftigkeit ihrer Beurteilungsmaßstäbe hinsichtlich der Entscheidung darüber, was Grund für was ist, die Zustimmung verweigern.

Ohne so etwas, wie einen ‚minimalen Platonismus‘, ist, so scheint mir zumindest, objektive Vernunft nicht zu haben. Und ich denke, im Gegensatz zu Brandom, auch nicht, dass wir bei dem Versuch, die unterschiedlichen Formen der Möglichkeit des objektiv Richtigen zu verstehen, sowohl auf die wechselseitige Unterstellung von als gemeinsam möglichen Bedeutungsverständnissen, als auch auf die generelle *Intuition einer*, aus unterschiedlichen Perspektiven und auf unterschiedlichen kognitiven

⁷⁹ Vergl. etwa E. E. Evans-Pritchard: *Hexerei, Orakel und Magie bei den Zande* (1937/1978).

Niveaus erfahrbaren, singulären Welt *mit intrinsischen Eigenschaften*, mit der wir uns sowohl kognitiv als auch praktisch (möglichst erfolgreich) auseinandersetzen *müssen*, verzichten können.

Zusammenfassung und weitergehende Erläuterung der Argumentationsstrategie

Als Kern meiner Kritik am Dogmatischen Naturalismus hatte ich eine Form von Kategorienverstoß-Einwand bestimmt, der im Grundsatz besagt, dass naturalistische Theorien (Beschreibungen von) Instantiierungen von ‚Intentionalität‘ unzulässigerweise mit natürlichen Ereignissen oder Zuständen identifizieren. Genauer gesagt sollen ‚intentionale‘ Beschreibungen, je nach den wissenschaftstheoretischen, bzw. metaphysischen Hintergrundannahmen naturalistischer Theorien, entweder kausalwirksame physische Ereignisse oder kausalwirksame physische Eigenschaften identifizieren. Reduktive Theorien behaupten darüber hinaus, dass sich diese ‚intentionalen‘ oder semantischen Eigenschaften in einer Weise reformulieren lassen, die keiner ‚intentionalen‘, bzw. semantischen Begrifflichkeit bedarf. Dabei wird vorausgesetzt, dass – in einem vorausgehenden, ersten Schritt – sprachliche ‚Intentionalität‘ auf geistige, bzw. ‚semantische Tatsachen‘ auf ‚psychologische‘ zurückführbar seien.

Die durchzuführende Kritik wichtiger Positionen des Reduktiv-materialistischen Programms setzt vorrangig beim zweiten Reduktionsschritt an und soll zeigen, dass alle Versuche einer nicht-semantischen oder ‚nicht-intentionalen‘ Charakterisierung der Gehalte von propositionalen Einstellungen (paradigmatischerweise Überzeugungen) am Ende entweder doch wieder auf eine ‚intentionale‘ Begrifflichkeit zurückgreifen oder aber nicht mehr beanspruchen können, überhaupt noch in irgendeinem plausiblen Sinne von geistigen, bzw. semantisch evaluierbaren Phänomenen zu reden. Die raffinierteren Varianten des reduktiven Materialismus, die einigen Schwierigkeiten der kausalen Versionen des Funktionalismus unter Inanspruchnahme einer teleologischen Begrifflichkeit entgehen, trifft dagegen eine andere Form der Kritik. Denn diese setzen bei ihrem Versuch, propositionale Einstellungen oder Gehalte sozusagen erst über einen ‚teleologischen Umweg‘ kausalistisch zu bestimmen, eine Welt von Tatsachen voraus, die wir doch ihrerseits, so behaupte ich zumindest, nur in Abhängigkeit davon verstehen können, was es heißt, eine wahre Behauptung aufzustellen, bzw. eine wahre Überzeugung zu haben.

Analoges gilt am Ende auch für den Interpretationismus, obgleich die Dinge hier insofern etwas anders liegen, als insbesondere Davidson nicht mal mehr eine

abgeschwächte Form von Reduktionismus, wie der Teleofunktionalismus, verteidigen will. Zudem ist auch Davidsons *Tokenidentitätsthese* bezüglich mentaler und physischer Ereignisse mit gravierenden Problemen konfrontiert. Da Davidson aber eine wechselseitige Abhängigkeit von Zuschreibungen propositionaler Einstellungen und Bedeutungstheorien für Sprecher vertritt, werde ich mich vor allem auf seine Überlegungen zur Theorie der Bedeutung konzentrieren⁸⁰. Es soll gezeigt werden, dass Davidsons bedeutungstheoretische Vorschläge nicht nur wichtige Fragen zur kognitiven Struktur unseres Bedeutungsverstehens außen vor lassen, sondern in ihrer Konsequenz auch einigen zentralen Selbstverständnissen von Verständigungsteilnehmern nicht gerecht werden kann; Selbstverständnisse, die einen stärkeren Sinn von ‚Normativität‘ voraussetzen, als er zugestehen möchte. Es führt nämlich, wie ich meine, grundsätzlich in die Irre, die Fähigkeit zu sprachlicher Verständigung als dem Verfügen über eine empirische Theorie entsprechend hinzustellen.

Ein angemesseneres Bild der Lage erhalten wir, so der Gegenvorschlag, wenn wir zunächst davon ausgehen, dass das durch eine sprachliche Praxis etablierte Sinnpotential komplexe Verhältnisse der Möglichkeit der Übernahme oder Zuschreibung von Verantwortung konstituiert. Zur besonderen kognitiven Struktur dieser Praxis und den zur Partizipation erforderlichen Kompetenzen habe ich ja oben schon einige Hinweise gegeben. Diese, im Standardfalle an sprachliche Fähigkeiten gebundenen, kognitiven Kompetenzen bilden eine wesentliche Bedingung für die Möglichkeit nicht nur kognitiver, sondern auch sozialer oder moralischer Verantwortung: Wir können Menschen erst dann als Personen wirklich ernst nehmen, wenn sie über ein gewisses Maß an situationstranszendierender (Re-) Identifikations- und Unterscheidungskompetenz, einem dynamischen Vermögen zur Differenzierung von Sinn und Bezug und, damit einhergehend, an materialen und formalen inferentielle Fähigkeiten, an Zweck-Mittel-Rationalität und allgemeiner Vernunft (wie etwa Einsichtsfähigkeit fürs bessere Argument, ein Sinn für Angemessenheit) verfügen. Ein angemessenes Verständnis der Bedingungen jener unterschiedlichen Facetten möglicher Verantwortbarkeit sprachlicher Wesen, die sich wechselseitig als Personen anerkennen, erfordert also, wie ich plausibel machen will, eine besondere Berücksichtigung der kognitiven Funktion der Struktur von Propositionalität, die unter den epistemischen und methodologischen Voraussetzungen wahrheitssemantischer Analysen im Stile Davidsons, bzw. naturalistischer Theorien ganz allgemein, gar nicht in den Blick geraten kann.

⁸⁰ Zu diesem Unterschied, vergl. M. Dummett: ‚Was ist eine Bedeutungstheorie?‘ (1975).

Ich habe vorgeschlagen, Verantwortbarkeit, also die Möglichkeit der ‚Haftung‘ im Rahmen von normativen Verhältnissen, in einen intersubjektiven und einen objektiven Sinn zu differenzieren, nämlich als Unterschied zwischen einer Verantwortbarkeit aufgrund sozial instituiertes, lokal geltender Normen einerseits und einer Verantwortbarkeit auf der Grundlage von solchen Normen, die eben nicht aufgrund sozialer Institutionen gelten (insofern zu ‚entdecken‘ sind) andererseits. Intersubjektiver Verantwortbarkeit unterliegt die ‚harte‘, empirische Realität von Personen oder Institutionen, die uns, im vollen Wortsinne, gegenübertreten und, bei Verstößen, bestrafen können. Objektiver Verantwortbarkeit unterliegt die ‚weiche‘ Idealität des Vernünftigen, an der sich die Richtigkeit, Konsistenz oder Angemessenheit unserer nicht nur individuellen, sondern auch konsensuellen Überzeugungen oder Handlungen (bzw. unserer Urteile, Begründungsversuche, Kontextverständnisse und ähnliches) bemessen lässt.

Während es für normative Orientierungen unter der Maßgabe *intersubjektiver Korrektheit* hinreichen kann, mit einer gewissen Zuverlässigkeit eigene Handlungen als den sozialen Erwartungen *gemäß* oder nicht einschätzen zu können, erfordern normative Orientierungen unter der Maßgabe ihrer der *objektiven Vernünftigkeit* eine Fähigkeit zur *kritischen* Beurteilung jeweiliger faktischer Handlungs- oder Urteilserwartungen. Um sozial geltende Normen, etwa im Hinblick auf bestimmte Begriffsverständnisse, in Frage zu stellen, müssen sich Individuen auf externe Instanzen berufen, also auf Maßstäbe, sich nicht jenen intersubjektiv bereits etablierten Normen entnehmen lassen. Insofern können neue Erkenntnisse nicht (immer) lediglich neu entdeckte *Konsequenz* des Alten sein.

Den starken Intuitionen, die der Unterscheidung zwischen intersubjektiver und objektiver Gültigkeit zugrunde liegen, gerecht zu werden, halte ich im Hinblick auf ein angemessenes Verständnisses dessen, was im Umfeld von Begriffen wie „Handlung“, „Bedeutung“, „Überzeugung“ und ähnlichem thematisch werden kann, für maßgeblich. Wie oben exemplarisch zu zeigen versucht wurde, müssen deshalb auch solche *nicht-naturalistischen* Ansätze als inadäquat beurteilt werden, die entweder den normativen Aspekt von Bedeutung und Geist neutralisieren oder aber objektive Richtigkeit (Berechtigung) auf der Basis *intersubjektiver* Verantwortbarkeit erklären wollen: Während Glüer Geist (‚Intentionalität‘) und Bedeutung, in gewisser Weise, nur unter dem Aspekt objektiver Gültigkeit betrachtet und dabei Normativität sowohl im Sinne intersubjektiver, als auch im Sinne objektiver Verantwortbarkeit ausklammern will, betont Brandom zwar den normativen Aspekt von Geist und Bedeutung, vermag es aber weder, den (im transzendentalen Sinne) konstitutiven Charakter einiger zentraler Vernunftregeln zu berücksichtigen, noch den Zusammenhang von intersubjektiver und

objektiver Verantwortbarkeit in einer überzeugenden Weise herzustellen. Daraus ergeben sich jeweils spezifische Schwierigkeiten, die wir diskutiert haben.

Um *beiden* Aspekten von Geltung und Verantwortbarkeit gerecht zu werden, sollten wir von Handlungen, propositionalen Einstellungen und Bedeutung nur dann reden, wenn wir deren Instantiierung durch Subjekte im Rahmen einer *sozialen* Praxis meinen, die wir so beschreiben, dass die durch diese Praxis – in einem kontingenten Sinne – konstituierten *normativen Verhältnissen strukturell* nicht nur die Unterscheidung zwischen Richtig und Falsch und (darüber hinaus) zwischen Sein und Schein *ermöglichen*, sondern auch *situations-* und *perspektivenübergreifende Übereinstimmungen* einerseits und *soziale Lernprozesse* andererseits. Nur dann können wir nämlich letztlich das wenig erbauliche Trilemma von Platonismus, Relativismus oder Naturalismus im Hinblick auf Bedeutungen und Überzeugungen vermeiden.

Nun wird von Seiten naturalistischer Ansätze allerdings der, hier bereits mehrfach ins Spiel gebrachte, gemeinsame Sprachgebrauch als für die Möglichkeit propositionalen Gehaltes, mitunter sogar auch als für die Möglichkeit sprachlicher Verständigung, ‚sekundär‘ betrachtet. Ich werde dagegen, insbesondere in der Auseinandersetzung mit der Position Davidsons, argumentieren, dass ein als verbindlich verstandener, gemeinsamer – oder doch, gegebenenfalls, als gemeinsamer möglicher – Sprachgebrauch die einzig plausible Grundlage für die *Möglichkeit intersubjektiver Übereinkunft hinsichtlich übernommener oder zugeschriebener Verantwortung* bildet. Die Übernahme oder Zuschreibung von *Verantwortung* muss *allgemein spezifizierbar* und *beurteilbar* sein. Das setzt die Möglichkeit der *öffentlich kontrollierbaren* Gemeinsamkeit in Richtigkeits- oder Gültigkeitsurteilen voraus. Diese Funktion kann aber offenbar nur eine *wechselseitig verbindliche Zeichenverwendung*, als *intersubjektiver und situationstranzendenter Maßstab der Gleichheit*, erfüllen⁸¹.

Da nun dieses Argument auch im Hinblick auf Zuschreibungen propositionaler Einstellungen gilt, kann deren Verhältnis zu sprachlicher Bedeutung nicht so bestimmt werden, wie in naturalistischen Theorien weitgehend üblich. Die Zuschreibung von entsprechenden Einstellungen ist dann nämlich nur gegenüber solchen Subjekten im vollen Sinne gerechtfertigt, die an einer sprachlichen Praxis teilhaben, welche die

⁸¹ Die Integration einer Form *sozialer* Praxis, die nur auf der begrifflichen Ebene von Verantwortbarkeit angemessen beschrieben werden kann, setzt eine *allgemeine* Beurteilbarkeit richtigen und falschen Tuns voraus. Dazu aber, so werde ich argumentieren, bedarf es eines *Mediums*, das durch Eigenschaften wie *Öffentlichkeit*, *Standardisierbarkeit* und *Reproduzierbarkeit* jene allgemeine Beurteilbarkeit erst hergibt. Dadurch wird, über die Unterscheidung von Sinn und Referenz (bei einem richtigen Verständnis dieser Unterscheidung), eine, bezüglich jeweils *situationsrelativer Perspektiven von Akteuren situationsunabhängige* (nicht an ‚Ich-Du-Jetzt-Hier-Situationen‘ gebundene), aber allgemein verbindliche (sich an jeden potentiellen Adressaten richtende) *Spezifizierungen wechselseitiger Erwartungen*, bzw. gemeinsamer *Handlungssituationen*, möglich.

Spezifizierungen von und Stellungnahmen zu Verantwortbarkeiten in *allgemein kontrollierbaren Verständigungssituationen* ermöglicht. In diesem Sinne ist propositional oder begrifflich qualifizierter Geist als derivativ zu einer geteilten, sprachlichen Praxis zu verstehen – und bildet nicht umgekehrt deren Grundlage⁸².

Wesentliches Ziel dieser Arbeit ist es also, interne Zusammenhänge zwischen der Propositionalität von Verständigungsmöglichkeiten und Einstellungszuschreibungen, einem als allgemein verstandenen Sprachgebrauch, unserem intuitiven Verständnis objektiver Richtigkeit (bzw. Erkenntnis) und den beiden Formen von Verantwortbarkeit aufzuzeigen. Die Thesen, für die ich in der Folge in Auseinandersetzung mit naturalistischen Positionen argumentieren werde, sind, zusammengefasst, (1.) dass propositional spezifizierte Einstellungen, bzw. die Möglichkeit von Überzeugungen, Wissen, Wünschen, Absichten, usw. in ihrem primären Sinne eine bestimmten Form sprachlicher Verständigungspraxis voraussetzen; (2.) dass die Subjekte einer entsprechenden Form sprachlicher Verständigungspraxis gemeinsam verbindliche Begriffs- und Bedeutungsverständnisse – zumindest idealisierend – unterstellen müssen; und (3.), dass eine Rekonstruktion der Bedingungen einer solchen Praxis als einer, die dem kognitiven Potential der Propositionalität, also primär der Wahrheitsfähigkeit und Inferentialität von Gehalten im Sinne der Möglichkeit intersubjektiver Überprüfbarkeit von mit Behauptungen durch Subjekte mit situations- oder interessenrelativen Perspektiven verknüpften Wahrheitsansprüchen (und deren Konsequenzen), aber auch der Möglichkeit von Wissensfortschritt, gerecht wird, in einer Begrifflichkeit des kategorialen Rahmens von Verhältnissen intersubjektiver und objektiver Verantwortbarkeit erfolgen muss.

Das im einzelnen noch aufzuzeigende Scheitern naturalistischer Theorien führe ich also im Grundsatz darauf zurück, dass die unterschiedlichen Phänomene ‚intentionalen‘ Charakters nur auf der Grundlage einer Verständigungspraxis, die intrinsisch normativ verfasst ist, also im kategorialen Rahmen einer Begrifflichkeit von Stellungnahme und Verantwortbarkeit gedacht werden muss, adäquat verstanden werden können. Dabei müssen wir aber einerseits im Auge behalten, dass die vorgeschlagene Unterscheidung zwischen intersubjektiver und objektiver Verantwortbarkeit eine ‚analytische Idealisierung‘ der tatsächlichen Verhältnisse zu systematischen Zwecken darstellt. Dies

⁸² Ein Aspekt dieser Umkehrung, betrifft das naturalistische Verständnis propositionaler Zuschreibungen. Nach meiner Konzeption können solche Zuschreibungen erst sekundär – und nur in einem sehr begrenzten Sinne, als Einsetzungen in Schemata subjektbezogener, Erklärungen oder Prognosen (über zukünftiges Verhalten) gedeutet werden. Primär entsprechen sie einer *Einordnen von Personen und ihren möglichen oder tatsächlichen Handlungen in ein normatives Muster*, das einerseits (sozusagen auf der ‚empirischen Seite‘) zwar so etwas wie *Einschränkungen der Bandbreite möglicher Handlungserwartungen* ermöglichen soll, andererseits aber vor allem eine *Spezifizierung von Verantwortlichkeiten* erlaubt.

gilt, wie gesagt, insbesondere für begriffliche- und Bedeutungsregeln. Und andererseits sind auch die Formen normativer Orientierung, die vorrangig unter dem Aspekt intersubjektiver Verantwortung beschrieben werden können, auf Vernunftkompetenz angewiesen (und schon deshalb nicht auf statistische Verhaltenserwartungen reduzierbar). Auch die Verwendung oder Beurteilung der Verwendung eines sprachlichen Ausdruckstyps im Sinne intersubjektiver Verantwortbarkeit, also im Sinne der vorrangigen Orientierung an einem etablierten Sprachgebrauch, bedarf einer vernünftigen Urteilskompetenz hinsichtlich der Frage, ob oder wie, im Hinblick auf jenen Gebrauch, in relevanter Hinsicht das Gleiche getan wurde bzw. zu tun ist. Vernunft kommt also auch im Bemühen um die Erfüllung der normativen Erwartungen anderer immer schon ins Spiel, nicht erst, wenn wir – gewissermaßen gegen die mehr oder minder ‚eingespielten‘ normativen Erwartungen – die jeweils geltenden Standards kritisieren, bzw. modifizieren⁸³. Allerdings verschiebt sich im Falle der Kritik oder Modifikation unter Aspekten objektiver Verantwortbarkeit sozusagen die ‚Betonung‘: weg von jeweiligen gebrauch-internen Maßstäben des Vernünftigen (in relevanter Hinsicht gleiche Verwendung im Hinblick auf den üblichen Gebrauch, Widerspruchsfreiheit) hin zu allgemeinen, gebrauch-externen Maßstäben (etwa explanative oder systematische Fruchtbarkeit, umfassende begriffliche Kohärenz).

Ein wichtiges Argument für die These, dass die Begrifflichkeit von Bedeutung und Geist einem kategorial normativen Rahmen angehört, lautet demnach, dass jede Instantiierung eines sprachlichen Handlungstyps (auch ‚intentionale‘ Zuschreibungen), mit der ein Geltungsanspruch auf eine Form von Richtigkeit (etwa Wahrheit) erhoben wird, zugleich *eine implizite Stellungnahme* hinsichtlich *in relevanter Hinsicht gleicher Ausdrucksverwendung* darstellt. Dieser, dem Regelbegriff eigene, normative ‚Kick‘ ist nur aus einer Perspektive verantwortbarer Subjekte, bzw. aus einer Teilnehmerperspektive von ‚Anwendern‘ oder ‚Beurteilern‘ zugänglich und macht den eigentlichen Unterschied zu Regelmäßigkeiten aus. Aus der Teilnehmerperspektive stellt sich gewissermaßen in jeder Situation der Instantiierung einer Regel aufs Neue die Frage, wie fortzufahren ist, etwas der Regel entsprechend als gleich behandelt oder beurteilt werden sollte. Unsere entsprechende *Urteilspraxis* ist *regelkonstitutiv* in dem Sinne, dass mit jeder intersubjektiv *akzeptierten Stellungnahme* in Form von tatsächlichen praktischen Instantiierungen oder deren Beurteilung erst bestimmt wird, worin die Regel besteht. Dies ist nicht nur etwas anderes, als für eine Menge von

⁸³ So kann man sich etwa für eine richtige Verwendung des Ausdruckes „Fisch“ an einem allgemeinen Gebrauch orientieren, der Wale und andere unter Wasser lebende Tiere zur Extension dieses Begriffs zählt. Auch hier stellte sich dann (unabhängig von der Kritizierbarkeit dieses Gebrauchs anhand von gebrauch-externen Vernunftmaßstäben) die Frage nach der richtigen Verwendung des Ausdrucks in neuen Fällen.

Ereignissen oder Zuständen die Frage zu stellen, welcher Regelmäßigkeit oder Gesetzmäßigkeit sie unterliegen, bzw. welche induktive Verallgemeinerung die richtige oder angemessene ist. Es ist auch etwas anderes, als, aus einer reflexiven Perspektive auf unsere Handlungspraxis, die Frage zu stellen, an welcher Regel sich eine bestimmte Menge von Handlungen orientiert⁸⁴. Aus der Perspektive solcher Fragestellungen unterstellen wir die der *Beschreibung* der Regel oder Regelmäßigkeit *zugrundeliegenden* Regelverständnisse ‚objektivierend‘ – bzw. platonistisch – als gegeben und fragen uns, welche Regel zu einer bestimmten Reihe von Handlungen oder auf bestimmte Aspekte einer Reihe natürlicher Vorkommnisse passt. Regelinterpretationen oder -anwendungen aus der Teilnehmerperspektive, die für das konstitutiv sind, was wir unter einer bestimmten begrifflich-materialen Regel verstehen, entsprechen dagegen Urteilen im Hinblick auf relevante Gleichheit (bzw. etwas in derselben Weise zu tun), sie sind nicht gleichsam ‚mechanisch‘ vorgegeben, etwa wie – in Wittgensteins Bild – Gleise, die den Weg richtiger Ver- bzw. Anwendung durch die Raum-Zeit vorab eindeutig bestimmen, und auf denen wir einfach ‚entlangfahren‘ müssen, wenn wir die Regel richtig anwenden.

Dies gilt offensichtlich erst recht dann, wenn wir auch den Aspekt objektiver Verantwortbarkeit für sprachliche Äußerungen berücksichtigen. Jede Regel- oder Normkritik, bzw. jede begründete Modifikation (obwohl es hier zur Regelinterpretation fließende Übergänge geben mag) ist eine *explizite Stellungnahme* hinsichtlich eines vernünftigen Verständnisses unserer begrifflichen Verhältnisse, also etwa der Extension eines Begriffs. Ich werde dafür argumentieren, dass das Wissen um die *Möglichkeit* solcher Modifikationen Teil unserer begrifflichen- bzw. Bedeutungskompetenz ist. Dieser Aspekt von Bedeutung – sozusagen eine ‚Normativität zweiter Stufe‘ – lässt sich aus der deskriptivischen Perspektive naturalistischer Theorien prinzipiell nicht einholen, da die Möglichkeit der Berufung auf praxis-externe Maßstäbe des Vernünftigen eine *wertende* Komponente hat, die gewissermaßen von Haus aus auf die Möglichkeit der Transzendierung sozialer *Praktiken* verweist. Anders ausgedrückt, es kann keinen endgültigen ‚Algorithmus der Vernunft‘ geben, den wir gleichsam ‚mechanisch‘ auf vernünftige Individuen oder soziale Praxis anwenden könnten⁸⁵.

Es ist demnach auch, entsprechend dem im vorletzten Absatz Gesagten, in keinem naturalistisch verwertbaren Sinne ‚von sich her‘ eindeutig bestimmt, ob wir es jeweils mit einem neuen oder anderen Begriff zu tun haben, bzw. ob ein Ausdruck seine

⁸⁴ Was etwa der Form entspricht, in der Kripke Wittgensteins sogenanntes ‚Regelfolgeproblem‘ diskutiert.

⁸⁵ Was damit genauer gemeint ist, wird unten, in der Kritik des Funktionalismus, deutlich werden

Bedeutung geändert hat. Wir sind es, die gegebenenfalls Bedeutungsidentität oder Bedeutungswandel jeweils nach vernünftigen Kriterien (etwa der interessenrelativen Zweckmäßigkeit) bestimmen müssen. Bedeutungen und begriffliche, bzw. propositionale Gehalte sind nichts, was als beschreibbare, irgendwie in der Welt realisierte Eigenschaften verstanden werden können, denn ihr ‚Inhalt‘ hängt von unseren normativen Einstellungen ab.

Dass wir häufig auch Dinge oder Vorgänge in einem ‚intentionalen‘ Vokabular beschreiben, die den hier genannten Anforderungen an Verantwortbarkeit offensichtlich nicht gerecht werden, kann Anlass zur Skepsis gegenüber der vorgeschlagenen Dichotomie von Verantwortbarkeit und Geschehnishaftigkeit geben. Dabei gilt es aber zunächst einmal zu berücksichtigen, dass mögliche Fragen nach einer explanativen Angemessenheit von ‚intentional aufgeladenen‘ Beschreibungen von Naturvorgängen von Fragen danach unterschieden werden müssen, ob ‚intentionale‘, im ‚Reich der Gründe‘ angesiedelte, bzw. auf Verantwortbarkeit verweisende, Phänomene in einer Terminologie von Naturdingen oder – Vorgängen beschrieben oder erklärt oder doch zumindest als mit natürlichem Geschehen in irgendeinem verständlichen Sinne ereignis- oder tokenidentisch gedeutet werden können. Letzteres zurückzuweisen, heißt offensichtlich nicht zugleich, auch ersteres abschlägig zu beurteilen. Es wurde oben ja bereits darauf hingewiesen, dass wir uns kaum zu tilgender Formen anthromorphisierenden Redens über die Natur bedienen. Solche Redeweisen dienen offenbar schon common-sensisch meist dem *anschaulichen Verstehen* von Naturvorgängen. Dies gilt jedoch ebenso für wissenschaftliche Vokabulare, insbesondere in der Biologie. Aber auch die Physik bedient sich in kreativer Weise anthromorpher Metaphern, etwa im Sinne ursprünglich handlungsbezogener Begrifflichkeiten, um Erklärungen von Naturvorgängen zu veranschaulichen (wenn beispielsweise vom „Anziehen“ und „Abstoßen“ gegensätzlicher elektrischer Ladungen gesprochen wird). Dies ist wohl unproblematisch, solange man die Begriffe zu den jeweiligen Zwecken technisch definieren, bzw. quantitativ oder operational bestimmen kann, und sich des veranschaulichend-metaphorischen, nicht originär sinngebenden Charakters ihrer Verwendung in wissenschaftlichen Kontexten bewusst bleibt. Genau an diesem Bewusstsein aber scheint es, wie wir sehen werden, naturalistischen Positionen typischer Weise zu mangeln.

Auf der anderen Seite kann auch Sinnhaft-Strukturiertes, also ‚kulturell überformte‘ menschliche Aktivitäten, in einem eingeschränkten Sinne, als Geschehen verstanden

werden kann. Dies ist ja bereits dann der Fall, wenn etwa von allgemeinmenschlichen Neigungen zu bestimmten Einstellungen oder Verhaltensweisen die Rede ist. Eine entsprechende Perspektive auf menschliches Verhalten bildet eine wesentliche Voraussetzung soziologischer Verallgemeinerungsversuche. Dabei wird entweder die Zuschreibung von Verantwortbarkeit im objektiven Sinne ausgeblendet, in dem als Tendenz, bzw. statistisch erfassbare Regel- und Gesetzmäßigkeiten hinsichtlich Einstellungen oder Handlungsorientierungen von Subjekten erfasst oder strukturelle Eigenschaften von sozialen Phänomenen beschrieben werden, die sozusagen hinter dem Rücken der Akteure wirken⁸⁶. Oder wir blenden objektive *und* intersubjektive Verantwortbarkeit in der Beschreibung struktureller sozialer Regelmäßigkeiten ganz aus oder suspendieren diese in Fällen des Mangels an oder der von Störung von autonomer Handlungskompetenz. In solchen Fällen haben wir es mit einer Perspektive auf Subjekte zu tun, bei der vernunftorientierte, moralische, sittliche oder rechtliche Bewertbarkeiten von Handlungen im Extremfall überhaupt keine mehr Rolle spielen, sondern Verhalten (individuelles oder von Gruppen) als Instantiierung eines Verhaltensmusters beschrieben oder erklärt wird.

Dazu will ich hier nur so viel sagen, dass solche, *von objektiven oder intersubjektiven Verantwortbarkeitszuschreibungen befreiten* Sichtweisen auf ‚Vorgänge‘ im Bereich des sinnstrukturierten Teils der Wirklichkeit die normative Begrifflichkeit des kategorialen Rahmens von Verantwortbarkeit am Ende *nicht substituieren* können⁸⁷: Wir würden diese Wirklichkeit nämlich in entsprechenden Beschreibungen gar nicht als unsere soziale Praxis *wiedererkennen*, wäre etwa eine systemische Beschreibung des Sozialen nicht letztlich durch nur aus einer ‚Teilnehmerperspektive‘ von Zuschreibern und Adressaten möglicher Verantwortbarkeit verständliche Begriffe (wie z.B. Schuld, Wunsch, Wissen oder Grund) vermittelt. Solche Begriffe und eine in ihrem kategorialen Rahmen beschriebene Wirklichkeit zu verstehen heißt, sich selbst als potentiellen Adressaten einer entsprechenden Verantwortbarkeitszuschreibung, bzw. als Subjekt einer Verantwortungsübernahme wahrnehmen zu können. Dass es zwischen den Perspektiven der Ereignishaftigkeit und Verantwortbarkeit (im starken und schwachen Sinne) in der Praxis ‚fließende‘ Übergänge oder unklare Fälle geben mag – man denke etwa an unsere ambivalenten Einstellungen gegenüber primitiven Kulturformen (bzw. primitiven Zügen unserer eigenen) oder zu Verhaltensweisen von Kleinkindern, Menschen in emotionalen Ausnahmesituationen, psychisch Schwerkranken oder geistig Behinderten – , ändert nichts an deren systematischer Differenzierbarkeit.

⁸⁶ Eine bekannte Metapher hierfür ist das von Adam Smith sogenannte *Walten einer ‚invisible hand‘*, aber gemeint sind z. B. auch alle Gemeinsprüche über allgemeinmenschliche Neigungen

⁸⁷ Vergl. hierzu: P. F. Strawson (1962) und (1985/1987).

Teil II REDUKTIVER MATERIALISMUS

Als gemeinsamen Nenner all jener Theorien der Bedeutung und des Geistes, die hier als „naturalistisch“ bezeichnet werden sollen, hatte ich zunächst die grundsätzliche Voraussetzung bestimmt, unsere ‚intentionalen‘ Redeweisen, also Äußerungen über die Bedeutung von Äußerungen oder Sätzen und vor allem die Zuschreibung von propositionalen Einstellungen oder Handlungsabsichten, als *empirische Aussagen über bestimmte Arten von Ereignissen oder Zuständen zu verstehen*. Dementsprechend wird unsere Kompetenz, sprachliches und nicht-sprachliches Verhalten zu verstehen, meist grundsätzlich als eine Fähigkeit (oder einer solchen entsprechend) gedeutet, adäquate empirische Theorien zur Erklärung oder Prognose von Verhaltensereignissen auf der Grundlage der Kenntnis entsprechender kausaler Zusammenhänge mit mentalen Zuständen oder Ereignissen zu bilden⁸⁸.

Darüber hinaus gibt es allerdings eine ganze Menge an Unterschiedlichkeit. Ein zentraler, aber doch nicht von allen Naturalisten geteilter Anspruch besteht darin, Erklärungen, die mit ‚intentionalen‘ Begriffen arbeiten, reduktiv⁸⁹ in etablierte naturwissenschaftliche Erklärungsweisen integrieren zu können. Das ausschlaggebende Argument hierfür ist, dass der offenbare ‚*prognostische Erfolg*‘ der *Alltagspsychologie* (nach dem Schema des praktischen Syllogismus) eine *reale physische Grundlage* haben müsse, um nicht schlichtweg mysteriös zu erscheinen. Diesem Argument liegt seinerseits die gängige Unterstellung zugrunde, dass alle Zustände und Vorgänge in der Welt durch einen ‚kausalen Kitt‘ verbunden sind, der lediglich physischer Natur sein kann.

Zusammenfassend verbergen sich hinter der allen naturalistischen Theorien *gemeinsamen* Ausgangsbasis einer empirischen Theoretisierbarkeit von Bedeutung und Geist^{9E} mindestens folgende Annahmen:

1. ‚Intentionale‘ Zuschreibungen dienen – im Rahmen der Anwendung des Schemas des praktischen Syllogismus – der (kausalen) Erklärung, bzw. der Vorhersage von Handlungen (bzw. Verhalten).

⁸⁸ Die Funktion dieser Fähigkeiten kann dann etwa so verstanden werden, dass sie sozial organisierten Wesen, wie dem Menschen, gegebenenfalls ein im Sinne der individuellen oder gemeinschaftlichen Bedürfnisbefriedigung vorteilhaftes Verhalten ermöglicht.

⁸⁹ In welchem Sinne „reduktiv“ zu verstehen sein kann, wird natürlich noch zu erläutern sein. Die Verwendung dieses Begriffs in der einschlägigen Literatur ist jedenfalls uneinheitlich.

2. Die Bedeutung von Äußerungen oder sprachlichen Gebilden wird durch propositionale Einstellungen und Mitteilungsententionen von Individuen generiert. Sprachhandeln oder -verhalten zu verstehen, heißt, über eine adäquate empirische Theorie der Bedeutungsintentionen eines Sprechers bzw. seiner propositionalen Einstellungen zu verfügen. (Konventionelle Regelungen sprachlichen Symbolgebrauchs sind immer als derivativ zu ursprünglichen subjektiven Bedeutungsintentionen anzusehen.)

3. Die physische Welt ist kausal geschlossen. Zustände und Ereignisse unterliegen der Kausalität nur *qua* physisch zu verstehender Eigenschaften.

Mir geht es um eine kritische Auseinandersetzung mit den Annahmen (1.) und (2.), wobei der Schwerpunkt auf (2.) liegen wird. Dagegen soll (3.) nicht bestritten werden. In diesem Kapitel werde ich mich zunächst den im Anspruch stärkeren, das heißt, den reduktiven Theorien des DOGNA zuwenden. Den Hintergrund für diesen stärkeren Anspruch bildet das (von Stephen Schiffer so genannte) „*Intention-Based-Semantics-Program*“ (IBS).

Nach IBS soll sprachliche Bedeutung – verstanden als eine Eigenschaft, die primär physischen Token, wie Lautfolgen oder Linien auf Papier, zukommt – auf ‚psychologisch-intentionale‘ Eigenschaften von intern entsprechend komplex organisierten, physischen Systemen reduziert werden können. Diese Eigenschaften sollen dann ihrerseits irgendwie auf rein physisch beschreibbare Eigenschaften eines betreffenden Systems *reduzierbar*, bzw. durch diese *erklärbar* sein. Das IBS-Programm ist demnach immer mit der ein oder anderen Form des *Reduktiven Materialismus* verbunden. Da reduktiv ambitionierte Theorien, die sich im Rahmen von IBS bewegen, nicht nur als die anspruchsvollste, sondern zugleich auch als die einflussreichste Form des naturalistischen Projektes anzusehen sind, wird ihnen in der Folge auch der meiste Raum gewidmet. Davon profitiert jedoch auch das Verständnis der anderen Positionen.

Dass IBS das anspruchsvollste Projekt ist, lässt sich anhand einer Formulierung Fodors verdeutlichen:

„If the Gricean program (d. h. die Rückführbarkeit ‚semantischer‘ auf ‚intentionale Fakten‘ /W.-J. C.) and the Naturalisation program can be carried through, then IBS will have solved one of the Great Metaphysical Problems: it will have found a place for meaning in the natural order.“ (Fodor 1989, S. 177)

Dass, was Fodor hier mit „einem der Großen Metaphysischen Probleme“ meint, nämlich *Bedeutung* irgendwie in der natürlichen Welt ‚unterbringen‘ zu können, wäre demnach dann gelöst, wenn ‚semantische Tatsachen‘ in nicht-zirkulärer Weise auf propositionale Einstellungen (inklusive Mitteilungsabsichten) von Sprechern im Sinne ‚psychologischer Tatsachen‘ zurückführbar wären und solche Tatsachen dann ihrerseits ohne Rückgriff auf eine ‚intentionalistische‘ Begrifflichkeit, also rein physisch, verstanden und in unser naturwissenschaftliches Weltbild integriert werden könnten. Mir wird es vorrangig um eine kritische Auseinandersetzung mit Überlegungen zur Möglichkeit der Durchführung jenes zweiten Schrittes gehen, den Fodor das „Naturalization program“ nennt. Um aber die wesentlichen Intuitionen des IBS-Programms im Ganzen nachvollziehbar zu machen, müssen wir immerhin klären, welches Verständnis von ‚Bedeutung‘ von Vertretern von IBS als grundlegend angesehen wird. Denn ein bestimmtes, im Hinblick auf eine intentionalistische Analyse im Stile Gricens augenscheinlich präjudizierendes Verständnis dieses Begriffs – nämlich im Sinne des doppeldeutigen englischen ‚*to mean something*‘ (was ja bekanntlich sowohl als ‚etwas (mit etwas) meinen‘ oder als ‚etwas bedeuten‘ übersetzt werden kann) –, ist, wie ich denke, eine der wesentlichen Voraussetzungen des Programms einer physischen Reduktion von Bedeutung in der oben angedeuteten Form. Andererseits können mögliche Schwierigkeiten in der Durchführung des zweiten Reduktionsschrittes, also einer Reduktion oder Erklärung ‚intentional-psychologischer Fakten‘ oder ‚Eigenschaften‘ auf bzw. durch physische, auch Konsequenzen für die Plausibilität eines intentionalistischen Bedeutungsverständnisses haben.

Bevor wir uns aber auf eine Auseinandersetzung mit einflussreichen Vorschlägen für materialistische Verständnisse propositionaler Einstellungen bzw. Gehalte einlassen können, müssen, neben der intentionalistischen Analyse des Bedeutungsbegriffs, auch noch andere wichtige Fragen diskutiert oder zumindest angeschnitten werden – schon um die Bedingungen der Möglichkeit jenes zweiten reduktiven Schrittes besser verstehen zu können. Dazu zählt zunächst die Frage, wodurch sich Kausalerklärungen im Allgemeinen und wissenschaftlichen Erklärungen im Besonderen auszeichnen. Die Unterstellung einer *kausalen Erklärbarkeit von Verhalten* ist ja die methodologische Ausgangsbasis naturalistischer Theorien des Geistes^{PE}. Vor allem aber muss sondiert werden, was von dem weitergehenden Anspruch auf eine *reduktive Erklärung des Geistes* bzw. von ‚psycho-physischen‘ Identifizierungsstrategien eigentlich zu erwarten wäre. Grundgedanken dieses weitergehenden, reduktiven Anspruches werden daraufhin insbesondere anhand des Computermodells des Geistes, also der Annahme, dass ‚Semantische Maschinen‘ als ‚Syntaktische Maschinen‘ realisiert sein können, diskutiert. Diese Erörterungen liefern schließlich die Grundlage für ein besseres

Verständnis der unterschiedlichen Varianten einer, im weitesten Sinne, *kausalistischen* Analysen begrifflichen, bzw. propositionalen Gehaltes.

Ich werde dabei versuchen, den Reduktiven Materialismus (bzw. IBS), zumindest bezüglich einiger grundsätzlicher Überlegungen, als so etwas wie ein einheitliches Projekt zu darzustellen. Dafür kommt es nicht auf eine umfassende Wiedergabe einzelner Positionen an. Dies wäre den hier verfolgten grundlagenkritischen Absichten auch eher abträglich und würde zu einer (noch weiter gehenden) Aufblähung dieser Arbeit führen. Ich werde also Abweichungen, die mir für meine als eher grundsätzlich beabsichtigten Einwände gegen die Möglichkeit einer rein deskriptivischen bzw. empirischen Herangehensweise an das Phänomen von Bedeutung und Geist nicht wesentlich erscheinen, vernachlässigen und die entweder einflussreichsten oder jeweils am plausibelsten anmutenden Vorschläge herauspicken⁹⁰. Sowohl die folgenden, allgemeineren, als auch die anschließenden, spezifischeren Überlegungen zum IBS-Programm und zum Reduktiven Materialismus haben aber auch hinsichtlich eines besseren Verständnisses der anderen, später zu diskutierenden Strategie, also des Interpretativen Materialismus, einen klärenden Sinn, und werden deren Einordnung innerhalb des hier als naturalistisch abgesteckten Terrains erleichtern.

⁹⁰ Eine Schwierigkeit dieser Vorgehensweise besteht darin, naturalistische Argumentationslinien zu rekonstruieren, die sich nur an der Frage nach der Naturalisierbarkeit propositionalen Gehaltes orientieren: So werden innerhalb der einschlägigen Debatten um die Möglichkeit ‚psycho-physischer‘ Identifikation die Beispiele für „mentale Phänomene“ oft wechselweise so gewählt, dass sie der jeweils vorgeschlagenen Strategie entgegenzukommen scheinen. Vor allem bleibt meist völlig unklar, in welchem Zusammenhang das gern diskutierte Beispiel der Empfindungen, wie etwa Schmerzen, mit propositionalen Einstellungen stehen soll.

1 ‚Intention Based Semantics‘: Der Versuch einer Reduktion von Semantik auf ‚materialistische Psychologie‘

Es ist eine durchaus bemerkenswerte Tatsache, dass viele der Beiträge, die sich an der neueren Debatte um ein naturalistisches Verständnis des Geistes^{PE} beteiligen, auf das Thema ‚sprachliche Bedeutung‘ erst gar nicht mehr eigens eingehen. Grund hierfür scheint eine stillschweigende, weitgehende Akzeptanz eines grundsätzlich ‚intentionalistischen‘ bzw. sich auf die einschlägigen Griceschen Überlegungen stützenden Verständnis der Bedingungen sprachlicher Bedeutung zu sein. Dabei wird es oft als mehr oder minder selbstverständlich angesehen, dass Fragen aus dem Problembestand der vormals die philosophische Szene beherrschenden Sprachphilosophie nun im Rahmen der Philosophie des Geistes zu bearbeiten seien.

In seinem berühmten Aufsatz ‚Meaning‘⁹¹, hatte Grice bekanntlich ‚Bedeutung‘ zunächst von ‚natürlicher Bedeutung‘ (im Sinne eines kausal bedingten Anzeichens, wie etwa dem, dass Rauch Feuer ‚bedeutet‘) unterschieden. Grundlage einer angemessenen Analyse des Begriffes (nicht-natürlicher) *sprachlicher* Bedeutung sei *nicht* – und dies ist bereits eine wichtige Vorentscheidung – die Wendungsform ‚X bedeutet P (bzw., dass p) in einer Sprache L‘, sondern ‚Sprecher S meint P (bzw., dass p) mit X (bzw., in dem er X äußert)‘. (Dass es sich hier überhaupt noch um denselben Bedeutungsbegriff handelt, mag allerdings im Englischen eher plausibel erscheinen, als beispielsweise im Deutschen, da dort die Unterscheidbarkeit der Bedeutungen von ‚(to) mean(s)‘, in ‚(mit etwas) etwas meinen‘ und ‚bedeuten‘, nicht wie im Deutschen, durch unterschiedliche Worte gekennzeichnet ist.)

Grice gibt damit der ‚Gelegenheits-, oder ‚Situativen-Bedeutung‘, also derjenigen (wörtlichen) Bedeutung, die ein Sprecher seiner Äußerung gewissermaßen ‚gibt‘, einen rekonstruktionslogischen Vorrang gegenüber der allgemeinen oder konventionellen Bedeutung von sprachlichen Zeichen⁹². Mit Stephen Schiffer⁹³ – einem vormaligen Grice-Schüler – können wir diese Unterscheidung durch die Ausdrücke ‚Sprecherbedeutung‘ (*speaker-meaning*) und ‚Ausdrucksbedeutung‘ (*expression-meaning*) kennzeichnen und uns in der Folge auch an dessen prägnant formulierter Beschreibung der wesentlichen Ziele des IBS-Programmes orientieren:

„Its hallmark is that it seeks to define all public-language semantic notions in terms of propositional-attitude concepts that themselves presuppose nothing about meaning in a public language. If the program is correct, then

⁹¹ Grice (1957).

⁹² Grice spricht von „*occasional meaning*“ und „*timeless meaning*“

⁹³ S. Schiffer (1987), Kap. 9.

- (1) *speaker-meaning* – essentially a persons meaning that such-and-such is the case or that so and so is to do such and such - is noncircularly defineable as a certain kind of behaviour done with the intention of activating belief or action in another;
and
- (2) *expression meaning* – essentially, the semantic features of natural language expressions - is noncircularly defineable as certain kinds of correlations between marks and sounds and types of acts of speaker meaning.“ (a. a. O., S. 242)

Bevor wir auf diese Punkte genauer eingehen, ist es interessant, zunächst einmal zu sehen, wie der Vorrang der (situativen oder) Sprecherbedeutung gegenüber (allgemeiner oder) Ausdrucksbedeutung eigentlich begründet wird. Schiffers Argument etwa lautet, dass jemand mit einer Äußerung, wie zum Beispiel der ‚Geste‘ eines Knurrelauts, etwas meinen kann (z.B. dass man ärgerlich ist), ohne dass diese Äußerung deshalb eine konventionelle Bedeutung haben muss⁹⁴. Nun wäre dies nach dem Griceschen Kriterium zwar ein Fall nicht-natürlicher Bedeutung, da man ja durchaus knurren kann, ohne tatsächlich ärgerlich zu sein, aber es erscheint doch fraglich, ob sich solche Fälle ohne weiteres im Hinblick auf unsere üblichen Verständnisse sprachlicher Verständigung verallgemeinern lassen. Sie entsprechen nämlich einer vergleichsweise eingeschränkten Form von *Ausdrucksverhalten*, welches prinzipiell auch ohne sprachliche Kompetenz im hier gemeinten Sinne auskommt und zudem instinktiven Ausdrucks- bzw. Kommunikationsformen gleicht oder ähnelt. Jonathan Bennett sieht ein wichtiges Argument für den genannten Vorrang in der (vermeintlichen) Tatsache, dass es bisher einfach niemandem gelungen sei, in überzeugender Weise die Wendung ‚was *x* bedeutet‘ zu erhellen, ohne auf ‚was *S* mit *x* meint‘ zurückgreifen zu müssen⁹⁵. Aber dieses Argument setzt natürlich voraus, dass *überhaupt* eine nicht-zirkuläre definitonische Analyse des Bedeutungsbegriffs möglich ist und wir ihrer bedürften, um im Hinblick auf Fragen im Zusammenhang mit dem Phänomen der Bedeutungshaftigkeit von Sprache weiterzukommen.

Doch wie dem auch sei, ein wesentlicher systematischer Grund für die Behauptung des Vorranges der Sprecherbedeutung vor der Ausdrucksbedeutung innerhalb des IBS-Programmes liegt offensichtlich in der Überzeugung, dass damit eher die Chance einer materialistischen Reduktion von ‚Intentionalität‘ gegeben sei⁹⁶. Sprecherbedeutung soll ja auf komplexe propositionale Einstellungen zurückführbar sein und solche Einstellungen lassen sich, so scheint es, umstandslos als *Zustände* von oder *Ereignisse* in Subjekten verstehen. Und ‚subjektive‘ (bzw. individuelle) Zustände oder Ereignisse

⁹⁴ S. Schiffer: *Meaning* (1972), S. 7.

⁹⁵ J. Bennett: *Linguistic Behaviour* (1976), § 3.

⁹⁶ Vergl. etwa Bennett (1976), Einleitung u. Kap. 1, sowie Loar (1981), Einleitung u. Kap. 1 und Schiffer (1987), Vorwort u. Kap. 1.

müssen ja, wie immer sie beschrieben sein mögen, letztendlich auch physischer Natur sein. Würde hingegen Ausdrucksbedeutung einen rekonstruktionslogischen Vorrang vor der Sprecherbedeutung eingeräumt, kämen vorrangig so etwas Konventionen, Regeln, Normen oder Sinn – und damit eine ‚intentionales‘ Vokabular – ins Spiel. Auf solche Phänomene ist die Begrifflichkeit von kausalwirksamen, raum-zeitlich individuierbaren und vielleicht rein physisch beschreibbaren Zuständen oder Ereignissen offensichtlich nicht ohne weiteres anwendbar. In diesem Sinne gibt es also sicherlich ein starkes ‚forschungslogisch‘ begründetes Motiv für die Präferenz interner, psychischer Zustände vor so etwas, wie (vermeintlichen) ‚platonischen Entitäten‘ bzw., vor etwas scheinbar Mysteriösem, wie der ‚normativen Kraft‘ der Bedeutung von Symbolen.

Aber es scheint auch ein plausibel klingendes, entwicklungsgeschichtliches Argument für den Vorrang subjektiver ‚Intentionalität‘ vor der intersubjektiven Bindekraft von Normen oder Konventionen zu geben, das von der Einnahme des Standpunktes des PHINA unabhängig ist: Können Gedanken, Wünsche oder Absichten nicht erst dann *mitgeteilt* werden, wenn es sie schon gibt? Musste es nicht *erst* Repräsentationen von der Welt geben, die Überzeugungen, Wünsche, Handlungs- bzw. Mitteilungsabsichten individuieren, *bevor* Kommunikation überhaupt einen Sinn machen konnte? Müssen sich nicht ganz generell kognitive Vermögen, wie Intelligenz, instrumentelle Rationalität und Repräsentationsfähigkeit, bereits bis zu einem gewissen Grade individuenintern entwickelt haben, bevor Wesen in der Lage gewesen sein können, sich im Bedarfsfalle symbolgestützt mitzuteilen oder zu verständigen?

Zu diesem Argument will ich hier nur anmerken, dass zwar sicherlich vieles dafür spricht, dass die Entwicklung bestimmter kognitiver Vermögen naturgeschichtlich der Möglichkeit sprachlicher Verständigung vorausgegangen sein muss. Doch daraus folgt noch keineswegs ein genealogischer Vorrang des Denken-Könnens vorm Bedeutung-Haben – oder, genauer gesagt: von begrifflich-propositional strukturierten Einstellungen vor begrifflich-propositionaler Bedeutung von Sprache. Meine kritischen Überlegungen richten sich nicht gegen die Möglichkeit einer empirischen Erforschung kognitiver Vermögen. Sie richtet sich vor allem gegen die Behauptung der Möglichkeit einer naturalistische ‚Erklärung‘ von Geist und Bedeutung, bzw. gegen die Behauptung, Geist und Bedeutung im Wesentlichen aus einer Perspektive empirischer Beschreibung verstehen zu können.

Ich werde an dieser Stelle jedoch noch nicht gegen den vom IBS-Programm behaupteten *rekonstruktionslogischen* Vorrang subjektbezogener ‚Intentionalität‘ gegenüber geteilten Bedeutungsverständnissen argumentieren. Hier kommt es zunächst nur darauf an, die analytische Ausarbeitung der grundlegenden Unterstellung des

intentionalistischen Modells nachzuzeichnen, nämlich, dass man prinzipiell unabhängig von der Voraussetzung allgemeiner Bedeutungen mit sprachlichen Ausdrücken etwas meinen kann. Denn diese Unterstellung bilden ja die Grundlage des reduktionistischen Programms im Ganzen⁹⁷.

Schiffer rekonstruiert nun die ursprünglich von Grice vorgeschlagene Analyse (wörtlicher) deklarativer Sprecherbedeutung wie folgt:

- „(S) *S* means that *p* in uttering *x* iff, for some person *A* and a feature Φ , *S* intends
- (1) *A* to recognize that *x* has Φ ;
 - (2) *A* to think, at least partly on the basis of thinking that *x* has Φ , that *S* uttered *x* intending *A* to think that *p*;
 - (3) *A* to think, at least partly on the basis of thinking that *S* uttered *x* intending *A* to think that *p*, that *p*.“ (a. a. O. S. 243)

Dabei wird nun nicht etwa verlangt, ein Sprecher müsse sich solcher Art komplexer Intentionen tatsächlich bewusst sein, wenn er etwas in kommunikativer Absicht sagt (meint). Schiffer versteht (im Anschluss an Loar) *S*' Intentionen als im Regelfalle stillschweigende Erwartungen. Für ein genaueres Verständnis der genannten Bedingungen muss aber offensichtlich die Eigenschaft Φ (von *x*), die einem Hörer Hinweis auf des Sprechers Intentionen sein soll, erläutert werden können. Worin könnte diese Eigenschaft bestehen, wenn wir uns nicht auf Fälle ‚ikonischer‘ Gesten, wie Knurren oder andere Formen der Nachahmung natürlicher Anzeichen, beschränken? Die auf der Hand liegende Antwort, nämlich, dass sie in der Bedeutung besteht, bedarf aus der Perspektive einer intentionalistischen Semantik schon deshalb einer besonderen Qualifizierung, weil sie uns ja hinsichtlich der Frage nach einem rekonstruktionslogischen Prioritätsverhältnis zwischen Sprecherbedeutung und Ausdrucksbedeutung nicht weiterbringt. ‚Bedeutung‘ bzw. die entsprechende Eigenschaft Φ von *x*, müsste aus bedeutungsintentionalistischer Perspektive daher vollständig in ‚psychologischen‘ Begriffen bestimmt werden können. Bedeutung sollen Äußerungen oder Sätze ja letztlich nur auf Grund der jeweiligen, ‚kommunikativen‘ Sprecherintentionen haben; die Eigenschaft Φ von *x* soll dem Hörer lediglich einen *Hinweis auf diese Intentionen* liefern und darf dabei, gemäß den Voraussetzungen, nicht selbst als von den jeweiligen Sprecherintentionen unabhängige

⁹⁷ Schiffer (1972) und (1987) Kap. 9. Ähnlich, Bennett (1976) und Loar (1981). Auch Searle (ders.: *Speech Acts* (1969); u. *Intentionalität. Eine Abhandlung zur Philosophie des Geistes* (1983/1987) bekennt sich explizit zu den Grundvoraussetzungen eines intentionenfundierte[n] Verständnisses der Semantik. Ich orientiere mich vorrangig an Schiffers jüngster, zusammenfassender Version (1987) (wenngleich diese bereits aus einer Perspektive kritischer Distanz geschrieben ist).

Ausdrucksbedeutung verstanden werden. So kommt Schiffer zu folgender Charakterisierung der Rolle von Φ :

„In fact, we may simply take it as definitive of IBS that, for every sentence σ of every natural language L , there is some feature Φ , constitutive of the meaning of σ in L , such that

- (a) Φ has a specification in wholly psychological terms, which include, in particular, those propositional attitudes deemed by the IBS theorist to be constitutive of speaker-meaning;
- (b) by virtue of having Φ , σ is an especially efficacious device among speakers of L for performing acts of speaker-meaning of a certain type;
- (c) σ 's having Φ entails that σ has whatever meaning σ happens to have in L .“ (a. O., S. 244)

Nach diesem Modell kommt ein Hörer also irgendwie zu der Überzeugung, dass σ die Eigenschaft Φ hat (welche die Bedingungen (a)-(c) erfüllt) und schließt daraus (zusammen mit anderen Überzeugungen) auf des Sprechers kommunikative Absichten. Das ‚irgendwie‘ wird nun von Griceianern als *wechselseitiges Wissen* von Sprecher und Hörer verstanden, dass Φ eine Eigenschaft ist, die eine Äußerung von σ in einer Sprachgemeinschaft *SG per Konvention* mit einem propositionalen Gehalt p verbindet⁹⁸

Den Versuch einer (naturalistischen) Ausarbeitung dieser Überlegungen hat David Lewis⁹⁹ unternommen. Lewis Bestimmung einer (sprachlichen oder nicht-sprachlichen) Handlung als durch eine Konvention(en) geregelt, läuft darauf hinaus, dass eine Konvention von den meisten Leuten befolgt wird und sich stabilisiert, weil die meisten Leute erwarten, dass die anderen sie auch befolgen, weil die es für vorteilhaft halten, sie zu befolgen, wenn die anderen es auch tun würden und sie davon ausgehen, dass andere genauso kalkulieren. Im speziellen Fall sprachlicher Handlungen beherrschen kompetente Teilnehmer einer Sprachgemeinschaft *SG* nach Lewis ein rekursives Verfahren, nach dem sich aus der (logischen) Grammatik der Sätze und den konventionell bestimmten Bedeutungen der Satzkonstituentien für jede mögliche Äußerung (strukturierte Lautsequenz) x der von *SG* gesprochenen Sprache L die jeweilige Äußerungsbedeutung ergibt. Dadurch könne einerseits ein kompetenter *Hörer* aus einer Äußerungsbedeutung die Sprecherbedeutung erschließen; andererseits sei ein kompetenter *Sprecher* befähigt, seine kommunikativen Intentionen erfolgreich

⁹⁸ Schiffer (1972) dehnt das IBS-Modell auf alle illokutionären Varianten von Äußerungsmöglichkeiten aus. Wir beschränken uns hier der Einfachheit halber auf assertorische Äußerungen.

⁹⁹ D. Lewis: *Conventions* (1972).

umzusetzen. Aus der Perspektive einer rationalen Rekonstruktion, halten sich Mitglieder einer Sprachgemeinschaft dann deshalb an sprachliche Konventionen, weil dies für sie zur Koordinierung nur gemeinsam zu realisierender, praktischer Absichten vorteilhaft ist und sie davon ausgehen können, dass dieses Wissen ein wechselseitiges ist. Die entsprechenden Einstellungen werden von Lewis also spieltheoretisch analysiert: Konventionen stabilisieren sich aufgrund der aus den individuellen Perspektiven durchgeführten Vorteilskalkulationen zugunsten der Befolgung allgemein verbindlicher, sprachlicher Regelungen.

Es kommt hier nicht auf die technischen Details und eine exakte Wiedergabe der IBS-intern diskutierten Schwierigkeiten und Verbesserungsvorschläge einer intentionengestützten Analyse von ‚Bedeutung‘ an – weder hinsichtlich der Lewisschen Überlegungen, noch hinsichtlich einer Liste notwendiger oder hinreichender Bedingungen von *Sprecherbedeutung* im Griceschen Sinne¹⁰⁰. Mir wird es im Folgenden darum gehen, die einflussreichsten Vorschläge für eine *naturalistische Reduktion* propositionaler Zuschreibungen (Überzeugungen, Intentionen usw.) zu kritisieren. Dabei sollen vor allem auch die ‚realistischen‘ Unterstellungen bezüglich solcher ‚Zustände‘ oder ‚Ereignisse‘ aufs Korn genommen werden – ein für das IBS-Programm zentraler Punkt. An dieser Stelle genügt es vorläufig, festzuhalten, dass IBS‘ler sprachliche Wort- oder Ausdrucksbedeutung grundsätzlich auf Sprecherbedeutung reduzieren wollen und erstere als (durch ihre *qua* konventionell geregelter Praxis standardisierten Eigenschaften) für sprachlich sozialisierte Kommunikationsteilnehmer erkennbarer, sicherer (aber dennoch nicht-natürlicher) *Hinweis* auf letztere verstanden wird. Die entsprechenden Fähigkeiten der Kommunikationsteilnehmer – ob in der Sprecher- oder in der Hörerrolle – sind nun dasjenige, was unter naturalistischen Voraussetzungen erklärt werden müsste. Um aber zu verstehen, was sich der Reduktive Materialismus hier tatsächlich vornimmt, kommen wir um einen Abstecher wissenschaftstheoretische Grundsatzfragen nicht herum.

¹⁰⁰ Vergl. hierzu u. a. Schiffer (1972) und (1987), sowie Loar (1981). Aus einer naturalismus-externen Perspektive hat bereits Habermas (in: ders.: ‚Intentionalistische Semantik‘ (1975/1976)) auf wesentliche, problematischen Grundannahmen des Grice-Programms hingewiesen, nämlich dass (1.) die Voraussetzung, dass das, was gemeint wird (die Sprecherbedeutung), „in keiner Weise durch das, ‚was gesagt wird‘, also durch den Wert von ‚x‘ bestimmt ist“ (S. 335), nur durch die Einführung eines problematischen Begriffs von ‚mutual knowledge‘ gerettet werden kann und dass (2.) IBS‘ler von der Möglichkeit ‚vollwertiger‘ propositional ausdifferenzierter Intentionen, als von der Existenz einer Praxis sprachlicher Verständigung unabhängig, ausgehen müssen.

2. Erklärung, Kausalität, Reduktion

Aus naturalistischer Perspektive ist das Verhalten oder ‚Handeln‘ von menschlichen Organismen nur eine (wenn auch für uns besonders interessante) Art von *sich zutragenden Ereignissen*, so wie auch Wetterumschwünge, Fortpflanzungsakte, räumliche Veränderungen von Körpern, Supernovae oder Niedergänge von Kulturen sich ereignen. Mitunter wollen wir Ereignisse zu *erklären*, vor allem um dadurch in die Lage versetzt zu werden, Ereignisse des gleichen Typs (mehr oder minder genau) *vorhersagen* oder *technisch beherrschen* zu können (oder uns zumindest besser auf sie einzustellen). Dies tun wir, in dem wir uns bemühen, Regelmäßigkeiten im Naturablauf herauszubekommen und – darüber hinaus – Vermutungen über die zugrundeliegende Mechanismen der durch uns wahrgenommen Veränderungen anzustellen. Schon in unserer alltäglichen Praxis sind wir auf eine Unmenge solcherart verallgemeinernden Wissens angewiesen.

Wie ich bereits in der Einleitung ausführte, beinhaltet NANÜCS eine, auch von mir nicht bestrittene, allgemeine Annahme, dass alle Zustände und Ereignisse der natürlichen Welt, Verhältnissen von Ursache-Wirkung zu anderen Zuständen und Ereignissen ausgesetzt sind. Wir können Regelmäßigkeiten bezüglich bestimmter solcher Verhältnisse anhand von allgemeinen Korrelationen von Typen von Ereignissen oder Zuständen in Form von Gesetzen erfassen. DOGNA ist nun durch die zusätzliche Annahme bestimmt, alle Formen unserer konstativen Ausdrucksmöglichkeiten seien, in einem grundsätzlich analogen Sinne, als empirische Beschreibungen von Zuständen oder Ereignissen zu verstehen. IBS will, wie wir gesehen haben, mentale, bzw. ‚intentionale‘ Redeweisen als in – am Ende – empirische Aussagen über physische Phänomene ‚übersetzbar‘, verständlich machen, da all jenes, dem eine spezifische, kausale Rolle in der Welt zufallen kann (was ja für ‚intentionale‘ Zustände unterstellt wird) rein physischer Natur sei und deshalb jede Art empirischen Kausalwissens in einen systematischen explanativen Zusammenhang mit unserem Wissen über die physische Natur gebracht werden können muss. (Würde sich ‚Intentionalität‘ dauerhaft einer materialistischen Reduktion, bzw. Integration widersetzen, müssten Naturalisten folglich umgekehrt davon ausgehen, dass entsprechende Eigenschaften nicht tatsächlich als kausalwirksame instantiiert sind und deshalb, in einem wissenschaftlich orientierten Weltbild, keine Rolle mehr spielen dürften.)

DOGNA unterstellt damit, dass *alle* Arten von kognitiven Fähigkeiten, über die wir als kompetente Teilnehmer einer Alltagspraxis verfügen, als Formen empirischen Verallgemeinerungswissens über (letztlich physische) Ereignisse oder Zustände zu deuten seien – auch unsere Sprach- bzw. Handlungskompetenz. Wenn wir erfolgreich

agieren, interagieren oder kommunizieren, haben wir, nach diesem Bild, jeweils nicht nur unsere Umwelt, sondern auch das Verhalten anderer im Hinblick auf die dem Verhalten zugrundeliegenden Ursache-Wirk-Verhältnisse richtig erklärt oder die Effekte unseres eigenen Verhaltens richtig prognostiziert; wir verfügen jeweils über eine entsprechende, wahre Theorie. Wie aber lässt sich diese Fähigkeit zu wahren, theoretischen Erklärungen jener besonderen Formen von Zuständen und Ereignissen wiederum selbst unter materialistischen Voraussetzungen, erklären? – Was wird überhaupt, im Rahmen des IBS-Programms, genauer unter einer ‚kausalen Erklärung‘ von (sprachlichem und nicht-sprachlichem) Verhalten, einer ‚naturalistischen Erklärung‘ oder ‚Reduktion‘ des Geistes verstanden, bzw. was müssten wir von solchen Erklärungen erwarten?

Die Frage, was, ganz allgemein, unter Erklärungen (von Tatsachen, Ereignissen, Zuständen oder Eigenschaften) zu verstehen ist, bietet natürlich Stoff für eine mehrbändige Abhandlung. Ich werde mich zunächst – entlang einer Vergegenwärtigung einschlägiger wissenschafts- und kausalitätstheoretischer Positionen – mit einigen grundsätzlichen Überlegungen begnügen, die dann im Hinblick auf Fragen nach Möglichkeiten eines naturalistischen Verständnisses des Phänomens der Bedeutung, im Sinne propositionalen Gehaltes, spezifiziert werden. Vor diesem Hintergrund lassen sich die unterschiedlichen naturalistischen Positionen insbesondere auch durch ihre jeweiligen Antworten auf die Frage bestimmen, *wie und ob* ‚Alltagspsychologie‘ – in deren Rahmen propositionale Gehalte, nach naturalistischer Auffassung, ja als Grundlage einer Erklärung von Verhalten postuliert werden – in eine wissenschaftliche Theorie transformiert werden kann.

Was und wie erklären Erklärungen?

Für die neuere, aus dem Programm des *Logischen Empirismus* entspringende Debatte um den Erklärungs begriff, war und ist sicherlich der Ansatz K. G. Hempels einer der einflussreichsten¹⁰¹. Hempel ging es allerdings nur um die ‚Logik‘ (bzw. die Adäquatheitsbedingungen) von Erklärungen eines bestimmten Typs, nämlich um Erklärungen *singulärer Ereignisse*. Nach Hempel entsprechen alle vollständigen

¹⁰¹ Hempels eigene Darstellung am ausführlichsten in: ders. (1965/1977). Zusammenfassende und kritische Darstellungen der Debatten um den Begriff der Erklärung finden sich u. a. bei W. Stegmüller: *Wissenschaftlicher Erklärung und Begründung* (1972); Beckermann (1977) und ders.: ‚Handeln und Handlungserklärungen‘ (1977a); von Wright (1971/1984); und Keil (2000).

Erklärungen singulärer Ereignisse einem deduktiven oder induktiven Schema (Hempel redet nur von Ereigniserklärungen). In diesem Schema wird das zu erklärende Ereignis (das *Explanandum*) durch eine Konklusion beschrieben, die aus den singulären Sätzen, welche die Anfangsbedingungen beschreiben, und den generellen Sätzen, in denen die Anfangsbedingungen mit dem zu erklärenden Ereignis verknüpft werden, entweder *deduktiv-nomologisch* oder *induktiv-statistisch* folgt. Die singulären Konstatierungen und die verallgemeinernde Verknüpfung bilden zusammen das *Explanans*.

Deduktiv-nomologische Erklärungen lassen sich, nach Hempel, durch folgende Adäquatheitsbedingungen charakterisieren:

(Adn₁) Das *Explanandum* muss logisch deduktiv aus dem *Explanans* folgen.

(Adn₂) Das *Explanans* muss mindestens ein generelles Gesetz wesentlich enthalten, d.h. dieses Gesetz muss zur Folgerung des *Explanandums* aus dem *Explanans* unentbehrlich sein.

(Adn₃) Das *Explanans* muss empirisch gehaltvoll sein.

(Adn₄) Die Aussagen, die das *Explanans* bilden, müssen wahr sein.

Die entsprechenden Bedingungen für induktiv-statistische Erklärungen lauten:

(Ais₁) Die Aussagen des *Explanans* müssen dem *Explanandum* eine hohe induktive Wahrscheinlichkeit verleihen.

(Ais₂) Das *Explanans* muss mindestens ein statistisches Gesetz wesentlich enthalten.

(Ais₃) Das *Explanans* muss empirisch gehaltvoll sein.

(Ais₄) Die Aussagen, die das *Explanans* bilden, müssen wahr sein.

(Ais₅) Im *Explanans* müssen alle für das *Explanandum* statistisch relevanten Fakten berücksichtigt werden.

Ohne auf all diese Bedingungen im Einzelnen eingehen zu müssen, können wir uns fragen, ob dies tatsächlich eine befriedigende Erläuterung unseres Verständnisses kausaler Ereignis-Erklärungen ist. Ein grundsätzlicher, von unterschiedlichen Kritikern vorgebrachter Einwand läuft darauf hinaus, dass Hempel eher ein Schema der *Rechtfertigung*, denn ein Schema der Erklärung vorgelegt hat. Wie Stegmüller bemerkt, sind die von verschiedenen Autoren gegen Hempels zusätzliche These von der strukturellen Gleichartigkeit von Erklärung und Prognose vorgebrachten Einwände zugleich Argumente gegen ein Verständnis des Hempelschen Schemas als eines der *kausalen* Erklärung¹⁰²: So kann beispielsweise das auf einem Barometer ablesbare

¹⁰² Stegmüller (1972), S. 760 f.

Fallen des Luftdruckes das Ende einer Schönwetterperiode nicht erklären, weil das Fallen des Luftdrucks zwar ein guter *Indikator* für Heraufziehendes Schlechtwetter sein mag, *nicht* jedoch dessen *Ursache* (in ähnlicher Weise können Fieber und rote Flecken auf der Haut nicht als erklärende Ursache von Masern angesehen werden). Gemäß dem Hempel-Schema für Erklärungen würde jedoch das in singulären Konstatierungen beschreibbare Vorliegen von Tatsachen, die wir tatsächlich lediglich als Indikatoren für andere (möglicherweise zukünftige) Tatsachen ansehen, letztere ‚erklären‘. Hempel selbst hatte vorgeschlagen, zwischen *epistemischen* und *Erklärung-heischenden Warum-Fragen* zu unterscheiden. Nach Stegmüller sollten wir allerdings auch Hempels Modell der Erklärung eher als Versuch der Explikation wissenschaftlichen Begründens verstehen. Wissenschaftliche Erklärungen, die darauf antworten, warum sich ein Ereignis zugetragen hat (oder warum ein Zustand besteht), wollen uns, so Stegmüller, nicht sagen, warum es *vernünftig ist, zu glauben*, dass *A* weil *B*, sondern dass *tatsächlich A* weil *B*. Kausale Erklärungen sollten also nicht mit ‚Rationalisierungen‘ oder begründeten Prognosen gleichgesetzt werden, da zu ersteren offenbar mehr gehört, als das Schema Hempels hergibt.

John Mackie hat darauf hingewiesen, dass die genannten Einwände ebenso auf die dem Hempelschen Modell zugrundeliegende *Humesche Regularitätsanalyse von Kausalität* Anwendung finden, da eine solche Analyse nicht zwischen echten Verursachungsbeziehungen einerseits und zufälligen Regelmäßigkeiten, Epiphänomen (Symptomen, Indikatoren) oder verhinderten Ursachen andererseits unterscheidet¹⁰³. Eine weitere Schwierigkeit einer regularistischen Analyse der Kausalität besteht darin, dass diese nicht die Richtung der Abhängigkeitsbeziehung identifizieren kann, die für kausale Beziehungen typisch ist. Es drängt sich daher eine *kontrafaktische* Analyse des Kausalitätsbegriffes auf, also eine Analyse, die sich an der Frage orientiert, was gewesen wäre, wenn die vermeintliche Ursache nicht vorgelegen hätte. Wenn wir, im Sinne einer kausalen Erklärung singulärer Ereignisse, sagen, dass *A* die Ursache von *B* war, dann meinen wir damit üblicherweise auch, dass es einen Unterschied im Lauf der Dinge macht, ob *A* stattfand oder nicht. Wir meinen damit ungefähr, dass die Wirkung *B*, unter den gegebenen Umständen, wohl nicht eingetreten wäre, wenn *A* sich nicht ereignet hätte, bzw. der Fall gewesen wäre.

Dieser Intuition will Mackie Rechnung tragen, indem er Kausalität als eine bestimmte Form von Bedingungsrelation versteht¹⁰⁴. Sein anschauliches Beispiel ist ein gelöschter Wohnungsbrand, der natürlich die Frage nach dessen Ursache aufkommen lässt. Was bedeutet es, so fragt sich Mackie, wenn Experten zu dem Schluss kommen,

¹⁰³ J. Mackie: ‚Causes and Conditions‘ (1965).

¹⁰⁴ So auch von Wright (1971/1984).

dass ein Kurzschluss die Ursache des Feuers war? Da das Feuer sicher auch auf andere Weise hätte entstehen können, war der Kurzschluss keine notwendige Bedingung des Brandes. Er war andererseits aber auch keine hinreichende Bedingung, denn wenn zum Beispiel kein brennbares Material in der Nähe gewesen wäre, hätte der Kurzschluss nicht zu diesem Brand führen können. Wie lässt sich die genannte Ursachenbehauptung dann aber überhaupt noch als Bedingungsrelation verstehen? Mackie meint dazu:

„At least part of the answer is that there is a set of conditions (of which some are positive and some are negative), including the presence of inflammable material, the absence of a suitably placed sprinkler, and no doubt quite a number of others, which combined with the short-circuit constituted a complex condition that was sufficient for the house catching fire - sufficient, but not necessary, for the fire could have started in other ways. Also, of this complex condition, the short circuit was an indispensable part: the other parts of this condition, conjoined with one another in the absence of the short circuit, would not have produced the fire. The short circuit which is said to have caused the fire is thus an indispensable part of a complex sufficient (but not necessary) condition of the fire. In this case, then, the so called cause is, and is known to be, an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result. The experts are saying in effect, that the short circuit is a condition of this sort, that it occurred, that the other conditions which conjoined with it to form a sufficient condition were also present, and that no other sufficient condition of the house's catching fire was present on this occasion.“ (Mackie (1965), S. 16)

Sowohl die Form, als auch Möglichkeit einer Analyse der Kausalität im Sinne eines Bedingungsverhältnisses zwischen Ursache und Wirkung sind allerdings in der Diskussion um den Kausalitätsbegriff, bzw. den Begriff der kausalen Erklärung, strittig geblieben¹⁰⁵. Von Wright etwa unterscheidet kausale Erklärungen in Fragen des Types *Warum notwendig?* und Fragen des Types *Wie möglich?*¹⁰⁶. Während in ersteren hinreichende Bedingungen eine wesentliche Rolle spielten und entsprechende Erklärungen zu Voraussagen verwendet werden könnten, sei dies beim zweiten Typ von Fragen, bei dem notwendige Bedingungen die wesentliche Rolle spielten, nicht möglich. Darüber hinaus ist von Wright auch der Ansicht, dass der Kausalitätsbegriff nicht ohne die *Voraussetzung der Möglichkeit manipulativen Eingreifens in Vorgänge in der Welt* erläutert werden kann und kommt zu dem Schluss: „(...) dass *p* die Ursache von *q* ist, *bedeutet* (...), dass ich *q* herbeiführen könnte, wenn ich *p* tun könnte“¹⁰⁷. Auch eine solche Analyse ist allerdings nicht ohne Schwierigkeiten. So lässt sich u. a. bezweifeln, ob ein Begriff des Herbeiführens von bzw. des Eingreifens in

¹⁰⁵ Vergl. hierzu etwa den Sammelband *Causation*, hrsg. v. Sosa, E./ Tooley, M. (1993);

¹⁰⁶ Von Wright (1971/1984), Kap. II.

¹⁰⁷ Ebd., S. 75. Diesen Standpunkt teilt auch Searle (in: Searle (1983/1987), Kap. 4.

Naturvorgänge(n) seinerseits erläutert werden kann, ohne auf den der Kausalität zurückzugreifen. Die Frage, ob Kausalität am Ende nur über die Voraussetzung der Möglichkeit von Handlungen analysiert werden kann, ist an dieser Stelle aber nicht unser Thema. Der Kausalitätsbegriff soll uns hier lediglich im Zusammenhang mit der Frage interessieren, was wir unter kausalen Erklärungen verstehen können.

Donald Davidson hat, wohl nicht zu Unrecht, darauf hingewiesen, dass es nicht ganz einleuchtet, wenn Mackie sagt, *das* Feuer hätte auch auf andere Weise entstehen können¹⁰⁸. Hätte *dieses* Feuer wirklich anders verursacht werden können? Nach Davidson ist es ein Irrtum, Kausalität als eine Relation zwischen (physischen) Ereignis- oder Zustandstypen, statt zwischen entsprechenden *Token*, zu verstehen. Kausalität gäbe es allerdings, auch zwischen Ereignistoken, nur dann, wenn es eine Beschreibung *gibt* (wenn sie also Eigenschaften, bzw. Typen *instantiieren*), nach der diese Token unter ein *strikt*es (ausnahmeloses) Gesetz fallen. Während *Kausalerklärungen*, nach Davidson, *intensional* sind, stellt Kausalität eine *extensionale* Relation dar. Davidson vertritt damit die durchaus gängige Position, dass der Begriff der kausalen Erklärung auf den Begriff des Naturgesetzes angewiesen ist, der seinerseits von nicht-gesetzesartigen Korrelationen von Ereignissen unterschieden werden muss, um den Einwänden zu entgehen, die gegen das sogenannte ‚H-O-Schema‘ kausaler Erklärungen vorgebracht wurden.

Davidsons Analyse des Kausalitätsbegriffs spielt auch für seine Position innerhalb der ‚Philosophie des Geistes‘ eine wichtige Rolle, auf die wir noch zurückkommen werden. Hier geht es jedoch zunächst ganz grundsätzlich um Versuche einer adäquaten Explikation des Begriffs der kausalen Erklärung, und nur im Hinblick auf diesen Kontext gilt es daher, Davidsons Vorschlag etwas genauer zu erläutern. Unter Naturgesetzen versteht man üblicherweise universell quantifizierende, kontrafaktische, wahre Konditionalsätze, die keine räumlichen oder zeitlichen Einschränkungen enthalten. Durch die Kontrafaktizitätsbedingung und die Uneingeschränktheitsbedingung werden zufällige faktische Regelmäßigkeiten ausgeschlossen. Unter nicht-strikten Gesetzen versteht man solche, die behaupten, dass eine bestimmte Ursache (nur dann) eine bestimmte Wirkung haben wird, wenn keine störenden Einflüsse vorliegen. Der kausale Zusammenhang zwischen Ereignissen oder Zuständen könnte ja durch andere kausale Effekte sozusagen ‚überlagert‘ werden (z.B. verursacht die Tatsache, dass *x* leichter als Luft ist, dass *x* fliegt – wenn es nicht von

¹⁰⁸ D. Davidson: ‚Kausale Beziehungen‘ (1967). Für eine neuere Fassung von Davidsons Überlegungen zur Kausalität, vergl. ders.: ‚Thinking Causes‘ (1993). Auf die Schwierigkeiten dieser Position, vor allem in Hinblick auf das Problem ‚mentaler Verursachung‘ und auf die damit zusammenhängende Diskussion über den Status von *ceteris-paribus*-Gesetzen, werde ich unten zurückkommen.

jemandem festgehalten wird, durch starken Wind zu Boden gedrückt wird usw.). Man nennt dies den *ceteris-paribus-* (*cp-*) Charakter kausaler Gesetze. Solche Gesetze erlauben, wenn überhaupt, nur statistische Prognosen über das Eintreten zukünftiger Ereignisse, da man praktischerweise nicht alle möglichen störenden Einflüsse antizipierend berücksichtigen kann. Die Meteorologie ist hierfür ein prominentes Beispiel.

Wir gesehen hatten, lassen sich aus statistisch quantifizierende *cp*-Gesetzen keine kausalen, *deduktiv*-nomologischen *Erklärungen* im Sinne Hempels herleiten. Häufig wird davon ausgegangen, dass nur physikalische Gesetze keinen *cp*-Charakter haben, also strikte sind¹⁰⁹. Nun ist der besondere Status physikalischer Gesetze sicherlich schon dadurch gegeben, dass *alle* in ihnen genannten Eigenschaften durch quantitative Begriffe (wie etwa „Volumen“, „Geschwindigkeit“ oder „elektrische Ladung“) charakterisiert sind. Grundsätzlich gilt allerdings für alle empirischen Wissenschaften, dass ihre möglichen ‚Gegenstände‘ empirisch operationalisierbaren Kriterien der Individuierbarkeit (Identifizierbarkeit) von Ereignissen, Zuständen und Dingen genügen und die gegenstandsspezifischen Eigenschaften einen gewissen Grad an Quantifizierbarkeit (Messbarkeit von Größen) ermöglichen müssen. Unabhängig davon, müssen wir uns aber fragen, ob selbst physikalische Gesetze tatsächlich strikt in dem Sinne sind, dass sie ausnahmslose Korrelationen zwischen Ereignistypen zu beschreiben vermögen. Die scheint nämlich, wenn überhaupt, nur für einige basale Gesetze der Elementarphysik (etwa „ $E=mc^2$ “) zu gelten, die jedoch keine Kausalgesetze sind. Ansonsten kommt es auch im Bereich von physikalisch beschriebenen, regelmäßigen Ereigniskorrelationen – jedenfalls in der ‚freien Wildbahn‘, das heißt, außerhalb von speziellen Experimentalsituationen – normalerweise zu kausalen Überlagerungen. Es spricht offenbar vieles dafür, den Begriff des Naturgesetzes – und damit auch den Kausalbegriff – nicht über den Begriff der Regelmäßigkeit zu analysieren, selbst wenn entsprechende Allsätze durch zusätzliche Bedingungen, wie etwa Kontrafaktizität oder Uneingeschränktheit spezifiziert werden. Es bietet sich offenbar eher eine Charakterisierung im Sinne von *Potentialen* oder *Dispositionen* von Dingen mit bestimmten Eigenschaften an¹¹⁰.

Nun kann nach Davidson von kausalen Erklärungen durch nicht-strikte Gesetze, bzw. von Kausalität generell, nur deshalb die Rede sein, weil die in entsprechenden Erklärungen oder singulären Aussagen korrelierten Ereignisse unter irgendeiner anderen (physikalischen) Beschreibung auch unter strikte Gesetze fallen. Dies ist aber nach dem bisher Gesagten zweifelhaft, selbst wenn man, wie die meisten Theoretiker,

¹⁰⁹ So auch bei Davidson.

¹¹⁰ Vergl. A. F. Chalmers: *Wege der Wissenschaft* (2001), Kap 14.

grundsätzlich der Ansicht ist, dass für das Vorliegen von Kausalrelationen gesetzesartige Zusammenhänge nötig sind. Kausalgesetze beschreiben nämlich, nach dem hier nahegelegten potentialistischen oder dispositionalen Verständnis, gar keine Ereignissequenzen, bzw. sie korrelieren keine Ereignistypen. Es sind die *Aktualisierungen* (gegebenenfalls unterschiedlicher und sich möglicherweise überlagernder) kausaler Potentiale, die als Relationen zwischen Zustands- oder Ereignistoken aufgefasst werden müssen¹¹¹. Kausale *Erklärungen*, im bisher diskutierten Sinne, nehmen ja auf Wirkungen und ihre Ursachen als Ereignistoken Bezug: Bei einer entsprechenden Kausalerklärung geht es beispielsweise nicht um Wohnungsbrände und ihre Ursachen im Allgemeinen, sondern um die Tatsache eines bestimmten Brandes.

Es gibt in der Diskussion um den Begriff der Kausalität und um die kausale Erklärung von singulären Ereignissen noch eine andere Verwirrung. So wird – auch von Naturalisten – häufig behauptet, Kausalerklärungen (oder gar Kausalität) seien relativ auf unsere jeweiligen Erkenntnisinteressen: Es sei beispielsweise möglich, als Ursache dafür, dass das Licht im Zimmer anging, sowohl das Einschalten des Lichtes, als auch das Herunterdrücken des Lichtschalters oder die Schließung des Stromkreises anzugeben. Dies klingt zunächst einleuchtend. Doch wir sollten uns einmal fragen, inwiefern das Angehen des Lichtes überhaupt ein angemessener Kandidat für eine Kausalerklärung im üblichen Sinne eines erklärungsbedürftigen Ereignisses ist. Nach einer solchen Erklärung würden wir doch eigentlich nur dann fragen, wenn wir uns darüber wundern, dass das Licht angegangen ist; wenn also das betreffende Ereignis nicht dem erwarteten Ablauf der Dinge entspricht. Und in solch einem Fall wären alle der oben als mögliche angeführten ‚Erklärungen‘ keine vollständigen Erklärungen. Wir erklären eine Tatsache, wie die, dass das Licht anging, möglicherweise dadurch, dass jemand es einschaltete, von dem wir nicht bemerkten, dass er oder sie früher nach Hause kam, als wir erwarteten oder dadurch, dass der Schalter einen Wackelkontakt hat, weshalb der Stromkreis ungewollt geschlossen wurde.

Doch selbst wenn wir die Antwort „Jemand hat das Licht eingeschaltet“, auch in gewöhnlichen Fällen, bereits als eine wirkliche Erklärung eines Ereignisses des Lichtangehens akzeptierten, wäre dies, wie mir scheint, nicht eine unter möglichen anderen interessenrelativen Erklärungen. Denn das jemand den Schalter herunterdrücken musste, um das Licht einzuschalten, ist, unter den gegebenen Umständen, eher eine *notwendige Bedingung* dafür, dass er die für das Angehen des Lichtes ursächliche Handlung des Lichteinschaltens ausführen kann, aber keine

¹¹¹ Der Begriff der Kausalität, bzw. der Verursachung scheint überdies mit der Vorstellung einer Energieübertragung von der Ursache auf die Wirkung verbunden zu sein. Vergl. hierzu Quine: *Die Wurzeln der Referenz* (1974/1976), § 2).

Erklärung. Dies gilt erst recht für die physikalische Variante der ‚Erklärung‘ (der sich schließende Stromkreis): Alle Bedingungen, die für das Angehen des Lichtes eine Rolle spielen, ob sie das Nervensystem des Einschaltenden oder den Stromkreis und die Lichtquelle betreffen, erklären eigentlich nicht (kausal), *warum* das Ereignis des Lichtangehens im konkreten Fall stattgefunden hat. Sie können höchstens erklären, *wie* es *möglich* war, dass dieses Ereignis stattgefunden hat, indem sie, wie im konkreten Fall, etwa beschreiben, wie der Vorgang des Lichtanschaltens physisch instantiiert war. – Dies ist allerdings eine andere Art von Erklärung, als die eines singulären Ereignisses¹¹². Denn auch im Falle der Erklärung *des Ereignisses* des Lichtangehens durch einen Wackelkontakt wäre es irrelevant, *wie* dieser, im konkreten Fall, physikalisch realisiert ist. Das *Erklärende* wäre hier nämlich die konkrete Dysfunktionalität einer bestimmten Art von funktional verstandenem Mechanismus, wie immer er physisch beschaffen ist, und auch dies ist in keiner Weise interessenrelativ. Die Beschreibung der *physikalischen Realisierung* der Dysfunktion konkurrenziert hier sozusagen nicht auf der *gleichen Ebene möglicher Erklärung*.

Schon Mackies Beispiel weist darauf hin, dass *ursächliche* Erklärungen von *einzelnen Ereignissen* bestimmte Aspekte eines Gesamtzusammenhangs von Ereignissen oder Vorgängen hervorheben, die in einer bestimmten Situation *ungewöhnlich* waren. Wir verlangen grundsätzlich zunächst dann nach solchen Erklärungen, wenn die Dinge nicht so verlaufen, wie wir es üblicherweise erwarten würden; häufig, wenn etwas passiert, das wir – wenn möglich – zukünftig entweder vermeiden oder wiederholen möchten. Und dafür wäre eine Auflistung aller Bedingungen oder gar Randbedingungen, die wir zur Ermöglichung der zu erklärenden Tatsache als notwendig erachten (wenn eine solche vollständige Auflistung überhaupt möglich wäre) schlicht uninteressant. Die Interessenrelativität besteht, wenn überhaupt, also nicht in der Möglichkeit unterschiedlicher Erklärungen, sondern in der Bestimmung dessen, was *erklärungswürdig* erscheint.

Allerdings gehen *wissenschaftlich professionalisierte* Formen von Erklärungs Bemühungen (bzw. Fragens) über kausale Ereigniserklärungen hinaus. Systematisches, wissenschaftliches Forschen beginnt doch auch gerade dort, wo nicht mehr nur das Ungewöhnliche erklärt werden soll, sondern das Gewöhnliche (also zum Beispiel, warum der Apfel zu Boden fällt, warum es warme und kalte Jahreszeiten gibt, warum bestimmte Vögel schwarz sind oder warum die phänomenalen Eigenschaften des Feuers so sind, wie sie sind) – Vorgänge, Zustände oder Eigenschaften, bei denen wir normalerweise keinen Anlass sehen, uns zu wundern (zumindest dann nicht mehr,

¹¹² Vergl. die oben angeführte Unterscheidung von Wrights zwischen Kausalerklärungen des Typs ‚warum notwendig?‘ von solchen des Typs ‚wie möglich?‘.

wenn wir einer bestimmten Kindheitsphase notorischen Warum-Fragens entwachsen sind). Hier wird jedoch nicht nach der Ursache bestimmter Einzelereignisse gefragt, sondern danach, warum Ereignisse eines bestimmten Typs überhaupt stattfinden, also etwa, warum Äpfel ganz allgemein (früher oder später) zu Boden fallen.

Offenbar spielen also auch andere Arten von Erklärungen, als die der Ursache von Einzelereignissen, eine wichtige Rolle. Wenn wir etwa eine Erklärung für die jahreszeitlichen Temperaturschwankungen suchen (Unterschiedliche Erwärmung der Luft, hervorgerufen durch unterschiedliche Sonneneinstrahlung, hervorgerufen durch die Neigung der Achse der die Sonne umrundenden Erde usw.), verlangen wir keine kausalen Erklärungen von Einzelereignissen, sondern von bestimmten Regelmäßigkeiten in der Natur. Eine Erklärung einer solchen allgemeinen Tatsache bringt diese mit bestimmten natürlichen Gegebenheiten und Gesetzmäßigkeiten in Zusammenhang. Dabei hat die *Richtung* der Erklärung meist eine bestimmte *Form*, im Sinne eines Fortschreitens vom Phänomenalen zum Physischen.

Ein wichtiger Aspekt wissenschaftlicher Forschung ist die ‚Jagd‘ nach Erklärungen durch Theorien, die R. Cummins „Eigenschaftstheorien“ („*Property Theories*“) nennt¹¹³. Solche Erklärungen geben Antworten auf Fragen des Typs „Wodurch hat ein X (Ding, System) die Eigenschaft P“ oder „Wodurch instantiiert X die Eigenschaft P“. Sie unterscheiden sich damit eindeutig von, die wissenschaftstheoretische Debatte dominierenden, Erklärungen von Ereignissen, bzw. Zustandsveränderungen, welche Cummins „Transitionstheorien“ („*Transition Theories*“) nennt:

„Understood as an answer to the latter questions, the kinetic theory of heat (and the molecular theory of gases, that it presupposes) is not a transition theory but a property theory: it explains temperatur in gas by explaining how temperatur is instantiated; it does not by itself, explain changes in temperatur.“ (a. a. O., S. 15)

Im Hinblick auf Erklärungsweisen ‚oberhalb‘ physikalisch oder chemisch beschriebener Zustände oder Ereignisse, müssen darüber hinaus weitere Differenzierungen vorgenommen werden. Brian Ellis¹¹⁴ beispielsweise unterscheidet immerhin vier Formen wissenschaftlicher Erklärungen, denen unterschiedliche Theorieformen korrespondieren:

„A *causal explanation* is information about the causal history of something or about the causal processes wich result in something. A *functional explanation* is information about something in an ongoing system – about the contribution it makes in sustaining it. A *model theoretic* explanation is information about how

¹¹³ R. Cummins: *The Nature of Psychological Explanation* (1983).

¹¹⁴ B. Ellis: ‚What Science Aimes to Do‘ (1985).

(if at all) the actual behaviour of some system differs from that which it should have ideally if it were not for some perturbing influences and, where necessary, includes some information about what perturbing influences may be causing the differences. A *systematic explanation* is information about how the fact to be explained is systematically related to other facts.“ (Ellis 1985, S. 55)

Unabhängig von der Frage, ob diese Unterscheidungen adäquat und vollständig sind, kann jedenfalls dort, wo Begriffe wie „System“ und „Verhalten“ ins Spiel kommen, vor allem auch die Frage interessieren, *wie die innere Struktur eines Systems beschaffen sein muss, um erklären zu können, warum oder wodurch dieses System bestimmte Kausaleigenschaften hat*. In diesem Sinne unterscheidet etwa Dretske zwischen einer *auslösenden* und einer *strukturierenden Ursache* eines Verhaltens¹¹⁵. Dretske bestimmt Verhalten als einen *Prozess des Verursachens* eines Zustandes oder Ereignisses durch eine dem sich Verhaltenden *interne* Ursache, als etwas, das einem Ding oder System nicht einfach, durch eine äußere Ursache, zustößt. Wenn wir diese eher ‚liberale‘ Bestimmung dessen, was Verhalten sein kann, mal als akzeptabel stehen lassen, könnten wir sagen, dass wissenschaftliche Theorien, die Verhalten erklären sollen, häufig nicht auf eine Erklärung der *auslösenden* Ursache von Einzelereignissen abzielen, wie etwa das Sich-Schließen einer bestimmten Blüte (etwa dadurch, dass es an Licht einer bestimmten Wellenlänge mangelte), sondern danach fragen, *warum* sich Blüten *überhaupt* unter entsprechenden Umständen schließen – was also die *strukturierende* Ursache dieses Verhaltens ist. Sie geben dann etwa eine *funktionalistische* Erklärung, die ihrerseits auf bestimmten evolutionstheoretischen Annahmen beruhen kann. Mit solchen Formen der Erklärung, die besonders auch für einige naturalistische Philosophien der Bedeutung und des Geistes^{DE} eine wichtige Rolle spielen, werden wir uns noch eingehender beschäftigen.

Den unterschiedlichen Formen von Erklärungen entsprechen jeweils unterschiedliche Arten von Gesetzen: Neben Kausalgesetzen (bzw. nomischen Korrelationen, die durch Ursache-Wirkung-Paare instantiiert werden) haben wir es auch mit *nicht-kausalen* nomischen Korrelationen zu tun. So können auch Korrelationen, die als Generalisierungen induktiver Evidenz über natürliche Zusammenhänge zu verstehen sind, *nomische Korrelationen* sein, die Erklärungen hergeben¹¹⁶. Darüber hinaus lassen sich noch andere Arten von Gesetzen anführen.

¹¹⁵ Dretske (1988).

¹¹⁶ Beispiele für nicht-kausale Korrelationen, die, anders als die von Barometeranzeige und Wetterentwicklung oder roter Flecken und Masern, nicht nur in Schemata eine Rolle spielen, die den Charakter einer Rechtfertigung von Erwartungen haben, sondern in genuinen Erklärungen, sind typische Naturgesetze, wie jenes, das thermische und elektrische Leitfähigkeit korreliert oder jenes, welches die inverse Variation von Gasdruck und Gasvolumen konstatiert.

Cummins unterscheidet zusätzlich zwischen *nomischen Attributionen* (gesetzesartige Prädikationen der Form „Alle x haben die Eigenschaft E “), *Instantiierungsgesetzen* (z.B. „Temperatur ist in einem Gas als die durchschnittliche mittlere kinetische Energie der Moleküle des Gases instantiiert“) und *Kompositionsgesetzen* (gesetzesartige Feststellungen im Sinne einer spezifischen Analyse eines bestimmten Typs von System, wie etwa Aussagen über den atomaren Aufbau des Wassermoleküls)¹¹⁷.

Dies müssen wir hier nicht im Einzelnen vertiefen. Wichtig für das Verständnis der reduktionistischen Ambitionen naturalistischer Theorien ist es, zu sehen, dass naturwissenschaftliche Kausalerklärungen in den sogenannten ‚Spezialwissenschaften‘ (wie z.B. der Biologie, der Geologie oder der Meteorologie) insbesondere auch für eine Antwort auf die Frage offen sein müssen, warum – im Sinne von *wie* oder *wodurch* – Ereignisse mit der Eigenschaft F die Ursache dafür sein können, dass Ereignisse mit der Eigenschaft G auftreten – zum Beispiel, warum die Bestrahlung von Pflanzen mit Tageslicht in diesen photosynthetische Prozesse verursacht, warum sich Wasser in gefrorenem Zustand ausdehnt oder warum eine Substanz mit einer anderen exotherm reagiert, nicht aber mit einer dritten.

Fodor nennt solche Wodurch-Fragen „Fragen nach dem *Mechanismus der Implementierung*“ nicht grundlegender Gesetze¹¹⁸. Grundlegend sind, nach Fodor, jene Gesetze, für die es (nach dem gegenwärtigen Stand der Forschung) keine Implementierung mehr gibt, die also den Punkt bilden, an dem sich bei reduktiven Erklärungsversuchen (durch Eigenschaftstheorien) der Spaten zurück biegt und die einzig angemessene Reaktion auf eine weitere Wodurch-Frage in der Antwort bestünde, dass die Natur eben offenbar diese Eigenschaften hat. Nicht nur alle spezialwissenschaftlichen, sondern überhaupt alle nicht-physikalischen, nomologischen Erklärungen wären in diesem Sinne nicht grundlegend.

Die Mechanismen, durch die spezialwissenschaftliche Gesetze implementiert werden können, sind typischerweise durch Eigenschaften der *Mikrostruktur* jener ‚Dinge‘ (Ereignisse, Zustände) bestimmt, die jeweils zum Bereich der Variablen des zu implementierenden Gesetzes gehören. So kann die Frage, wodurch Strahlung einer bestimmten Wellenlänge (wie das Sonnenlicht) in Pflanzen die Photosynthese bewirkt, auf der Ebene der chemischen Struktur der Pflanzen beantwortet werden. Entsprechend könnten Wodurch-Fragen hinsichtlich chemischer Kausal- oder ‚Transitions Gesetze‘ auf der Ebene physikalischer Eigenschaften erklärt werden. Dies ist allerdings nur dann möglich, wenn eine (wahre) Theorie existiert, die erklärt, *wie* das Vorliegen der Eigenschaften auf der *zu* implementierenden Ebene jeweils für das Vorliegen der

¹¹⁷ Vergl. Cummins (1983), Kap. 1.

¹¹⁸ J. Fodor: *The Elm and the Expert* (1994a); s. a. ders. (1987).

Eigenschaften auf der implementierenden Ebene hinreichend sein kann. Das heißt, dass Erklärungen durch implementierende Mechanismen nur auf der Grundlage von Eigenschaftstheorien möglich sind, die die entsprechenden, auf Wodurch-Fragen in Cummins Sinne antwortenden, Instantiierungsgesetze liefern.

Nach Fodor können Eigenschaftstheorien zwei Arten von Relationen zwischen den zu implementierenden Gesetzen und den implementierenden Mechanismen behaupten: entweder *Typenreduzierbarkeit* oder *multiple Realisierbarkeit*. Im ersten Fall würde die Eigenschaft auf der höheren Ebene mit einer Eigenschaft auf der explanativ nächst tieferliegenden (der implementierenden) Ebene *identifiziert* (beispielsweise: Wasser zu sein mit H₂O zu sein). Im anderen Fall sollen dagegen *unterschiedliche Arten* von Mechanismen ein und denselben Ereignistyp implementieren können. Multiple Realisierbarkeit liegt dementsprechend dann vor, wenn es eine Disjunktion von Eigenschaften auf der implementierenden Ebene derart gibt, dass die Instantiierung der Eigenschaft auf der zu implementierenden Ebene hinreichend für die Instantiierung der Disjunktion, aber nicht für die Instantiierung irgendeines ihrer Disjunkte ist. Dies gilt normalerweise dann, wenn die Eigenschaft auf der zu implementierenden Ebene *funktional* definiert ist und die realisierende Disjunktion alle und nur die Mechanismen enthält, welche die Funktion ermöglichen (beispielsweise kann die funktional bestimmte Eigenschaft Blut zu pumpen durch die Kontraktion eines Herzmuskels *oder* die mechanischen Bewegungen eines Kunstherzens realisiert sein).

In Fällen multipler Realisierbarkeit würde „*Erklärung durch Implementierung*“ also heißen, dass, wenn zwei Eigenschaftsdisjunktionen ED^{iM1} und ED^{iM2} , auf der Ebene des implementierenden Mechanismus, jeweils hinreichende Bedingungen für zwei (funktional bestimmte) ‚höherstufige‘ Eigenschaften E^{h1} und E^{h2} bildeten und ED^{iM1} eine Ursache von ED^{iM2} darstellte, nach diesem Modell *erklärt* wäre, *wodurch* E^{h1} E^{h2} verursacht hat. Theorien über Mechanismen der Implementierung von spezialwissenschaftlichen Gesetzen bewegen sich demnach auf mindestens zwei Ebenen von Vokabularen (welche durch ihre jeweils ebenenspezifischen Prädikate, Erklärungsweisen und Problemlösungen bestimmt sind). Allerdings spielt die Möglichkeit jeweils weiterer Ebenen der Implementierbarkeit für eine Antwort auf die ursprüngliche Wodurch-Frage häufig keine große Rolle. So ist beispielsweise für die Erklärung gewisser Vorgänge in Pflanzen ein Kenntnis physikalischer Gesetzmäßigkeiten nicht mehr in gleicher Weise aufschlussreich, wie die Kenntnis bestimmter chemischer Gesetzmäßigkeiten. Die Implementierungsmechanismen, die eine wesentliche Rolle in wissenschaftlichen Standarderklärungen spielen, beschränken sich meist auf eine nächsttieferliegende Ebene wissenschaftlicher Vokabularen.

Die Anwendung des deduktiv-nomologischen Erklärungsschemas auf Handlungen

Wir können naturalistische Positionen, wie gesagt, auch anhand der Typen wissenschaftlicher Erklärung einteilen, die sie jeweils beanspruchen zu können glauben. Theorien, die es auf die ‚harte Tour‘ probieren, wollen nicht nur Handlungen als mögliche Gegenstände deduktiv-nomologischer Ereigniserklärungen zu verstehen, sondern entsprechende ‚intentionale Gesetze‘ auch als Gegenstand von Eigenschaftserklärungen – im Sinne einer Reduktion auf physisch beschriebene Mechanismen der Implementierung. ‚Softies‘, wie insbesondere Davidson, halten dagegen keines der genannten wissenschaftlichen Erklärungsmuster auf der Ebene ‚intentionaler‘ Beschreibungen für angebracht. Gemeinsame Grundvoraussetzung ist aber, dass die sogenannte ‚Alltagspsychologie‘ als eine *empirische* Theorie des Verhaltens zu verstehen ist und meist auch, dass diese zumindest passable *Kausalerklärungen* von Vorkommnissen komplexer Verhaltensformen (bzw. Handlungen) liefert.

Wenden wir uns nun zunächst dem Versuch zu, Handlungen als Gegenstand ‚subsumtionstheoretischer‘ Erklärungen nach dem H-O-Schema plausibel zu machen. Wenn davon ausgegangen wird, dass propositionale Einstellungen, bzw. ‚intentionale‘ Zustände und Ereignisse mit Handlungen oder Verhalten in kausalen Beziehungen stehen, muss ja geklärt werden, was dies genauer heißen soll. Und die gegenwärtig noch immer gängigste Form, den Begriff der Kausalität oder der kausalen Beziehung zu analysieren, läuft über den Begriff der Gesetzesartigkeit, bzw. über gesetzesartigen Erklärungen von Ereignissen.

Der elaborierteste Versuch zur Verteidigung der These, Handlungen ließen sich kausal nach dem oben genannten, deduktiv-nomologischen Schema erklären, ist nicht Hempels eigener, sondern stammt von Paul Churchland¹¹⁹. Churchland macht sich dabei jene überraschende, auf Davidsons berühmten Artikel ‚Handlungen, Gründe und Ursachen‘ von 1963 zurückgehende, Idee zunutze, die in Form praktischer Syllogismen artikulierbaren *Rationalisierungsbeziehungen* als besondere *Form der Beschreibung kausaler Beziehungen* zu deuten. Wie wir oben bereits gesehen hatten, wollen die Vertreter kausaler Handlungstheorien mit diesem Zug dem ‚sprachendualistischen‘ Einwand begegnen, die genannten Formen von Beziehungen wären insofern gegensätzlich, als die eine begrifflicher (und damit notwendiger), die anderen aber

¹¹⁹ P. Churchland (1970/1977). Allerdings hat sich Churchland mittlerweile zum Hauptvertreter des Eliminativen Materialismus gewandelt.

empirischer (und damit kontingenter) Natur sei. Da die Beschreibung kausaler Beziehungen aber, nach der Standardauffassung, die Möglichkeit einer gesetzesgestützten Erklärung voraussetzt, müssen uns Vertreter der kausalen Handlungstheorie etwas darüber sagen, wie solche Gesetze aussehen könnten. Churchland ist nun, anders als Davidson, der Auffassung, dass es der praktische Syllogismus selbst ist, der die Form eines entsprechenden empirischen Gesetzeschemas angibt. Churchland formuliert sechs Bedingungen, denen die korrekte Formulierung des entsprechenden Gesetzes genügen müssen (a. a. O., S. 313):

- (1) X will \emptyset .
- (2) X glaubt, dass der Vollzug von H unter den gegebenen Umständen ein Mittel ist, \emptyset zu erreichen.

Es gibt keine weitere Handlung, von der X glaubt, dass er mit ihr unter den gegebenen Umständen \emptyset erreichen würde, und für die er eine wenigstens gleich große Präferenz hat wie für H.

X hat keinen anderen Wunsch, der ihn unter den gegebenen Umständen von seinem Wunsch nach \emptyset abbringt.

X weiß, wie man H tut.

X ist in der Lage, H zu tun

Diese sechs Bedingungen versteht Churchland als vollständige Menge von Prämissen des praktischen Syllogismus, so dass sich („praktisch“) folgern lässt:

X tut H.

Ein entsprechendes ‚Handlungsgesetz‘ („L₁“) erhält man, in dem man über X („X“ steht für eine Person), \emptyset („ \emptyset “ steht für einen Zustand in der Welt) und H („H“ steht für eine Handlung) allquantifiziert:

$$(L1) \quad \forall X \forall \emptyset \forall H [(1)-(6) \rightarrow (7)]$$

Über die genannten Bedingungen und die Frage ihrer Vollständigkeit hat es in der Folge eine rege Debatte gegeben. Die für uns wichtigen Punkte sind jedoch davon unabhängig. Das eigentliche Problem, mit dem sich Churchlands Formulierung eines Schemas für Handlungsgesetze konfrontiert sieht, wird, so denke ich, an der Frage deutlich, ob sich entsprechende Gesetze als strikte verstehen lassen sollen, oder als solche, die Ausnahmen zulassen. Deuten wir L₁ als striktes, also ausnahmeloses,

universelles Gesetz, drängt sich nämlich der von den ‚Sprachdualisten‘ vorgebrachte Analytizitätseinwand auf: Die im Vordersatz angegebenen Bedingungen als erfüllt anzunehmen, heißt einfach, X‘ens Tun *als* eine *Handlung H zu verstehen*. Die Unterstellung, dass die Bedingungen (1)–(6) erfüllt sind, ist ein Kriterium dafür, dass überhaupt gehandelt wurde. Und ihr Erfülltsein lässt sich unabhängig von Fällen entsprechenden Handelns prinzipiell auch gar nicht feststellen (bzw. mit Gründen zuschreiben)¹²⁰. Für empirische Gesetze muss aber gelten, dass sie falsch sein können, das heißt, dass die im Vordersatz angegebenen Zustände oder Ereignisse bestehen *könnten*, *ohne* dass die in der Konklusion genannten Zustände oder Ereignisse eintreten.

Wenn man dagegen, anders als Churchland, versuchte, L_1 als nicht-striktes Gesetz zu verstehen, um so dem Analytizitätseinwand zu entgehen, müsste man spezifizieren können, wann das Vorliegen der durch die Prämissen bezeichneten Zustände oder Ereignisse nicht zu den durch die Konklusion angegebenen Zustände oder Ereignisse führt, um falsifizierende von nicht-falsifizierenden Fällen unterscheiden zu können. Mit einer schlichten Ergänzung im Sinne einer *cp*-Klausel ist es also nicht getan, denn dies käme im Falle von L_1 einer ‚analytischen‘ Selbstimmunisierung gegenüber der Möglichkeit falsifizierender Instanzen gleich. Würden wir nun aber die Bedingungen angeben können, unter denen die bisherige Formulierung von L_1 nicht zur beschriebenen praktischen Konklusion führt, müssten wir, wie ich meine, davon ausgehen, dass die Prämissen des praktischen Syllogismus nicht vollständig waren, und nicht etwa annehmen, dass wir seinen Gültigkeitsbereich tatsächlich in spezifischer Weise eingeschränkt hätten. Der praktische Syllogismus ist nämlich – im Sinne eines Rationalitätsprinzips – unser *Kriterium* dafür, bestimmte Überzeugungen oder Wünsche zuzuschreiben, bzw. eine Handlung als solche zu verstehen. Wir haben gewissermaßen keine Wahl: Wir müssen ‚rationalisieren‘, um Handlungen als Handlungen, Überzeugungen als Überzeugungen, Wünsche als Wünsche zu *verstehen*. Der praktische Syllogismus *muss* gelten; falsch sein können Varianten seiner (unzureichenden) Formulierung. Fänden wir eine *Ursache* eines *Verhaltens* von X, die wir nicht im Sinne eines Grundes ‚rationalisieren‘ könnten, könnten wir natürlich auch nicht mehr davon sprechen, seine ‚Handlung‘ – bzw. sein Verhalten – *durch seine Gründe* kausal erklärt zu haben.

Churchland versucht sich allerdings gegen den Analytizitätseinwand zu verteidigen, indem er, bezüglich des „tief verankerten“ Stellenwertes von L_1 , eine Parallele zu einschlägigen fundamentalen Naturgesetzen ziehen zu können glaubt:

¹²⁰ In diesem Sinne auch von Wright (1971/1974), S.171 und Keil (2000), S. 63 f.

„Es ist schwierig, vielleicht sogar unmöglich, L_1 zu bestreiten, ohne die begriffliche Maschinerie zu unterminieren, die solch ein Verständnis ermöglicht oder, besser, *konstituiert*; aber mit all dem ist nicht gesagt, dass L_1 in irgendeinem Sinne ‚analytisch‘ ist, der damit unverträglich ist, dass es ein nomologisches Prinzip ist. Man könnte auch das Prinzip von der Masse-Energie-Erhaltung nicht bestreiten, ohne im begrifflichen Rahmen der modernen Physik eine ähnliche Verwüstung heraufzubeschwören. (...) Gesetzesartigkeit erfordert nicht, dass der betreffende Allsatz unabhängig von dem gesamten Prinzipien-Rahmen, zu dem er wesentlich gehört, falsifizierbar ist.“ (Churchland (1970/1977), S. 317 f.)¹²¹

Diese Erwiderung sticht allerdings nicht, da es beim Analytizitätseinwand gar nicht um eine von der Wirkung unabhängige Feststellbarkeit des Vorliegens von Ursachen oder um eine Falsifizierbarkeit von Gesetzen *unabhängig vom theoretischen Rahmen* geht. Es ist im Falle eines nomologischen Verständnisses des praktischen Syllogismus auch ‚*innerhalb* eines konzeptuellen Rahmens‘ unmöglich, *festzustellen*, dass die Prämissen (etwa die oben durch (1)-(6) benannten) erfüllt sind, ohne dass die entsprechende Handlung ausgeführt wird.

Es ist also nicht plausibel, den unterstellten Kausalcharakter von Handlungserklärungen auf einer gesetzesartigen Interpretation des praktischen Syllogismus gründen zu wollen. Davidson, der am nomologischen Charakter der Kausalität grundsätzlich festhält, beschreitet deshalb auch einen anderen Weg, indem er Verursachungen von ihrer gesetzesartigen *Erklärbarkeit* entkoppelt¹²². Nach Davidson müssen wir nur annehmen, dass es überhaupt irgend ein striktes Gesetz gibt, unter das die durch singuläre Kausalaussagen verknüpften Ereignisse *unter irgendeiner Beschreibung* fallen, wir müssen jedoch nicht davon ausgehen, solche Gesetze zu kennen. Ereignisse individuieren sich nach Davidson durch ihre kausalen Beziehungen zu anderen Ereignissen¹²³. Die Beschreibung dieser individuellen Beziehungen ist extensional, nur gesetzesartige Beschreibungen oder Erklärungen, die über Typen von Ereignissen quantifizieren, sind intensional. Diese ‚Ereignisontologie‘ ist auch der Hintergrund für Davidsons anti-reduktionistischen Materialismus, bzw. seinen sogenannten „Anomalen Monismus“. Davidsons ganze Konstruktion ist ebenso ingeniös, wie gewagt und wirft viele Fragen auf. Insbesondere habe ich große Zweifel,

¹²¹ So etwa auch Fodor in: ders.: *Psychological Explanation* (1968) und A. Beckermann (1977).

¹²² Davidson (1967) und (1970).

¹²³ Diese Bestimmung hat Davidson allerdings in der Zwischenzeit, aufgrund von Einwänden Quines und anderer, zurückgenommen oder zumindest relativiert (s. ders.: ‚Reply to Quine on Events‘ (1985)). Die Einwände liefen im Wesentlichen darauf hinaus, dass die wechselseitigen *Kriterien* der Identität von kausalen Beziehungen einerseits und Ereignissen andererseits quasi zirkulär seien. Davidson neigt nun offenbar zu einer Individuierung von Ereignissen durch Raum-Zeit-Zonen.

ob sich ein Begriff von *Ereignisidentität* verständlich machen lässt, der von bestimmten Weisen der Beschreibungen unabhängig sein soll. Da es in der Folge jedoch zunächst um reduktionistische Programme des Naturalismus gehen soll, stellen wir diese Fragen bis zur Diskussion von Davidsons eigener Position zurück.

Für reduktionistisch orientierte Naturalisten beschränkt sich die Gesetzesartigkeit ‚intentionaler Generalisierungen‘ nun allerdings nicht auf praktische Schlüsse, sondern sie betrifft prinzipiell alle Arten von inferenziellen Zusammenhängen, womit sich etwa Gesetze, wie „(X) (Wenn X glaubt, dass p & q , dann, *ceteris paribus*, glaubt X, dass p)“ ergäben. Dabei scheint grundsätzlich eine *cp*-Klausel geboten. Die unterstellte Fähigkeit zur Verhaltensvorhersage und –Prognose beruhte demnach auf einer weitgehenden Kenntnis solcher (empirischer) Gesetze. Als ein erster Schritt zur reduktiven Reformulierung der ‚Alltagspsychologie‘ kann dann eine, die Qualität entsprechender Erklärungen und Vorhersagen steigernde, Verbesserung der entsprechenden theoretischen Annahmen im Rahmen einer wissenschaftlichen ‚Intentionalen Psychologie‘ gesehen werden. Dies müsste, neben einer Eliminierung von Irrtümern oder falschen Schlüssen, insbesondere durch eine gegenüber der ‚alltagspsychologischer‘ Individuierung präzisere Bestimmung der jene Zustände oder Ereignisse individuierenden ‚intentionalen‘ Gehalte geschehen, über die ‚intentionale‘ Gesetze quantifizieren¹²⁴.

Vor allem aber muss die (vermeintliche) genuine Kausalität oder ‚Kausalkraft‘ mental-kognitiver Eigenschaften schließlich auf der Grundlage eines rein physischen ‚Substrats‘ verständlich gemacht werden können. Dazu bedarf es offenbar Eigenschaftstheorien, die eine materialistische Implementierung alltagspsychologischer Generalisierungen ermöglicht¹²⁵. Dies wird heute meist als im Sinne *multipler* physischer *Realisierbarkeit* verstanden, wobei dem Funktionsbegriff eine wesentliche Rolle zukommt. Dieser Begriff hat im Rahmen verschiedener naturalistischer Strategien allerdings nicht denselben explanativen Stellenwert und wird auch durchaus unterschiedlich gedeutet (und wir werden unter anderem fragen müssen, ob nicht manche derjenigen naturalistische Theorien des Geistes, die sich „funktionalistisch“ nennen, von vornherein eine inadäquate Bestimmung dieses Begriffs vornehmen).

Bevor wir uns nun aber mit solchen Theorien des Geistes auseinandersetzen, die sich in irgendeiner Weise am Funktionsbegriff orientieren, sollten wir aber die wichtigsten, das IBS-Programm bzw. den Reduktiven Materialismus kennzeichnenden Punkte, nochmal zusammengefasst werden: IBS ist als das offensichtlich

¹²⁴ Vergl. D. Dennett: ‚Three Kinds of Intentional Psychology‘ (1981).

¹²⁵ S. hierzu und zum Folgenden: Fodor (1994).

anspruchsvollste unter den Projekten einer naturalistischen Konzeption von Bedeutung und Geist anzusehen – immerhin soll *Bedeutung materialistisch erklärt* werden können. Dazu müsste zunächst sprachliche Bedeutung, als eine Eigenschaft von Token, in Begriffen von psychologischer Sprecher- und Hörereinstellungen analysierbar, bzw. Bedeutung *qua* ‚Sprecherbedeutung‘ auf ‚intentionale‘ Einstellungen zurückführbar sein. Verständigung soll dabei zwar nicht in Begriffen von physischen Reiz-Reaktions-Mustern, sondern auf einer Ebene ‚intentionaler‘ Begrifflichkeit beschrieben werden. Doch diese Begrifflichkeit soll schließlich ihrerseits in eine materialistische Redeweise überführbar sein. Eine reduktive Theorie, die ‚intentionale Eigenschaften‘ als physische erklären können will, müsste dementsprechend schließlich in der Lage sein, *naturalistisch hinreichende Bedingungen* dafür zu liefern, dass etwas in einem *bestimmten* ‚intentionalen Zustand‘ ist. Wäre dies möglich, würde unser jeweiliges Verständnis von bedeutungsvollen Äußerungen – beziehungsweise das Verständnis von ‚rationalem Verhalten‘ allgemein – der Kenntnis einer (mehr oder minder guten) empirischen Theorie entsprechen, die prognostisch (mehr oder minder) äquivalent mit einer empirischen Theorie über die kausal zugrundeliegenden ‚psychologischen Mechanismen‘ behavioraler Ereignisse wäre.

Nun geht zwar offenbar *jedes* als *naturalistisch* einzustufende Projekt davon aus, dass Erklärungen von Handlungen (bzw. Verhalten) Kausalerklärungen von Ereignissen sind, denen bestimmte Gesetze zugrunde liegen müssen. Es ist jedoch strittig, ob es, erstens, genuin ‚intentionale Gesetze‘ gibt und ob, zweitens, wenn es sie gibt, systematische (gesetzesartige) Zusammenhänge zwischen den ‚intentionalen‘ und den physischen Beschreibungen derjenigen Zustände oder Ereignisse, die in Kausalerklärungen des Verhaltens eine Rolle spielen sollen, gefunden werden können (also beispielsweise zwischen einer Beschreibung wie „X hat die Überzeugung *Ü*“ und einer Beschreibung wie „X ist in einem neuronalen Zustand *N*“). Nach Ansicht der Vertreter des IBS-Programmes sind ‚intentionale Gesetze‘ nicht nur genuin, sondern es lassen sich auch gesetzesartige psycho-physische Beziehung oder doch zumindest hinreichende Bedingungen einer physische Instantiierung ‚intentionaler‘ (semantischer) Eigenschaften angeben. Die Aufgabe besteht demnach darin, solche Bedingungen oder entsprechende Instantiierungsgesetze für ‚intentionale Eigenschaften‘ zu formulieren. Reduktive Materialisten orientieren sich dabei heute – im Gegensatz zu kruden Formen psycho-physischer Reduktion – vorwiegend an einer *funktional vermittelten*, multiplen Realisierbarkeit ‚intentionaler Eigenschaften‘. Die unmittelbar implementierenden Mechanismen für ‚intentionale Gesetze‘ mit funktionalistisch definierten Eigenschaften könnten dementsprechend sowohl neuronal, als auch ‚*computational*‘ bzw. syntaktisch definierte sein. Die Hoffnung, die sich mit dem sogenannten „Computermodell des

Geistes“ verbindet, besteht dabei darin, mit der Funktionsweise von Computern einen – im Gegensatz zum Gehirn – gut verstandenen Mechanismus zur Verfügung zu haben, mit dessen Hilfe die Ebene der implementierenden Mechanismen in einer Weise beschrieben werden könnte, die sowohl die semantisch-logischen, als auch die unterstellten kausalen Eigenschaften ‚intentionaler‘ Zustände auf sich zu vereinigen vermag. Doch bevor wir auf dieses Modell genauer eingehen, sind noch einige allgemeine systematische Bemerkungen zum Funktionsbegriff angebracht

3 ‚Funktion‘ als Schlüsselbegriff einer materialistischen Theorie des Geistes

Eine grundsätzliche Skepsis gegenüber einer direkten bzw. eindeutigen ‚psycho-physischen‘ Typenidentifizierbarkeit ist heute auch unter Naturalisten gang und gäbe. Dies lässt sich schon an der spätestens mit den siebziger Jahren einsetzenden Dominanz des Computermodells des Geistes¹²⁶ festmachen, denn ein Computer ist, physisch gesehen, offensichtlich völlig anders aufgebaut, als ein menschliches Gehirn. Zwar traten moderne Theorien direkter psycho-physischer Typenidentität¹²⁷ häufig mit Beispielen an, die ihrer These noch am ehesten genügen zu können schienen. Damit meine ich jene Aspekte des Mentalen, wie Empfindungen (z.B. Schmerzen), die *nicht* Thema dieser Arbeit sind. Ob eine entsprechende Identifizierbarkeit mit (neuronalen) Gehirnzuständen oder -Vorgängen plausibel gemacht werden kann, möchte ich bezweifeln, es ist jedoch nicht unser Thema. Die Typenidentitätsthese wird allerdings meist so verstanden, dass sie sich auf alle Bereiche mentaler Redeweise erstreckt. Die Unterstellung einer physischen Gleichheit all jener Zustände oder Ereignisse, die eine bestimmte ‚mentale Eigenschaft‘ realisieren sollen, ist aber gerade auch bezüglich propositionaler Einstellungen in Verruf geraten. Wenn es im Hinblick auf die Zuschreibbarkeit ‚intentionaler Einstellungen‘ zumindest zwischen Menschen und

¹²⁶ Seit Mitte der achtziger Jahre hat dieses Modell allerdings zunehmend Konkurrenz von sogenannten ‚konnektionistischen‘ Theorien kognitiver Phänomene bekommen.

¹²⁷ Z. B. J.J.C. Smart: ‚Sensations and Brain Processes‘ (1959). Für eine frühe, einflussreiche Kritik ‚psycho-physischer‘ Typenidentität, s. Putnam (1960).

Tieren, vielleicht aber auch Artefakten, keinen Unterschied geben soll, führte dies zu der unplausiblen Annahme, dass in jedem dieser unterschiedlichen Fälle die physische Instantiierung einer ‚Überzeugung‘, wie beispielsweise, dass dort etwas Essbares sei, typidentisch sein müsse. Vielen erschien vor allem diese Konsequenz der direkter Typenidentitätstheorie, nämlich dass *jede mögliche* Form der physischen Realisierung einer propositionalen Einstellung jener bei uns Menschen entsprechen müsste, als eine Art ungerechtfertigter ‚Humanimperialismus‘ (auch gegenüber vielleicht völlig andersartig aufgebauten ‚Marsianern‘). Es ist darüber hinaus allerdings auch fragwürdig, ob es plausibel ist, für alle Menschen eine entsprechende identische Realisierungsweise anzunehmen. Unter anderem schon angesichts der Unterschiedlichkeit der Umstände, unter denen wir unterschiedlichen Subjekten identische propositionale Einstellungen zuschreiben, spricht eigentlich nichts für eine ‚harte‘, physische Identität.

Nun spielen Typenidentitätstheorien in der gegenwärtigen Forschungspraxis der *Cognitive Sciences* ohnehin keine große Rolle, und zwar schon deshalb, weil eine ‚*Bottom-up*-Strategie‘, die bei der Erklärung kognitiver Phänomene *atomistisch*, also z. B. bei Neuronen, ihrer Funktionsweisen und ihren Vernetzungen ansetzte, insofern wenig Aussicht auf Erfolg verspricht, als der neuronale Aufbau des Gehirns einfach zu komplex ist und sich eine Untersuchung des Zusammenwirkens einzelner, sensorischer Nervenstränge als für jede Art *kognitiver* Beschreibung oder Erklärung unbrauchbar erweist¹²⁸. Die aussichtsreichere Alternative scheint eine ‚*Top-Down*-Strategie‘ zu sein, die mit einer abstrakter angelegten *Dekomposition* der höchsten Ebenen ‚psychologischer Organisation‘ beginnt und dann stufenweise fortfährt, „diese in immer detailliertere, kleinere Systeme oder Prozesse zu analysieren, bis man am Ende bei Elementen angelangt, die dem Biologen bekannt sind“¹²⁹.

Bei *Top-Down*-Erklärungsweisen von Systemen ist typischerweise auf einer ‚oberen‘ Ebene von *Funktion* (von Subsystemen) die Rede. Diese Redeweise können wir hier, ganz allgemein, vorläufig so erläutern, dass mit ihr von bestimmten Details der physischen Realisierung eines Dinges oder Vorganges zugunsten jener Aspekte abstrahiert wird, die in der jeweiligen Situation in irgendeinem Sinne relevant sind oder bestimmte Erklärungsinteressen befriedigen. Üblicherweise bedeutet eine Funktion zu haben dabei, unter bestimmten Bedingungen einen bestimmten Zweck zu erfüllen – oder, vorsichtiger ausgedrückt, zu einem bestimmten Ergebnis zu führen. Subsysteme sind durch die Funktion bestimmt, die sie zum Bestand, bzw. zur Möglichkeit des Funktionierens des übergeordneten Systems leisten. So kann das Herz – ebenso, wie

¹²⁸ Vergl. z. B. G. Roth: *Das Gehirn und seine Wirklichkeit* (1994).

¹²⁹ D. Dennett: ‚*Artificial Intelligence as Philosophy and as Psychology*‘ (1978b), S. 110 (meine Übersetzung).

andere Organe oder Vorgänge (etwa der Stoffwechsel) – als ein Subsystem unter anderen verstanden werden, welches einen funktionalen Beitrag zum Bestand des übergeordneten Systems „Menschlicher Organismus“ leistet¹³⁰. Um diesen Beitrag auf einer oberen Erklärungsebene verstehen zu können, kommt es nicht auf die jeweilige physische Realisierungsform eines Herzens an, sondern auf dessen funktionale Rolle innerhalb des Gesamtsystems (diese Rolle kann im Fall des Herzens, wie wir wissen, mittlerweile durch technische Apparate *funktional* ersetzt werden). Es ist für Funktionen, bzw. funktionale Eigenschaften, wesentlich, dass sie *multiple realisierbar* sind¹³¹.

Eine implementierende Erklärung von zunächst funktional beschriebenen Zuständen oder Ereignissen sollte, so die Grundidee eines *Top-Down-Ansatzes*, am ‚unteren Ende‘ etwa zu einer chemischen oder gar physikalischen Erklärung der jeweiligen physischen Realisierungen der Funktionen führen. Die in reduktiven Erklärungen funktionaler Ereignisse zum Zuge kommenden implementierenden Mechanismen sind dabei – im Sinne multipler Realisierbarkeit – nicht bereits durch die Art der Funktion auf bestimmte physische Eigenschaftstypen festgelegt. Außerdem würde eine etwa physikalistische Reduzierbarkeit nicht bedeuten, dass die höherstufigen (funktionalen) Eigenschaften des Systems explanativ überflüssig würden.

Allerdings müssen wir hinsichtlich des Funktionsbegriffs noch zwei Unterscheidungen beachten: So wird einerseits auch – anders als in unserer obigen Erläuterung – in einem *nicht-teleologischen* Sinne von Funktionen gesprochen. Eine Funktion zu haben soll dann bedeuten, eine spezifische *kausale* Rolle auf einer Ebene der Beschreibung einzunehmen, die für bestimmte Formen von Erklärung von Eigenschaften übergeordneter Systeme adäquat ist und bei der die jeweilige physische Realisierung der kausalen Rolle explanativ irrelevant ist¹³². Auf die Debatte zwischen Vertretern eines nicht-teleologischen oder kausalen und Vertretern eines teleologischen oder normativen Verständnisses des Funktionsbegriffs werden wir unten noch ausführlich eingehen müssen.

¹³⁰ Diese Erläuterung der Begriffe „Funktion“, „System“ und „Subsystem“ entspricht einer relativ neutralen Formulierung des gängigen *teleologischen* Verständnisses funktionaler Erklärungen. Wie sich solche Verständnisse naturalistisch ummünzen lassen könnten, was man also unter einer funktionalen Erklärung im Allgemeinen (wenn eine solche allgemeine Bestimmung überhaupt möglich sein sollte) und unter einer funktionalen Erklärung des Geistes im Besonderen zu verstehen hätte, ist aber, wie wir noch sehen werden, strittig (vergl. hierzu etwa V. G. Hardcastle: ‚Understanding Functions: A Pragmatic Approach‘ (1999)) – insbesondere hinsichtlich der Frage, ob funktionalistische Theorien, bzw. ihre begrifflichen Grundlagen, überhaupt teleologisch oder gar normativ sind.

¹³¹ „Multiple Realisierbarkeit“ bedeutet, formal ausgedrückt, dass die Instantiierung der zu implementierenden Eigenschaft hinreichend für eine bestimmte Disjunktion von Eigenschaften auf der Ebene der Implementierung, aber nicht für eines ihrer Disjunkte ist.

¹³² So etwa bei Cummins (1983)

Die andere Unterscheidung geht auf einen Vorschlag Schiffers zurück, nach dem es, im Rahmen der hier zur Debatte stehenden Fragen auch sinnvoll ist, zwischen funktionalen und funktionalistischen Theorien zu unterscheiden¹³³. Schiffer bestimmt *funktionale* Theorien dabei generell als Lösungsstrategien hinsichtlich sogenannter ‚Black-Box-Probleme‘, mit deren Hilfe versucht wird, zu Erklärungen von oder Voraussagen über das Verhalten von Systemen zu gelangen, von denen zwar die Inputs und Outputs bekannt sind, nicht aber ihr interner Aufbau, bzw. die Natur ihrer inneren Zustände und die kausalen Gesetze, denen sie unterliegen. Dies bedeutet für Schiffer, dass in funktionalen Theorien über das Verhalten solcher Systeme bestimmte Zustandstypen postuliert werden, die in diesen oder jenen kausalen oder transitorischen Weisen untereinander und mit den Inputs und Outputs des Systems verbunden sind. Wenn eine solche Theorie detailliert genug und richtig ist, können wir, aufgrund der Kenntnis des Inputs, den Output, bzw. das Verhalten eines betreffenden Systems vorhersagen¹³⁴. In diesem Sinne ist die Kenntnis eines Computerprogramms eine funktionale Theorie, die, bei gegebenen Inputs (z.B. einer bestimmten Eingabe) Vorhersagen über den Output (z.B. eine bestimmte Anzeige auf dem Bildschirm) ermöglicht, selbst wenn wir kaum etwas über den Aufbau der *Hardware* wissen.

Im Rahmen von *funktionalen Theorien über ‚Intentionale Systeme‘* können, nach Schiffer, Propositionen (bzw. propositionale Gehalte und Einstellungsmodi) die Funktion externer Indizes für die funktionalen Rollen übernehmen, welche den internen physischen Zuständen des Systems zugeschrieben werden sollen. Diese Zuschreibungen stellen dann empirische Mutmaßungen hinsichtlich möglicher Entsprechungen kausaler oder transitorischer Zustände des Systems mit (im weitesten Sinne) logischen Verhältnissen zwischen Propositionen dar. Diese Mutmaßungen lassen sich zunächst nur über ihren Erfolg bei Vorhersagen des Verhaltens eines ‚Intentionalen Systems‘ überprüfen. In funktionalen Theorien müssen jedoch

¹³³ Schiffer (1987), Kap. 2.

¹³⁴ Ein bekanntes und anschauliches Beispiel für ein System, dessen Zustände in diesem Sinne rein funktional bestimmt sind, stammt von Ned Block (in: ders.: ‚Troubles with Functionalism‘ (1978)): Denken wir uns einen Getränkeautomaten, der als möglichen Input Ein- oder Zweimarkstücke, als möglichen Output ein Bier und ein Einmarkstück hat. Nach einigen Versuchen können wir sein ‚Verhalten‘ dadurch erklären (und vorhersagen), dass wir zwei möglich innere Zustände postulieren. Der erste (Z_1) ist funktional dadurch definiert, dass der Automat ein Bier ausspuckt und dann in diesem Zustand bleibt, wenn zwei Mark eingeworfen werden und außerdem dadurch, er in (Z_2) übergeht, wenn nur eine Mark eingeworfen wird. Der zweite (Z_2) ist dann dadurch definiert, dass er ein Bier und ein Markstück ausspuckt und dann in (Z_1) zurückgeht, wenn zwei Mark eingeworfen werden und außerdem dadurch, dass er ein Bier ausspuckt und dann in (Z_1) zurückgeht, wenn eine Mark eingeworfen wird. *Wie* der Automat diese funktionalen Zustände genau realisiert, ist für eine entsprechende ‚Verhaltensklärung‘ irrelevant.

‚Propositionen‘ (bzw. propositional individuierte Einstellungen) bei der physischen Erklärung des Verhaltens des Systems keine Rolle spielen¹³⁵.

Im Sinne der ersten der hier in Frage stehenden Unterscheidungen des Funktionsbegriffs, können wir bereits hier festhalten, dass sich Schiffer bei seiner Charakterisierung funktionaler Theorien auf eine Sichtweise von Systemen und Funktionen beschränkt, in der auf intentionale (bzw. normative) Idiome verzichtet wird. Es ist aber fraglich, ob er damit auch den üblichen Verständnissen des Funktionsbegriffs in dem wichtigen Bereich biologischer Erklärungen gerecht wird. Denn in diesen scheint unter der Funktion eines (Sub-)Systems, Objektes oder Prozesses doch eher der Beitrag verstanden zu werden, den es/er normalerweise für die Funktionsweise oder den Bestand des übergeordneten Systems leistet. So geht es bei funktionalen Erklärungen auch nicht in erster Linie um Vorhersagen, sondern um ein besseres Verständnis dafür, wie die Bestandteile eines Systems zusammenwirken müssen, um das System zu dem zu machen, was es jeweils ist (wobei von den für die Erfüllung der Funktion unwesentlichen Eigenschaften ihrer jeweiligen Realisierung abstrahiert wird). Ein funktionales Verständnis ermöglicht es dann einerseits, bei eintretenden oder absichtlich hervorgerufenen *Veränderungen von Systembestandteilen* eine Abschätzung der *Folgen im Hinblick auf das Funktionieren oder die Stabilität des Gesamtsystems* vorzunehmen, bzw. das *Niveau der erforderlichen Typidentität hinsichtlich einer Ersetzbarkeit von Systembestandteilen* zu klären. Und es ermöglicht andererseits, bei *vorliegenden Störungen des Gesamtsystems*, deren mögliche *Ursache* durch die Feststellung einer *Dysfunktionalität* eines Systembestandteils in einer für das Verständnis der *Störung relevanten* Weise zu *spezifizieren*. Darauf werden wir noch zurückkommen.

Im Gegensatz zu funktionalen sollten *funktionalistische* Theorien, nach Schiffer, nicht als lediglich empirische, sondern als zunächst *philosophische* Unternehmen verstanden werden, die die Frage beantworten sollen, wie und in welchem Sinne ‚intentionale Eigenschaften‘ (also etwa in einem Zustand zu sein, der in einer bestimmten Relation zu einer Proposition steht) als funktionale Eigenschaften verstanden, bzw. mit diesen identifiziert werden können; sie sind *Theorien über die*

¹³⁵ Eine in diesem Sinne funktionale Theorie der Verhaltenserklärung von Personen stellt etwa R. Stalnakers (1984) dar. Die Idee, Propositionen als externe Indizes zu verstehen, wird durch Stalnaker mit der Analogie der Rolle von Zahlen bei der quantitativen Beschreibung der Eigenschaften physischer Dinge erläutert: Bestimmte Eigenschaften physischer Dinge, wie etwa ihr Gewicht, können in eine strukturelle Identitätsbeziehung zu Zahlen gesetzt werden und somit als bestimmte Eigenschaften (z. B. 5 kg) einer Eigenschaftsfamilie (z.B. Gewicht) herausgegriffen werden. Entsprechendes sollen Propositionen für ‚intentionale Eigenschaften‘ leisten: „The analogy suggests that to define a relation between a person or physical objekt and a proposition is to define a class of properties with a structure that makes it possible to pick one of the properties out of the class by specifying the proposition“ (ebd., S. 11).

Natur propositionaler Einstellungen (a. a. O., S. 26). Den Gegenstand meiner naturalismuskritischen Überlegungen sollen nun nicht nur *funktionalistische Theorien* propositionaler Einstellungen bilden, sondern auch *funktionale* Theorien, nämlich dann, wenn sie mit dem *philosophischen* Anspruch auftreten, reduktive Erklärungen von ‚Intentionalität‘ im Sinne von Eigenschaftstheorien darzustellen.

Auch bei seiner Charakterisierung philosophisch ambitionierter, also funktionalistischer Theorien beschränkt sich Schiffer wiederum auf ein nicht-telologisches Verständnis, bei dem propositionale Einstellungen mit bestimmten kausalen Eigenschaften (bzw. ihren kausalen Rollen) identifiziert werden. Die systematische Pointe der Schifferschen Unterscheidung ist jedoch von dieser Beschränkung unabhängig. So haben sich in jüngerer Zeit eher die teleologischen Varianten *funktionalistischer* Theorien der ‚Intentionalität‘ in den Vordergrund gedrängt, Theorien, die Funktionen ätiologisch, also etwa evolutionistisch bestimmen.

Grundsätzlich war der Aufstieg des Funktionsbegriffs im Zusammenhang mit modernen Varianten naturalistischer Theorien des Geistes allerdings zunächst eng mit der zunehmenden Orientierung am Computermodell des Geistes verknüpft. Um den Charme zu verstehen, den dieses Modell nicht nur für naturalistisch gesinnte Philosophen, sondern auch für Forscher im Bereich der Entwicklung von Systemen mit sogenannter ‚künstlicher Intelligenz‘ weiterhin hat, müssen wir es zunächst eingehender erläutern. Dabei wird sich allerdings zeigen, dass der Stellenwert dieses Modells in den verschiedenen Ansätzen durchaus unterschiedlich beurteilt wird, was nicht zuletzt mit der jeweiligen Verwendung des Ausdrucks ‚Funktion‘ im Sinne der oben gemachten Unterscheidungen zusammenhängt.

Das Computermodell des Geistes

Das Paradigma für ‚*Top-Down-Strategien*‘ im Bereich naturalistischer Theorien des Geistes ist die Implementierung eines Computerprogramms. Die Grundidee des Computermodells des Geistes (CMG) besteht dabei darin, dass das Verhältnis von menschlichem Geist^{PE} zu seiner materiellen Grundlage, dem Gehirn, grundsätzlich dem von Computern zugeschriebenen kognitiven Fähigkeiten zu den jeweiligen, irgendwie physisch realisierten, Computerprogrammen entsprechen soll. Computer werden dabei als ein Modellfall verstanden, der immerhin zeige, *wie* eine physische Realisierung von ‚Intentionalität‘ *überhaupt verstehbar* sein kann. Denn Computer führen Operationen aus, die wir fraglos als *kognitive* beschreiben können, und sie zeigen ein Verhalten,

welches, in Analogie zum Menschen, durchaus als *intelligentes, zielgerichtetes Lösen von Problemen* beschrieben werden kann. Die Zuschreibung solcher kognitiver Fähigkeiten aber geht Hand in Hand mit der Zuschreibung propositionaler Einstellungen. Andererseits hat der innere Aufbau von Computern, im Gegensatz zum Gehirn, nichts Mysteriöses an sich – sie sind ja von Menschen selbst konstruiert worden. Wenn wir „Geist“ als substantivierenden Begriff für kognitive Fähigkeiten oder Leistungen, wie etwas erinnern, planen, berechnen, Schlüsse ziehen usw. deuten und das Gehirn des Menschen als ein (wiederum selbst funktional zu analysierendes) Subsystem des menschlichen Organismus verstehen, dessen Funktion vor allem auch darin besteht, solche kognitiven Leistungen zu ermöglichen, dann könnten Computer (bzw. deren funktionale Systembestandteile), die entsprechendes zu leisten vermögen, als dem menschlichen Gehirn (bzw. dessen funktionalen Systembestandteilen) funktional äquivalente Systeme verstanden werden. Da nun kognitive Fähigkeiten beim Menschen häufig die Eigenschaften haben, begrifflich und propositional zu sein, würden wir – so die grundlegende Intuition der Vertreter des CMG – auch einem naturalistischen Verständnis von Bedeutung, Begrifflichkeit und Propositionalität auf die Sprünge helfen, wenn wir uns, anstatt uns damit abzumühen, das Gehirn mit seiner extrem komplexen Struktur zu untersuchen, darauf besännen, nach welchen Prinzipien wir (mit schrittweise steigendem Komplexitätsniveau) dasjenige selbst entwickeln, welches den kognitiven Vermögen des Gehirns inzwischen offenbar immer näher kommt, nämlich Computer (bzw. computergestützte Apparate).

Allerdings ist auch unter Naturalisten strittig, ob das CMG zu einer Erklärung der Natur ‚intentionaler‘ Einstellungen hinreichend ist. Denn aus der grundsätzlichen Idee, die ‚rein syntaktische‘ Funktionsweise von Computern, also eine regelgeleitete, schrittweise Umformung von Symbolen, als ein Paradigma der physischen Realisierung des Geistes^{PE} zu verstehen, ergibt sich ja nicht gleichsam automatisch auch schon die Antwort auf die Frage, wie der semantische Gehalt jener ‚Symbole‘ naturalistisch erklärt werden könnte. Ohnehin wird häufig angezweifelt, dass Artefakte, wie Computer, *per se* tatsächlich ‚originäre Intentionalität‘ generieren. Die den Artefakten zugeschriebene ‚Intentionalität‘ wird dann als lediglich von ihren Erbauern oder Benutzern hergeleitete relativiert; deren ‚ursprüngliche‘ Intentionalität wiederum sei aber durch Computermodelle nicht wirklich erklärbar. So dient das CMG für manche Theoretiker lediglich dazu, zu zeigen, dass ein Vorgang oder ein Verhalten, der bzw. das berechtigterweise auf der Ebene von ‚intentionalen‘ Eigenschaften beschrieben oder erklärt wird, auf rein physischen Mechanismen aufrufen, bzw. durch physische Gesetzmäßigkeiten implementierbar sein *kann*. Wenn aber die ‚intentionalen Eigenschaften‘ von Computern nicht genuin wären, müsste der Naturalist natürlich

andere Kriterien – etwa genealogische – für die *berechtigte* Zuschreibbarkeit von *originärer* Begrifflichkeit und Propositionalität liefern, als jene, die sich bereits aus einer Analyse der *Funktionsweise* von ‚Intentionalität‘ lediglich *simulierenden* Maschinen gewinnen lassen.

Dies können wir hier allerdings zunächst einmal zurückstellen. Denn wie auch immer man sich zum Status ‚künstlicher Intelligenz‘ verhält, klar ist, dass Reduktive Materialisten, die ‚intentionale Tatsachen‘ (wo auch immer diese originär realisiert sein mögen) *als genuine* ernst nehmen wollen – und insofern einen ‚realistischen‘ Standpunkt vertreten –, in einer ‚nicht-intentionalen‘ Terminologie erklären können müssten, wie oder wodurch ‚intentionale‘ Zustände oder Ereignisse jene semantischen Eigenschaften haben, die sie nun mal haben. Nach dem CMG wäre der wesentliche Schritt zu einer materialistischen Reduktion des ‚Intentionalen‘, getan, wenn es gelänge, ‚intentionale Gesetze‘ auf einer kalkülmäßigen bzw. computationalen Ebene zu implementieren; oder, anders ausgedrückt, „if the immediately implementing mechanisms for intentional laws are computational, then we need a property theory that provides computationally sufficient conditions for the instantiation of intentional properties“¹³⁶. Wäre dies in einem plausiblen Sinne zu leisten, hätten wir es quasi schon mit einer materialistischen Reduktion zu tun, denn Computerprogramme sind sowohl in ihrer Funktionsweise verstanden und erklärbar als auch physisch realisiert.

Dazu müssten Reduktive Naturalisten allerdings zunächst eine naturalistische Reformulierung von ‚intentionalen Eigenschaften‘, insbesondere von propositionalem Gehalt vorlegen. Denn um ‚Intentionale Gesetze‘ computational implementieren zu können, müssen intentionale Zustände oder Ereignisse ja ihrerseits in einer kausalen Terminologie definierbar sein. Entsprechenden Versuchen werden wir uns später noch ausführlich zuwenden. Da aber nicht alle Anhänger des CMG Reduktive Materialisten sind und einen realistischen Standpunkt gegenüber ‚intentionalen Tatsachen‘ einnehmen, können und sollten wir das CMG unabhängig von der Frage der Reduzierbarkeit ‚intentionaler Tatsachen‘ durch entsprechende Eigenschaftstheorien diskutieren. Wer unserer ‚intentionalen‘ Redeweisen für *loose talk* hält (wie etwa Dennett), braucht keine strenge Reduktion, sondern nur eine plausible Erklärung für einen mehr oder minder gelungenen systematischen ‚Gleichklang‘ zwischen den *logisch-semantischen* Eigenschaften propositionaler Einstellungen und den *kausalen* Eigenschaften ihrer physischen Realisierung.

Ich werde also den zweiten Schritt von dem ersten machen und mich zunächst um die Frage der physischen Realisierung von – *for the sake of the argument* als naturalistisch reformulierbar unterstellten – ‚intentionalen Gesetzen‘ kümmern. Vor

¹³⁶ Fodor (1994b), S. 298.

allem gilt es die zentrale These des CMG näher zu erläutern, wonach jede Realisierung von Geist_{pe}, also auch unsere spezifisch menschliche *qua* eines Gehirns (und Nervensystems), prinzipiell auf der Grundlage eines *Automatischen Formalen Systems* (AFS), bzw. einer *regelgeleiteten Manipulation von Symbolen* (oder *Repräsentationen*) verständlich gemacht werden könne¹³⁷. Diese These wird ja auch von solchen Philosophen vertreten, die keine *funktionalistische* Theorien propositionaler Einstellungen bzw. von ‚Content‘ vertreten und die Zuschreibung solcher Einstellungen nicht als eine Beschreibung ‚harter‘ Fakten deuten.

Formale Systeme lassen sich zunächst mit Spielen vergleichen, in denen Token (z.B. einzelne Spielfiguren oder Instantiierungen von Symbolen, bzw. ‚Zeichenträger‘) nach bestimmten Regeln manipuliert werden können, um neue Konfigurationen zu erreichen. Solche Systeme sind insofern *in sich geschlossen*, als alles, was nicht in der aktuellen Konfiguration des Systems enthalten ist, für die jeweils bestehenden Manipulationsmöglichkeiten keine Rolle spielt. Die Regeln, nach denen ein Token jeweils manipuliert werden kann (darf), entsprechen seinem Typ. Die aktuelle Konfiguration des Systems (z.B. eine bestimmte Stellung der Spielfiguren beim Schach) und der symbolische Typ, den ein Token instantiiert (z.B. „Turm“) legen fest, welche (erlaubten) Manipulationsmöglichkeiten jeweils bestehen. Formale Systeme sind *eindeutig bestimmt*, das heißt, es gibt keine Unklarheiten, Vagheiten oder Doppeldeutigkeiten hinsichtlich der jeweiligen ‚Position‘ oder der Manipulationsmöglichkeiten der Token. Außerdem müssen die *Verfahren der Feststellbarkeit* der Regelmäßigkeit der Manipulationen für jede mögliche Konfiguration *endlich* sein. Geschlossenheit, Bestimmtheit und endliche Feststellbarkeit bedeuten, dass Formale Systeme *digital* sind. Alle Fragen zur Konfiguration oder zur Zulässigkeit bestimmter Manipulationen lassen sich mit „Ja“ oder „Nein“ beantworten. Eine wichtige Eigenschaft Formaler Systeme ist auch ihre *analytische Systematizität*: Komplexe (zusammengesetzte) *Token* einer höheren Ebene können der Konfiguration einer tiefer liegenden Ebene entsprechen, wobei erstere sich aus einer relativ geringen Anzahl einfacher (atomarer) Typen nach bestimmten Regeln systematisch aufbauen lassen.

Aus dem Gesagten ergeben sich zwei wesentliche Eigenschaften Formaler Systeme: Erstens sind sie unabhängig von dem (physischen) Medium, in dem sie ‚verkörpert‘ bzw. realisiert sind. Zwei Formale Systeme sind dann äquivalent, wenn es zwischen

¹³⁷ Die folgenden Ausführungen orientieren sich vor allem an J. Haugeland: *Artificial Intelligence: The very Idea*.(1985); vergl. auch: J. Kim: *Philosophy of Mind* (1996) und Cummins (1983), Kap. II.

ihren Konfigurations- und Manipulationsmöglichkeiten eindeutige Entsprechungsbeziehungen gibt, also unabhängig davon, ob zwischen der jeweiligen materiellen Art und Weise ihrer Realisierung eine Ähnlichkeit besteht. Und zweitens sind *semantische* Fragen für die Eigenschaften Formaler Systeme insofern irrelevant, als die ‚Bedeutungen‘ (der jeweilige Typ) von Token rein syntaktisch, durch ihre formal bestimmten Manipulationsmöglichkeiten und die jeweilige internen Positionen festgelegt sind (was der Eigenschaft ‚interner Geschlossenheit‘ entspricht).

Formale Systeme sind in Computerprogramme (*Software*) umsetzbar, die sich ohne prinzipielle Schwierigkeiten über einen hinreichend komplexen Aufbau von binären Instanzen physisch (als *Hardware*) realisieren lassen. *Automatische* sind Formale Systeme dann, wenn sie durch eine Maschine realisiert sind, die die *Symboltoken* eines solchen Systems – nach Maßgabe eines formaler Algorithmus – regelmäßig manipulieren, bzw. erzeugen und verändern kann. Das heißt, alle Operationen eines solchen Systems laufen stur nach einem bestimmten Muster ab. Dies geschieht, im Hinblick auf ein klar definiertes Ziel, schrittweise und auf der Grundlage einer endlichen Menge einfacher Grundoperationen. Ein einschlägiges Beispiel bilden Schachcomputer, die zugleich auch ein für den Versuch der maschinellen Simulation geistiger Fähigkeiten wichtiges Phänomen verdeutlichen: Interessant für die Kognitionsforschung werden Formale Systeme nämlich eigentlich erst dadurch, dass sie, wie Spiele, nicht-deterministisch sind; das heißt, bei den meisten Konfigurationen gibt es nicht nur einen, sondern mehrere mögliche (‚erlaubte‘) Züge, wobei einige davon zur Erreichung eines bestimmten Zieles (bzw. der Erfüllung einer bestimmten Funktion) besser geeignet sein können, als andere. Die Frage ist dann, wie ein AFS vorgehen soll, wenn es, in einer Situation sich verzweigender Alternativen, auf der Suche nach der besten – oder immerhin einer ‚guten‘ – ‚Entscheidung‘ von den formalen *Erlaubnisregeln* ‚im Stich gelassen wird‘.

Eine jeweils befriedigende Strategie zur Lösung dieses wohlbekannten Problems (des sogenannten ‚Steuerungs-Problems‘) zu finden, ist die Hauptschwierigkeit bei der Schaffung eines AFS, welches Fähigkeiten haben soll, die mit bestimmten kognitiven Kompetenzen von Menschen zumindest vergleichbar sind. Die Schwierigkeiten beruhen einerseits auf der sich mit jedem Schritt exponentiell steigenden Zahl zu berücksichtigender, formal möglicher Alternativen, der sogenannten ‚kombinatorischen Explosion‘, wobei die Kapazitätsgrenzen digital arbeitender Systeme eine natürliche Beschränkung darstellen. Und sie beruhen andererseits auf der Tatsache, dass mögliche Zielvorgaben nicht immer so eindeutig definiert sind, wie etwa beim Schach. Der Anspruch jener Kognitionsforscher, die sich am CMG orientieren, besteht ja immerhin darin, kognitive Kompetenz im Sinne von Geist^{PE} nach dem Muster eines AFS zu

verstehen, und dieser Anspruch soll (unter anderem) eingelöst werden können, wenn es gelänge, Maschinen mit Fähigkeiten zu konstruieren, die wir in einem vollen Wortsinne „geistig“ nennen müssten, also etwa, eine normale Konversation zu führen¹³⁸. In vielen Alltagssituationen des praktischen Umgangs, wie etwa dem Austragen einer Meinungsverschiedenheit oder bei einem Einkaufsbummel, gibt es jedoch meist keine so *eindeutigen* Zielvorgaben (wie beispielsweise jene, den Gegenspieler ‚Matt‘ zu setzen). Es scheint damit für nicht-deterministische Systeme kein *algorithmisches* Verfahren zur Wahl einer jeweils *besten* – oder doch zumindest einer guten – ‚Entscheidung‘ zur Verfügung stehen zu können, weil einerseits die nachfolgenden Kombinationsmöglichkeiten exponentiell ‚explodieren‘ und andererseits oft gar nicht klar ist, was das Beste wäre, da hierfür das Ziel unzureichend bestimmt ist.

Eine üblicher Versuch, mit solchen Schwierigkeit fertig zu werden, besteht darin, sich an Faustregeln zu orientieren, bzw. in Formale Systeme eine *Heuristik*, bzw. heuristische Regeln einzubringen (wie etwa: „Tausche nie eine Königin gegen einen Bauern ab“ oder „Wechsle nicht ohne Grund unvermittelt das Thema“), die wiederum durch ein geschickt ausgedachtes Zusammenwirken ‚stupider‘ Algorithmen formalisierbar ist. Mit heuristischen Regeln können auch viele offensichtlich sinnlose oder kontraproduktive (aber gleichwohl formal mögliche oder ‚erlaubte‘) Alternativen von vornherein ausgeschlossen werden. Allerdings sind der Möglichkeit, heuristische Verfahren nun wiederum selbst zu ‚algorithmisieren‘ schon dadurch Grenzen gesetzt, dass Faustregeln zwar häufig, aber nicht generell, also nicht in *allen* möglichen Situationen, zur Wahl der jeweils (gemäß der Zielvorgabe) besten Alternative führen. Im Hinblick auf die Kompetenz zur Bewältigung eines komplexen Spektrums von möglichen (neuen) *Situationen*, mit denen wir es selbst in begrenzten Teilen unserer Handlungspraxis zu tun haben können, scheint die Suche nach einem *Set* von eindeutigen, vollständig aufzählbaren Regeln nicht besonders erfolgversprechend. Vielmehr kommt hier ein *nicht-thematisches*, bzw. *implizites* und zum Teil *diffuses Hintergrundwissen* ins Spiel, für das es keine klare Grenze zu geben scheint¹³⁹. Im Sinne welcher Regeln könnte dieses Hintergrundwissen berücksichtigt werden? Entsprechende Vorschläge in KI-Forschungsprogrammen laufen meist auf ein hierarchisch organisiertes, an sogenannten ‚Stereotypen‘ orientiertes enzyklopädisches

¹³⁸ Dies über einen bestimmten Zeitraum fertigzubringen, ohne dass der Unterschied zwischen Mensch und Maschine feststellbar wäre, entspricht dem berühmten ‚Turing-Test‘.

¹³⁹ Zu solchen Schwierigkeiten, s. bereits H. Dreyfus: *Die Grenzen künstlicher Intelligenz. Was Computer nicht können* (1972/1985) u. J. Weizenbaum: *Die Macht der Computer und die Ohnmacht der Vernunft* (1976/1978). Die Thematisierung eines nicht-thematischen, aber sinnkonstitutiven Hintergrundwissens geht allerdings mindestens bis auf Husserl zurück. In diesen Zusammenhang gehören auch viele der Überlegungen Wittgensteins (insbesondere in *Über Gewißheit*).

System hinaus, welches mit begrifflichen Querverweisen arbeitet, um so aus der Unmenge möglicher Hintergrundfakten das für die jeweilige (z.B. kommunikative) Situationsbewältigung relevante Wissen zu selektieren. Allerdings sind auch solche Ansätze mit grundlegenden Schwierigkeiten behaftet, nicht zuletzt wegen einer damit in Verbindung stehenden weiteren ‚harten Nuss‘ für KI-Tüftler, nämlich dem sogenannten ‚*Rahmenproblem*‘, welches vor allem darin besteht, die möglichen situationsverändernden Effekte abzuschätzen, die dann auftreten, wenn ein AFS, gemäß seinem menschlichen Vorbild, seine jeweilige Umgebung manipulieren können soll¹⁴⁰.

Die vieldiskutierten Probleme, mit denen Versuche einer Algorithmisierung heuristischer Orientierungshilfen, einer relevanzorientierten Berücksichtigung des Hintergrundwissens, der Antizipation situationsverändernder Effekte und anderer Aspekte unseres Verständigungswissens zu kämpfen haben, weisen möglicherweise auf *grundlegende* Schwierigkeiten hin, sprachliche und nichtsprachliche Verstehens- und Handlungskompetenz *formalistisch* (bzw. digital) zu simulieren¹⁴¹. Diese Frage möchte ich aber an dieser Stelle noch nicht vertiefen. Wir sollten zunächst erst einmal näher auf den im vorigen Abschnitt vollzogenen Übergang von Formalen Systemen *ohne externen Sinn* (wie dem Schach) zu Symbolsystemen mit externem Sinn, wie einer natürlichen Sprache, eingehen. Komplexe, syntaktisch strukturierte Symbolsysteme, wie natürliche Sprachen, haben ja einen materialen und empirischen Sinn, das heißt, ihr begrifflicher Gehalt ist *prima facie* nicht ‚rein formal‘, bzw. rein intern bestimmt (wie etwa, nach dem hier dargelegten Verständnis, der Sinn der Figuren im Schach lediglich durch interne syntaktische Transformationsregeln bestimmbar ist).

Damit *künstlich erschaffene* Zeichensysteme Weltbezug haben oder ‚Welt repräsentieren‘, bedürfen sie, gemäß der Tradition der Formalen Semantik, einer extensionalen semantischen ‚*Interpretation*‘. Darunter wird dort meist die eindeutige Zuordnung materialer oder empirisch gehaltvoller, wahrheitsrelevanter Eigenschaften zu den basalen Zeichentypen (‚Wörter‘) eines syntaktisch strukturierten Systems verstanden, wodurch erst eine wahrheitsbezogene *Beurteilung* syntaktisch wohlgeformter Anordnungen von Zeichentoken (‚Sätze‘) möglich wird. Eine solche semantische Interpretation lege – so drückt man es häufig aus – fest, wie sich ein syntaktisch strukturiertes System ‚zur Welt verhält‘. Wichtig ist dabei, dass zwar die Zuordnung von Bedeutung zu bestimmten einfachen Symboltypen willkürlich (konventionell) ist, die Bedeutung komplexer Symbole jedoch durch ihre Komposition systematisch determiniert ist.

¹⁴⁰ Für eine konstruktiv orientierte, ausführliche Darstellung und kritische Diskussion solcher Probleme, s. Haugeland (1985).

¹⁴¹ Vergl T. Winograd/ F. Flores: *Understanding Computers and Cognition* (1987).

Im Gegensatz zu entsprechend *gemachten* künstlichen Sprachen sind natürliche Sprachen historisch gewachsen. Für Versuche einer syntaktischen und semantischen Präzisierung natürlicher Sprachen im Sinne extensional interpretierter Formaler Systeme besteht eine altbekannte Schwierigkeit zunächst darin, dass sich die Bedeutung vieler ihrer Ausdrücke nicht ohne weiteres auf wahrheitsrelevante Eigenschaften reduzieren lässt. Eine Gleichsetzung des sprachlichen Sinns mit *extensionalen* Eigenschaften wie Bezug oder Wahrheitsbedingungen ist ja gerade in sogenannten ‚intensionalen Kontexten‘, in denen Ausdrücke wie „bedeutet“, „glaubt“ u. ä. eine Rolle spielen, aber etwa auch in indexikalischen Kontexten, problematisch (von den pragmatischen Dimension sprachlicher Verständigung mal ganz abgesehen). Und wenn, wie Vertreter des IBS-Programms behaupten, Sprache jeweils Ausdruck von, den Miteilungsabsichten von Sprechern zugrundeliegenden, Gedanken sein soll, dann gelten diese Schwierigkeiten natürlich auch für entsprechende extensional-repräsentationale Verständnisse des ‚subjektiven Geistes‘.

Nun lassen sich Theorien, die sich in ihren Naturalisierungsbemühungen (in der ein oder anderen Weise) am CMG orientieren, grob in drei Varianten unterscheiden: Gemäß der ersten Variante ist Geist *nichts anderes als ein syntaktisches bzw. Automatisches Formales System*¹⁴²; nach der zweiten ist Geist *ein intrinsisch semantisches System, das irgendwie durch ein syntaktisches System bzw. ein AFS physisch realisiert wird*¹⁴³; und nach der dritten ist Geist *ein interpretiertes syntaktisches System bzw. AFS*¹⁴⁴. Den hier gemeinten Standardfall des CMG bildet die zweite Variante. Die erste Variante kommt offenbar einem Eliminativen Materialismus sehr nahe und soll uns dementsprechend auch nicht gesondert beschäftigen. Mit der spezifisch interpretativen Komponente der dritten Variante werden wir uns noch auseinandersetzen. Jedenfalls haben alle drei Positionen ihre je eigenen Schwierigkeiten das CMG für eine Erklärung von Geist^{PE} fruchtbar zu machen, bzw., krude gesagt, Semantik auf der Grundlage von Syntax zu erklären.

Bevor wir uns jedoch auf die Schwierigkeiten stürzen, sollten wir versuchen, die Grundidee des CMG und die Suggestivkraft, die es auf viele Theoretiker ausübt, besser zu verstehen. Ich meine die Suggestivkraft, die in der Vorstellung liegt, dass jene ‚Eigenschaften‘, die wir zu meinen scheinen, wenn wir Personen ‚intentionale Zustände‘ oder ‚Ereignisse‘ zuschreiben, auf der Grundlage eines automatischen Systems, welches am Ende nach rein formalen Regeln arbeitet, verständlich gemacht werden können. So wie bei (interpretierten) Formalen Systemen wie einem

¹⁴² So etwa bei: S. Stich: *From Folk Psychology to Cognitive Science – The Case Against Belief* (1983).

¹⁴³ Etwa bei: Fodor (1994).

¹⁴⁴ Dennett (1987).

Axiomensystem aus einfachen, als wahr vorausgesetzten Sätzen (Axiomen) immer wieder wahre Sätze folgen, wenn man sich strikt an die formalen (,syntaktischen‘) Regeln hält, soll auch die ,Erdgeschoss-Architektur des Geistes‘ dem Formalistenmotto genügen: „Achte Du auf die Syntax, dann wird die Semantik für sich selber sorgen“. Formale Systeme aber lassen sich, wie das Beispiel des Schachcomputers zeigt, mechanisieren, indem man nicht nur die (algorithmischen) formalen Erlaubnisregeln automatisiert, sondern zusätzlich eine Heuristik, die so etwas wie ,Klugheit‘ ins Spiel bringt, aus einem Zusammenwirken einfacher, mechanisierbarer Algorithmen aufbaut. Geist_{pe} wäre dann, ebenso wie ein Computer, ein System, welches zusätzlich über weltbezogene, *semantisch interpretierte* ,Axiome‘, bzw. Zeichen verfügt.

Könnte dies nicht vielleicht doch den Weg zu einer möglichen Auflösung des alten Paradoxons einer ,mechanistischen Vernunft‘ aufzeigen, also des Paradoxons eines offenbar irgendwie bestehenden, aber doch zugleich scheinbar unmöglichen Zusammenhangs von Materie und Geist? Der anhaltende Eindruck eines solchen Paradoxons zeigt sich etwa in solchen Fragen, wie „Wie sind intentionale Eigenschaften auf Grund rein physischer Eigenschaften möglich?“ oder „(Wie kann ein) bloß physischer Vorgang die intentionalen Gesetze implementieren, denen ein semantisch kohärenter Geist gehorcht?“¹⁴⁵ Wie etwa sollen es, krude formuliert, die Kausaleigenschaften eines physischen Tokens einer ,mentalenen Repräsentation‘, wie etwa „Fido“, bewerkstelligen können, die semantischen Eigenschaften des sprachlichen Typs „Fido“ zu realisieren, also sich *auf Fido zu beziehen* und, als Bestandteil von Gedanken, material-logische Eigenschaften zu haben?

Die Idee, dass Geist^{pe} anhand eines Computermodells verständlich zu machen sei, wird häufig auch als „Maschinen-Funktionalismus“ bezeichnet und ist wesentlich von Überlegungen des Mathematikers Alan Turing inspiriert worden. Auch Kritiker des nicht-teleologischen Funktionalismus, die, wie etwa Dennett, die nicht daran glauben, dass die semantischen Eigenschaften des Geistes funktionalistisch reformulierbar sind, neigen zu der Ansicht, dass das CMG immerhin einen wichtigen Schritt zu einer naturalistischen Erklärung des Geistes^{pe} liefere, in dem es zeige, dass Systeme, denen wir berechtigterweise semantische Eigenschaften zuschreiben, *in einer Weise funktional organisiert und physisch realisiert* sein können, *die wir verstehen*. Um nun aber etwas genauer zu sehen, wie (semantisch interpretierbare) Formale Systeme *automatisiert* werden können, müssen wir kurz auf das Konzept der ,Turing Maschine‘ eingehen.

Turing hatte nachgewiesen, dass *jedes* Formale System, also auch jedes Computerprogramm, aus jedem beliebigen Stoff konstruiert werden kann, wenn dabei

¹⁴⁵ Fodor (1994a), S. 296 (meine Übersetzung).

nur ein entsprechendes System von wenigen und sehr einfachen Transformationsregeln realisiert wird¹⁴⁶. Vereinfacht dargestellt, lässt sich ein solches (automatisches) System von Transformationsregeln, eine Turing-Maschine, als eine Kontrolleinheit verstehen, die sich jeweils in einem bestimmten inneren Zustand (im Rahmen eines vordefinierten, endlichen Repertoirs von möglichen Zuständen) befindet, und diesen Zustand aufgrund von (rein formal beschreibbaren) Inputs schrittweise (digital) verändert. Der Output hängt dann einerseits vom ursprünglichen inneren Zustand der Kontrolleinheit und andererseits von den Eigenschaften des jeweiligen Input ab. Die inneren Zustände ändern sich in der Regel mit jedem Schritt einer neuen Input-Aufnahme. Die jeweils produzierten Outputs haben wiederum Auswirkungen auf die in der Folge möglichen Inputs. Die Vorgehensweise einer Turing-Maschine wird durch die in der Kontrolleinheit gespeicherte Menge an Instruktionen, dem sogenannten ‚Maschinen Tableau‘, festgelegt. Die physische Realisierung einer solchen Turing-Maschine kann man sich etwa so vorstellen, dass ein maschineller Schreib-Lese-Kopf über ein unbegrenztes Band läuft, welches in Felder unterteilt ist, auf denen ‚Informationen‘ im Sinne von Token wie „1“ oder „ “ (Leerfeld) eingestanz sind, die den Zustand des ‚Kopfes‘ (und damit seine kausalen Effekte auf das Band, den Output) in einer bestimmten Weise – je nach der physischen Beschaffenheit des ‚Kopfes‘ – kausal verändern können. Dabei bestehen die möglichen Operationen der Kontrolleinheit lediglich darin

- den Schreib-Lese-Kopf um ein Feld nach rechts oder links zu verschieben;
- ein Token aus dem vorgegebenen Repertoire von ‚Zeichen‘ (also etwa 1 und Leerfeld) auf das Feld, über dem sich der Kopf befindet, stanzen zu können;
- das Zeichen, das auf diesem Feld steht, löschen zu können;
- stoppen zu können.

Turings Grundthese lautet, dass für *jedes* beliebige deterministische AFS eine formal äquivalente Turing-Maschine existiert. Andererseits können Turing-Maschinen auf sehr unterschiedliche Art realisiert sein. Es kommt nur darauf an, die funktionalen Relationen zwischen den verschiedenen, im Maschinen-Tableau erwähnten Parametern (Typen von ‚Symboltoken‘, Zustände und Bewegungen des ‚Kopfes‘) durch passende Kausalrelationen zwischen den physischen Realisierungen dieser Parameter zu ersetzen. Wie ein Zustand der Kontrolleinheit physisch realisiert ist, ist egal, solange dabei nur diejenige kausale Rolle ermöglicht wird, durch die der Zustand funktional

¹⁴⁶ A. Turing: ‚Computing Machinery and Intelligence‘ (1950). Für eine gute, einführende Darstellung der Eigenschaften von Turing Maschinen, s. J. Kim (1996).

(kausal-relational) definiert ist. Es kann für zwei unterschiedliche physische Systeme eine gleichlautende Verhaltensbeschreibung, bzw. Input/Output-Korrelationen geben, wobei diese Systeme dennoch verschiedene Turing-Maschinen realisieren. Eine funktionale Theorie des Verhaltens eines Systems *S*, bzw. eine sogenannte ‚Maschinen-Beschreibung‘ von *S* (d.h., *S* realisiert eine Turing-Maschine), ist allerdings *immer als relativ auf eine Bestimmung möglicher Inputs und Outputs von S* zu verstehen. Dies gilt es bei der Frage, die uns im Zusammenhang mit *funktionalistischen* Theorien noch beschäftigen wird, im Auge zu behalten, nämlich ob und wie innere System- oder Maschinen-Zustände (oder Ereignisse) als *propositionale* Zustände (oder Ereignisse) – relativ auf Input/Output-Beschreibungen – *funktional definiert* werden könnten.

Um nun die eigentliche Pointe des ‚Maschinen-Funktionalismus‘ zu verstehen, müssen wir uns vergegenwärtigen, dass, nach dem hier dargelegten Verständnis, eine Syntax zwar einerseits ein *formales Regelsystem*, syntaktische Eigenschaften aber andererseits zugleich *physische Eigenschaften* ‚zweiter Stufe‘ („second order physical properties“¹⁴⁷) sein sollen, das heißt, als Eigenschaften, die intrinsisch im Haben von anderen Eigenschaften bestehen. So hat beispielsweise jedes Ding, welches der, nach formalen, bzw. syntaktischen Regeln definierbaren, Figur des Turms im Schachspiel entspricht *typische* physische Eigenschaften, also eine geometrische Form oder eine Farbe (bzw. eine gewisse Reflexion des Lichtes). Symboltoken eines syntaktischen Systems nehmen relativ zueinander bestimmbare, räumliche Positionen ein und treten mit einer bestimmten relativen Häufigkeit auf. Diese *physischen* Primäreigenschaften werden nun aber – so die Idee – in realisierten AFS, bzw. Turing-Maschinen durch eine bestimmte Art der Anordnung oder Strukturierung entsprechender ‚Eigenschaftsträger‘ zu *sekundären* kausalen Eigenschaften von Symboltoken. Die *qua* einer entsprechenden Strukturierung oder physischen Form von Symboltoken einer ‚Syntax‘ gegebenen *sekundären* Kausaleigenschaften entsprechen den funktionalen Rollen, welche Symboltoken im Rahmen eines AFS spielen.

Die *reduktive* Pointe des ‚Maschinen-Funktionalismus‘ läuft nun darauf hinaus, *Syntax* als ein *Bindeglied* zu verstehen, welches die Möglichkeit eines sozusagen ‚kategorienübergreifenden Gleichklanges‘ zwischen ‚intentionalen‘, bzw. semantisch-logischen Eigenschaften von *weltbezogenen* Symbolen mit den kausalen Eigenschaften ihrer physischen Instantiierung verständlich machen soll. Und was für Computer gilt, muss grundsätzlich auch für unser Gehirn gelten. In genau diesem Sinne meint etwa Dennett, dass

¹⁴⁷ J. Fodor: ‚Fodor’s Guide to Mental Representation‘ (1985), S. 22.

„(...) the brain (...) is just a syntactic engine; all it can do is discriminate its inputs by their structural, temporal, and physical features and let its entirely mechanical activities be governed by these ‚syntactic‘ features of its input“ (1981, S. 61).

Treffender, wenn gleich nicht sehr elegant, wäre deshalb wohl die Bezeichnung „*Eine intentionale-ist-eine-syntaktische-Maschine-Ansatz*“¹⁴⁸. Auf die spezifischen Schwierigkeiten eines solchen ambivalenten Verständnisses von „Syntax“ – Regeln der ‚Symbolmanipulation‘ einerseits, Kausaleigenschaft ‚zweiter Stufe‘ andererseits – werde ich gleich noch zurückkommen. Zunächst sollten wir aber ganz grundsätzlich die Frage nach der Plausibilität dieses Ansatzes stellen. Kritiker des Turing-Maschinen Funktionalismus haben betont, dass sich, selbst unter Verwendung noch so raffinierter Rezepte, am Ende aus rein formalen bzw. syntaktischen ‚Zutaten‘ kein ‚intentionaler Kuchen‘ backen lasse. Zu den prominentesten dieser Kritiker zählt sicherlich John Searle, dessen einschlägiges ‚*Chinese-Room-Argument*‘ *beweisen* soll, dass bei einer rein syntaktischen/formalen Beschreibung der internen Organisation von Systemen keine semantischen Eigenschaften herauspringen können¹⁴⁹.

Searle entwickelt ein Gedankenexperiment, bei dem er einen Raum mit einer Eingabe- und einer Ausgabeöffnung beschreibt, in dem jemand sitzt, der die Aufgabe hat, auf Eingaben von Täfelchen mit chinesischen Sätzen mit der Herausgabe von richtigen, bzw. adäquaten Täfelchen so zu reagieren, dass es scheint, als wäre der Raum samt Inhalt eine (chinesisch-)sprachlich kompetente Maschine. Allerdings beherrscht die Person in dem Raum kein Wort des Chinesischen. Sie verfügt lediglich über ein (ziemlich kompliziertes) Handbuch in ihrer Sprache, welches *allein aufgrund ‚syntaktischer‘ Eigenschaften, also der physischen Gestalt* der Zeichen und ihrer *Reihenfolge* auf den hereinkommenden Eingabetäfelchen (‚Fragen‘) angibt, welche Täfelchen (als ‚Antwort‘) herauszugeben sind. Diese Situation lässt sich natürlich entsprechend komplexer ausmalen. Die Analogie zum CMG liegt jedoch auf der Hand: Das Handbuch entspricht dem Computerprogramm (einem instantiierten Formalen System), die Zeichentäfelchen entsprechen der sogenannten ‚Datenbasis‘ und die Urheber des Handbuchs den Programmierern.

Es scheint nun intuitiv naheliegend, der Person in dem so beschriebenen Raum, selbst bei perfekter Ausführung der gestellten Aufgabe, keine echte Sprachkompetenz, bzw. Geist (oder ‚Intentionalität‘) zuzusprechen. Verallgemeinert, läuft Searles Argument darauf hinaus, dass Computer nicht die Möglichkeit von Geist^{PE} erklären können, weil ‚intentionale‘ Zustände einen semantischen Gehalt haben; Syntax für

¹⁴⁸ In diesem Sinne, s. etwa auch: Z. Pylyshyn: *Computation and Cognition: Toward a Foundation for Cognitive Science* (1984).

¹⁴⁹ J. Searle: *Geist, Hirn und Wissenschaft* (1984/1986), Kap. 2.

Semantik nicht ausreicht, Computerprogramme aber vollständig durch ihre formale, bzw. syntaktische Struktur bestimmt sind. Die semantischen Eigenschaften ‚Intentionaler Systeme‘ müssten also durch irgendwelche zusätzlichen Bedingungen zustande kommen¹⁵⁰.

In ganz ähnlicher Weise hat auch Ned Block zu zeigen versucht, dass es grotesk anmutende Konsequenzen hätte, wenn – nach den Voraussetzungen der Computer-Analogie des Geistes – die Realisierungsformen ‚Intentionaler Systeme‘ so zu verstehen sein sollen, dass man am Ende des funktionalen Teils der ‚Reduktionsgeschichte‘ jeweils bei lediglich binär kodierten Mechanismen angelangt¹⁵¹. Blocks Argument läuft darauf hinaus, dass, nach einem solchen Verständnis, eigentlich auch ein angeheuertes Heer von entsprechend zueinander positionierten (und vor moralischen Bedenken gefeiten) Leuten, die dafür entlohnt werden, im Rahmen eines adäquaten Input-Output-Settings auf empfangene Schläge auf den eigenen Kopf mit Schlägen auf den Kopf von bestimmten anderen Teilnehmern eines solchen ‚Klopp-den-Kopp-Systems‘ zu reagieren, *jeden in diesem Sinne* ‚funktional‘ beschriebenen kognitiven Prozess funktional simulieren könnte (Ähnliches ließe sich auch für Versuche einer rein formalen Bestimmung nicht-kognitiver Funktionen, wie etwa der Verdauung, anführen). Für Block ist dies nicht plausibel. Ein solches Heer von ‚Klopp-den-Kopp-ern‘ könnte keine kognitive Funktion erfüllen, weil für funktionale Identität offenbar *mehr erforderlich* sei, *als kausale Strukturisomorphie*. Letztere wäre damit für die Möglichkeit funktionaler Identität – jedenfalls im üblichen Sinne von „Funktion“ – genauso unnötig, wie physische (Typ-) Identität. Blocks Überlegung beruft sich allerdings lediglich auf einen Mangel an intuitiver Überzeugungskraft einer entsprechenden Vorstellung. Ein davon unabhängiges Argument bleibt er jedoch schuldig. Falls aber Geist tatsächlich im Sinne des (Turing-) ‚Maschinen-Funktionalismus‘ erklärbar wäre, warum sollte dann die Funktionsweise des Gehirns nicht prinzipiell durch ein solches System von ‚Klopp den Kopp-ern‘ ersetzt werden können, wenn wir eine gigantische Apparatur hinzudenken, die für die entsprechenden verhaltenssteuernden Verbindungen zu einem riesigen Körper sorgen würde¹⁵²?

¹⁵⁰ Tatsächlich glaubt Searle, dass ‚Intentionalität‘ so etwas wie eine höherstufige Systemeigenschaft des menschlichen Gehirns ist. Das Argument, warum Computer keine Intentionalität haben können, hat bei Searle allerdings später (in: ders.: *The Rediscovery of the Mind* (1992)) insofern eine bemerkenswerte Konkretisierung erfahren, als Searle nun die – für Artefakte nicht zu unterstellenden – Phänomene Bewusstsein und die Empfindsamkeit als Grundlage der Möglichkeit von ‚Intentionalität‘ (im ursprünglichen Sinne) ins Feld führt.

¹⁵¹ Block (1978).

¹⁵² Hierzu A. Beckermann: *Analytische Einführung in die Philosophie des Geistes* (1999), S. 164 f.

Doch auch unabhängig von einer eindeutigen Bewertung jener Form von Einwänden, wie sie von Searle und Block vorgebracht wurden, ist ein nicht unerheblicher Teil des naturalistische Lagers der Ansicht, dass artifiziellen AFS – ebenso wie anderen Artefakten, wie Büchern oder Korkenziehern – ein *Sinn* oder eine *Funktion durch uns* erst *verliehen*, bzw. durch eine bestimmte praktische Verwendung erst gegeben und *nicht* bereits etablierter Sinn nur beschrieben, entschlüsselt oder erklärt wird. So bemerkt etwa Haugeland, dass *alle* Artefakte

„...only have meaning, because we give it to them; their intentionality, like that of smoke signals and writing, is essentially borrowed, hence *derivative*. To put it bluntly: computers themselves don't mean anything by their tokens (anymore than books do) - they only mean what we say they do. Genuine understanding, on the other hand, is intentional 'in its own right' and not derivatively from something else.“ (1981, S. 32 f.)¹⁵³

Ebenso, wie also das physische Substrat, aus dem Grenzsteine, Korkenzieher oder Bücher bestehen, nur dadurch zu etwas ‚intentional Aufgeladenem‘, wie Grenzsteine, Korkenzieher oder Bücher würde, weil es ‚intentionale‘ Wesen gibt, *für die* sie Grenzsteine sind, bzw. die sie *als* Korkenzieher oder Bücher erschaffen (und benutzen), so würden auch diejenigen intrinsisch zunächst rein physikalischen Dinge, die wir auf einer Ebene von elektrischen Schaltkreisen, Displays, Tastaturen usw. beschreiben können, nur dadurch zu Flugüberwachungssystemen, Rechnern und Übersetzungsgeräten, weil *wir* sie zu diesem Zweck konstruieren. Und genauso generierten, so der Einwand, auch die von Computern (im Sinne bestimmter physikalischer Vorgänge) manipulierten Symboltoken ihren Sinn nicht aus ‚eigener Kraft‘. Damit aber wäre ihr Verhalten auch *nicht wirklich* zielgerichtet und intelligent, sondern nur ‚*als ob*‘. Wenn nun der Sinn eines AFS grundsätzlich etwas ihm Äußerliches wäre, dann könnte das Computermodell des Geistes natürlich auch nicht die Frage beantworten, auf welche Weise *ursprünglicher Sinn*, bzw. *Bedeutung* generiert, der dem lediglich abgeleiteten eines AFS zugrunde liegen muss. ‚Originäre‘ oder ‚intrinsische Intentionalität‘, die Menschen und höheren Tieren, nicht aber Artefakten zukäme, müsste sich dann durch andere oder zumindest zusätzliche, *nicht formal reduzierbare* Eigenschaften auszeichnen.

Auch Fodor meint, dass Eigenschaftstheorien, mit denen ‚intentionale Gesetze‘ implementiert werden sollen, semantische Eigenschaften nicht mit rein formalen Eigenschaften korrelieren können, es bedürfe auch einer (externalistischen) Theorie des ‚Informationsgehaltes‘; immerhin sei das CMG aber der einzige Ansatz, der zur

¹⁵³ Diese Position hat Haugland (1985/1987) später allerdings relativiert.

Hoffnung Anlass gibt, erklären zu können, wie Repräsentationen und ihre syntaktische Struktur überhaupt physisch realisiert sein – und damit kausale Wirksamkeit erlangen – können. Diese Überlegung ist nun für Fodor zugleich ein Hinweis auf die sprachanaloge Form ‚mentaler Repräsentationen‘:

„In computer design, causal role is brought into phase with content by exploiting parallelisms between the syntax of a symbol and its semantics. But that idea won't do the theory of *mind* any good unless there are *mental* symbols; mental particulars possessed of semantic *and syntactic* properties. There must be mental symbols because, in a nutshell, only symbols have syntax, and our best available theory of mental processes – indeed, the *only* available theory of mental processes that isn't *known* to be false – needs the picture of the mind as a syntax-driven machine“. (Fodor/1985, S. 23)

Um die These der Analogie des ‚mentalen Repräsentationsapparates‘ zu natürlichen Sprachen hinsichtlich des Verhältnisses von Syntax und Semantik zu betonen, redet Fodor auch ganz unverblümt von einer ‚*Sprache des Geistes*‘ oder einem ‚Mentalesisch‘¹⁵⁴. Der Sinn, bzw. der repräsentationale Gehalt des ‚Mentalesischen‘ müsste dessen *Token* allerdings *intrinsisch* zukommen, also nicht, wie bei ‚äußeren‘ Sprachen, durch eine Form ‚intentionaler Sinngebung‘ (wie etwa im Falle des Griceschen Verständigungsmodells) oder durch eine externes Interpretationsverfahren. Fodor glaubt deshalb, dass der Gehalt *angeboren* sein muss und durch auch *system-externe* Faktoren bestimmt ist¹⁵⁵. AFS, bzw. Computer können somit keine ursprüngliche ‚Intentionalität‘ besitzen.

Damit scheint Fodor von Einwänden im Sinne Searles oder Blocks insofern nicht betroffen, als er nicht behauptet, semantische Eigenschaften seien vollständig durch die internen formalen Eigenschaften eines ‚Intentionalen Systems‘ determiniert. Eine entsprechend erforderliche kausal-informationale ‚Theorie des Gehalts‘ muss dann aber wiederum erklären können, wie sie durch rein *interne* ‚syntaktische‘ Eigenschaften *realisiert* sein kann, wo sie doch einerseits nicht mehr, andererseits aber zugleich doch mehr sein soll, als das! Die ist allerdings eine Frage, die auch die ‚Natur‘ propositionaler Gehalte betrifft, und wir werden später darauf zurückkommen. Immerhin hat die Fodorsche Analogisierung der internen Konstitution ‚mentaler Repräsentationen‘ zur Struktur natürlicher Sprachen den Vorteil, dass sich so die

¹⁵⁴ Zuerst in: J. Fodor: *The Language of Thought* (1975). Zu den Anhängern der Hypothese einer ‚Sprache des Geistes‘ zählen (*inter alia*) auch H. Field (s. ders.: ‚Mental Representation‘ (1978)), W. Lycan (ders.: *Judgement and Justification* (1988)) und McGinn (1982).

¹⁵⁵ Wie wir noch sehen werden, lehnt Fodor funktionalistische (Eigenschafts-) Theorien des Gehaltes ab und setzt mittlerweile eher auf kausalistisch-externalistische Erklärungen (wobei aber skeptisch gegenüber evolutionistischen, bzw. teleologischen Erklärungen ist). S. etwa Fodor: ‚A Theory of Content‘ (I+II) (1990a).

Fähigkeit zu ‚intentionaler *Produktivität*‘ besser verständlich machen ließe¹⁵⁶. Damit ist das Phänomen gemeint, dass wir in der Lage sind, aus einem gegebenen Repertoire möglichen Sinns gegebenenfalls Gedanken, Wünsche usw. zu entwickeln, die nicht einfach als Ganzes erlernt wurden. Dies entspricht dem im Falle der Sprachkompetenz wohlbekannten Phänomen, Sätze äußern zu können, die wir vorher nie gehört haben; wir können ja aus einem endlichen Vorrat an Wörtern und grammatischen Regeln, prinzipiell unendlich viele Sätze bilden. Diese ‚Produktivität‘ sprachlichen Sinns wird üblicherweise durch die *Kompositionalität* der Satzbedeutung erläutert, also (vereinfacht gesagt) dadurch, dass die durch die Satzkomponenten getragenen Bedeutungskomponenten nach (mehr oder minder bestimmten) strukturellen Regeln kreativ variiert werden können. Nach dem gleichen Prinzip könnte man sich nun, nach der Vorstellung der Verfechter einer ‚Sprache des Geistes‘, auch die Fähigkeit zu ‚intentionaler Produktivität‘ vorstellen, also der Fähigkeit zur Generierung neuer Gedanken. Ein damit offensichtlich zusammenhängendes Phänomen nennt Fodor die „Systematizität“ der Sprache (bzw. propositionaler Einstellungen): Wer eine bestimmte Menge von Sätzen einer Sprache versteht, versteht *eo ipso* auch eine bestimmte andere Menge von Sätzen dieser Sprache. Die entsprechenden Eigenschaften ‚intentionaler Prozesse‘ lassen sich nach Ansicht der Verfechter der *Lingua-Mentis*-Hypothese nur erklären, wenn ‚mentale Zustände‘ *intern*, wie Sätze, eine syntaktische Konstitution aufweisen.

Die These, dass ‚intentionale Zustände‘ eine ihren ‚propositionalen Objekten‘ entsprechende *konstitutive Struktur* haben müssen, ist – wie gesagt – umstritten. Für solche Vertreter des ‚intentionalen Realismus‘, für die die interne Konstitution ‚repräsentationaler Zustände‘ keine funktional erklärende Rolle spielt (nennen wir sie „monadische Funktionalisten“), kommt es nur auf die semantische Bewertbarkeit solcher Zustände als Ganzer an¹⁵⁷. Für die unten zu diskutierenden Schwierigkeiten funktionalistischer oder kausalistischer Theorien propositionaler Einstellungen (bzw. Gehalte) ist die Frage des Für und Wieder einer ‚Sprache des Geistes‘ jedoch nicht ausschlaggebend; noch betrifft sie unmittelbar die Frage, ob Artefakten, wie Computern, eine lediglich ‚abgeleitete‘ oder doch eine ‚intrinsische Intentionalität‘ zuzusprechen sei. Worauf es mir hier zunächst ankommt, ist der Versuch Fodors und andere Vertreter des CMG, die Möglichkeit einer mechanistisch-physischen Reduktion des ‚Intentionalen‘ über den Begriff der Syntaktizität plausibel zu machen. Und es scheint mir offensichtlich, dass der eigentliche ‚Trick‘ des ganzen Unternehmens darin besteht, dass der Ausdruck „Syntax“ in einer äquivoken Weise verwendet wird,

¹⁵⁶ Fodor (1985), S. 16 f. u. ders. (1987) S. 147 f.

¹⁵⁷ So etwa für Loar (1981).

nämlich einerseits im Sinne von physischen *Regelmäßigkeiten* ‚zweiter Stufe‘, und andererseits im Sinne von *Konfigurationsregeln*. Die argumentationsstrategische Rolle dieses doppeldeutigen Verständnisses des Syntaxbegriffs ist nämlich gerade dadurch bestimmt, dass er keiner Seite (Regel oder Regelmäßigkeit) *exklusiv* zugeschlagen wird und so, durch eine scheinbar mögliche gleichzeitige Verwendung in beiden Kontexten, eine Brückenfunktion zwischen *prima facie* inkompatiblen begrifflichen Sphären übernehmen soll.

Nun ist es aber mehr als fraglich, ob man Syntax überhaupt als ‚physikalische Eigenschaft zweiter Stufe‘ verständlich machen kann. Denn in welchem Sinne kann eigentlich *ohne* den Rückgriff auf semantische Vorverständnisse von „syntaktischen Eigenschaften von *Symboltoken*“ oder „syntaktischen *Regeln*“ die Rede sein? Wie sind den *Symboltoken* überhaupt als solche individuier- bzw. identifizierbar, wenn nicht unter Rückgriff auf ihren intrinsisch semantischen Charakter? Anders ausgedrückt: Was macht ein *Token* zu einem *Symboltoken*? Ein Symbol verstehen wir ja zunächst als etwas, das gemäß, bzw. aufgrund einer Regel *richtiger Verwendung* etwas (für jemanden) bedeutet. Und es scheint doch so, als ob syntaktische Eigenschaften (im üblichen Sinne) nicht einfach als ‚höherstufige physische Eigenschaften‘ von Dingen beschrieben werden können, weil physische Dinge als solche nichts an sich haben, was sie überhaupt erst zu Symbolen, also einem *möglichen Gegenstand syntaktischer Konfigurationsregeln*, macht. Dies zeigt sich schon daran, dass es offenbar keine rein physische Bestimmung der *Gleichheit* von Symbolen gibt¹⁵⁸. Ob zwei Token als Instantiierungen *desselben* Symbols gelten können, hat zwar – jedenfalls im Rahmen einer bestimmten Sprache – auch etwas mit deren physischer Ähnlichkeit zu tun. Nichtsdestotrotz legen physische Eigenschaften aber *per se* keine entsprechenden Unterscheidungskriterien fest, da die Begriffe der Regel und der Gleichheit eine *nicht vorab determinierte Offenheit* ihrer Anwendung, bzw. der *Beurteilung* ihrer Instantiierungen, angesichts jeweils neuer Situationen beinhaltet. In diesem Sinne ist auch der Begriff der syntaktischen Regel auf einen Begriff der Gleichheit der Instantiierungen von *Symboltypen* angewiesen, der sich durch physische Identitätskriterien nicht einholen lässt. Regeln oder Paradigmen, bzw. die Orientierung an ihnen, sind ‚grammatisch‘ (im Sinne Wittgensteins) an die (freie Wahl-) *Möglichkeit des Nicht-Befolgens* durch kompetente Adressaten gekoppelt – eine Möglichkeit, die kausalen Mechanismen grundsätzlich verwehrt ist¹⁵⁹. Damit aber scheint sich die ganze

¹⁵⁸ Hierzu: Searle (1990).

¹⁵⁹ Ein Naturalist würde diesem Einwand womöglich entgegenhalten, dass die Pointe des ganzen Unternehmens am Ende doch gerade darin bestehe, eine ganze Kategorienfamilie, zu der, neben dem Begriff des Symbols, auch die mit diesem ‚verwobenen‘ Begriffe der Regel und der Gleichheit gehören, zu naturalisieren – etwa dadurch, dass man sie durch Begriffe wie

Idee einer Vermittlung von kausalen und semantischen Eigenschaften durch ‚syntaktische‘ Eigenschaften als Augenwischerei herauszustellen: Denn wenn man den Regelbegriff nicht loswird, hat man einen ‚normativen Virus‘ im terminologischen Gepäck, das alle anderen Bereiche anzustecken droht.

Ich denke, dass Anhänger des CMG, neben den oben bereits angeschnittenen Zweifeln an einer ‚Quasi-Algorithmisierbarkeit‘ von heuristischen- oder Vernunftregeln, ganz allgemein mit der Frage konfrontieren sind, wie jene so unterschiedlichen Eigenschaftskategorien, wie die semantisch-‚intentionale‘ einerseits und die, als *physikalische* Eigenschaft zweiter Stufe verstandene, ‚syntaktisch-formale‘ andererseits, im Sinne einer ‚Ist-eigentlich-nicht-mehr-als-Beziehung‘ verstanden werden können soll. Dies erscheint doch schon deshalb unplausibel, weil der Regelbegriff – und mit ihm das gesamte ‚intentionale‘ Vokabular – intern auf den Begriff eines *Subjektes* verweist: Symboltoken bedeuten bzw. vermitteln einen Sinn doch nicht *an und für sich!* Die ganze, von Naturalisten und KI-Forschern praktizierte, Rede von „Repräsentationen“, „Wahrnehmungen“, „Bedeutungen“, „regelgeleiteter Symbolmanipulation“ usw. ist von der Vorstellung eines Subjektes abhängig, für das etwas repräsentiert oder bedeutet, das etwas wahrnimmt, nach Regeln ‚manipuliert‘, bzw. diesen folgt. Da nun aber die gesetzesförmig beschreibbare *Regelmäßigkeit* physischer Ereignisse, die eine naturalistische Reduktion von ‚Intentionalität‘ am Ende liefern soll, mit der Regelmäßigkeit der Orientierung von Handlungssubjekten offensichtlich begrifflich inkompatibel sind, stellt sich die Frage, wie man denn, im Rahmen einer naturalistischen Reduktion ‚intentionaler‘ Redeweise, das intentionale Subjekt loswerden will?

„Regelmäßigkeit“ und „hinreichende physische *Ähnlichkeit*“ substituiert. Obwohl es also durchaus unterschiedliche physische Realisierungen von semantischen Eigenschaften gäbe, müssten die physischen Träger solcher Eigenschaften, zumindest innerhalb einer als funktionale Einheit individuierten Realisierungsform von ‚Intentionalität‘, bestimmten physischen Ähnlichkeits- bzw. Unterscheidbarkeitsbedingungen genügen, nämlich immerhin soweit, dass sie den *funktional* ‚geforderten‘ *gleichen*, bzw. unterschiedlichen *kausalen Rollen* gerecht werden könnten. Was dabei als „hinreichend ähnlich“ durchgeht, könnte man zum Beispiel als durch die selektiven Mechanismen evolutionären Naturgeschehens ‚ausgewählt‘ verstehen. ‚Fehler‘, also – ‚naturalistisch übersetzt‘ – zu starke Abweichungen in bestimmten Kausaleigenschaften, wären dann als dysfunktional im Sinne der Systemfunktion definierbar. Und der Naturalist könnte hinzufügen, dass damit der Begriff „Symbol“, zumindest im genannten Kontext, eine neue, entanthromorphisierende Bedeutung erhalten hätte – ein Vorgang, der im Rahmen wissenschaftlichen Erkenntnisfortschritts ja nicht unüblich sei.

Mag sein. Doch wir müssten dann einerseits immer noch unterstellen, dass zumindest der Funktionsbegriff als nicht-normativer verständlich gemacht werden kann – was zweifelhaft erscheint. Und andererseits hätten wir damit auch nicht so etwas, wie eine naturalistische Reformulierung *unserer Urteilspraxis* erreicht, denn diese können wir ja nicht gleichsam ‚von außen‘ betrachten, ohne deren *normatives Selbstverständnis* zu verlieren.

Um insbesondere mit dieser Schwierigkeit fertig zu werden, haben Fodor, Dennett und andere¹⁶⁰ vorgeschlagen, die *Top-Down*-Strategie einer reduktiv-funktionalen Erklärung des Geistes im Sinne einer ‚analytischen Verdummungsstrategie‘ aufzufassen. Die ‚Homunkuli-Theorie der Intentionalität‘ versteht den Geist (*mind*) als kognitives System hierarchisch strukturierter ‚Kooperationsgemeinschaften‘ von funktionalen Subsystemen mit nach unten hin schrittweise abnehmenden kognitiven Fähigkeiten – vom übergeordneten, ‚vollwertigen‘ Intentionalen System bis hin zu Einheiten mit einem Unterscheidungspotential von einem Bit. Damit wird unmittelbar an jene, an der Beschreibung von Computersystemen orientierten, Redeweise einer *Hierarchie* von (Informations-) Flussdiagrammen (*Flow-Charts*) angeknüpft. Die Homunkuli-Strategie versucht eine Art von ‚Sukzessiv-Entzauberung‘ mentaler Phänomene im Sinne einer schrittweisen kognitiven Depotenzierung jeweils funktionensebenenspezifisch beschriebener ‚Subjekte der Intentionalität‘, bei der die Möglichkeit eines mehr oder minder *kontinuierlichen Übergangs* von einer ‚intentionalen‘ zu einer ‚syntaktischen‘ und schließlich zu einer rein physischen Beschreibung (‚erster Stufe‘) plausibel gemacht werden soll. Dennett formuliert diese Idee in der für ihn typisch solopen, aber anschaulichen Weise:

„If we can get a team or committee of *relatively* ignorant, narrow minded, blind homuncoli to produce intelligent behaviour of the whole, this is a progress. A flow chart is typically the organizational chart of a committee of homuncoli (investigators, librarians, accountants, executives); each box specifies a homunculus by prescribing a function *without saying how it is to be accomplished* (one says, in effect,; put a little man in there to do the job). If we than look closer at the individual boxes we see that the function of each is accomplished by subdividing it via another flow chart into still smaller, more stupid homuncoli. Eventually this nesting of boxes within boxes lands you with homuncoli so stupid (all they have to do is to say yes or no when asked) that they can be, as one says, ‚replaced by a machine‘. One *discharges* fancy homuncoli from one’s scheme by organizing armies of such idiots to do the work.“ (Dennett (1978), S. 123 ff.)

Nach dem homunkulistischen Computermodell des Geistes beschreibt das *Master Flow-Chart*, auf der obersten Ebene, die wechselseitigen Einflussnahmen der systemintern auf Teilaufgaben spezialisierten, ‚intentionalen‘ Subsysteme oder Abteilungen. Diese Subsysteme und Abteilungen entsprechen *Black Boxes*, deren wechselseitiger *Informationstransfer*¹⁶¹ durch Pfeildiagramme, die mit entsprechenden

¹⁶⁰ Fodor (1968); Dennett (1978); W. Lycan (1988).

¹⁶¹ In diesem Begriff stecken einige Schwierigkeiten. Zur Erörterung eines naturalistischen, an der Shannon-Weaver-Theorie orientierten Informationsbegriffs, siehe: F. Dretske: *Knowledge and the Flow of Information* (1981). Auf Dretskes Position und auf die Schwierigkeiten mit

Kommentaren zu versehen sind, aufgezeigt werden können. Auf dieser Analyseebene sollen die *Black Boxes* etwa als ‚Ausführungs-Einheit‘, ‚Wahrnehmungsanalyse-Abteilung‘, ‚Sprachzentrum‘ oder ‚Zwischengedächtnis‘ bezeichnet werden. Mit jeder tieferliegenden Analyseebene müssten der ‚Gehalt‘ des ‚Informationstransfers‘ zwischen zunehmend spezialisierteren Subsystemen dann entsprechend primitiver, das heißt, als unterscheidungs-unspezifischer verstanden werden (also etwa im Sinne eines diskriminatorisch-informativen Gefälles, wie zwischen „läuft“ und „bewegt sich“). Schritt für Schritt sollen so die jeweiligen analyseebenenrelativen *Black-Boxes* ihrerseits funktional erklärt werden, indem man sie durch ein Flussdiagramm beschreibt, welches funktionsspezifische Sub-Subsysteme und ihr Zusammenwirken darstellt. Auf einer Ebene unterhalb des *Master Flo- Charts* wäre dann beispielsweise, wie in der Informatik üblich, von „Rezeptoren“, „Wegen“, „Speichern“, „Auslösern“, „Filtern“ oder „Dämpfmechanismen“ die Rede¹⁶². Eine solche ‚Verdummungsstrategie‘ gelangt schließlich auf eine Ebene der Beschreibung, auf der Operationen im Sinne einfachster Regeln durchgeführt werden, wie sie etwa durch die obige Beschreibung der Realisierung einer Turing-Maschine exemplifiziert sind. Mit dieser Vorgehensweise ist es, nach Ansicht der *Homunculi*-Theoretiker, schließlich möglich, den ‚vermeintlich‘ dichotomischen Gegensatz zwischen Geist und Physis, zwischen der Möglichkeit (subjektbezogener) Bedeutungs- Zweck- oder Sinnhaftigkeit einerseits und (subjektlosem) mechanistischem Geschehen andererseits, als einen bloß *graduellen* plausibel zu machen¹⁶³.

Ähnlich wie hinsichtlich des ambivalenten Verständnisses des Syntaxbegriffs muss aber auch hier die Frage aufgeworfen werden, ob nicht die ganze *homunculi*-theoretische Rede von Funktionen überhaupt nur solange den Anschluss an ‚intentionale‘ – oder auf Intentionalität (im Sinne zweckorientiert handelnder Subjekte) bezogene – Redeweisen als gewährleistet plausibel machen kann, als sie den Schritt zu einer Reduktion auf eine rein physische, bzw. mechanistische Redeweise nicht wirklich vollzogen hat, d. h. auf eine Redeweise, die Funktionen nur mehr über kausale Rollen, etwa im Sinne der Realisierung einer Turing-Maschine bestimmt. Entweder reden wir (noch) von Subjekten – und seien sie auch noch so dumm – oder wir beschreiben einen Vorgang in einer Weise (wobei die Rede von „dumm“ als eine lediglich metaphorische

einem als „naturalistisch“ verständlich zu machenden Informationsbegriff werde ich noch zurückkommen.

¹⁶² Lycan (1988), Kap. 1.

¹⁶³ Allerdings mag man diese Hoffnung als insofern eher programmatisch anzusehen, als das reale Paradigma dieser Überlegungen – die modernen Computer – bisher offensichtlich weit davon entfernt sind, menschlich-kognitive Allerweltskompetenzen, wie etwa alltagssprachliche Verständigungsfähigkeit, auch nur annähernd zu realisieren. Siehe etwa: T. Winograd/F. Flores (1987).

bewusst sein sollte), dass von Subjekten eben vernünftigerweise keine Rede mehr sein kann. Im *Homunkuli* Modell des Geistes scheinen am Ende sowohl der in Anschlag gebrachte Begriff von Syntax, als auch der der Funktion genau dort, wo diese Begriffe die eigentliche Bruchstelle zwischen physischen Eigenschaften oder kausalen Geschehnissen einerseits und Regeln, Sinn oder Zwecken – und damit Subjektbezogenheit – überbrücken sollen, zwischen beiden Seiten zu oszillieren. Auch unabhängig von der Frage, ob es sinnvolle kausalistische Verständnisse von „Syntax“ oder „Funktion“ geben kann, sollte aber klar sein, dass man die entsprechenden Begriffe entweder in dem einen oder dem anderen Sinne verstehen kann, jedoch nicht irgendwie in beiden zugleich.

Zur Frage einer angemessenen Definition des Funktionsbegriffs und zum Stellenwert teleologischer Erklärungen

Bevor wir ausführlicher auf die Schwierigkeiten eingehen, die sich aus der oben konstatierten äquivoken Verwendungsweise des Syntaxbegriffs und aus dem Homunkuli-Modell des Geistes ergeben, müssen wir zunächst die Frage behandeln, ob es für naturalistische Theorien auf der Grundlage des CMG nicht möglich – und daher ratsam – wäre, von vorneherein ganz auf die intentionale Terminologie von ‚Subjekten‘, ‚Regeln‘, ‚Zwecken‘ u. ä. m. zu verzichten. Dies müsste offenbar damit beginnen, sich erst gar nicht auf ein teleologisches Verständnis des Funktionsbegriffs einzulassen, einen Funktionsbegriff also, der sich auf eine genuine Form teleologischer Erklärung stützt.

Wir hatten ja Funktion oder Funktionalität oben vorläufig als so etwas, wie Zweckdienlichkeit, funktionale Identität von Zuständen oder Vorgängen als Austauschbarkeit hinsichtlich der Zweckdienlichkeit, bzw. – bei Subsystemen – hinsichtlich des Beitrags zur Aufrechterhaltung der Funktion eines übergeordneten Systems eingeführt. Häufig verwenden wir den Funktionsbegriff und teleologische Erklärungen in Sinne von Homöostase: Schwitzen etwa hat eine *kompensatorische* Funktion, nämlich auch im Falle einer durch Anstrengung hervorgerufenen Steigerung der ‚Kalorien-Verbrennung‘ oder bei erhöhter Außentemperatur die Temperatur des Körpers (innerhalb einer bestimmten Schwankungsbreite) konstant zu halten¹⁶⁴. Erklärungen in Begriffen kompensatorischer Funktionen liefern eine genuin teleologische, am Ergebnis orientierte Erklärung eines Ereignisses (also etwa des

¹⁶⁴ A. Collins: ‚Action, Causality and Teleology‘ (1984).

einsetzenden Schwitzens), keine ursächliche. Dementsprechend erzeugen beide Formen der Erklärung auch nicht dieselbe Art von Verstehen.

Wie wir gesehen hatten, vermeidet das Modell einer Turing-Maschine, anders als der ‚Homunculi-Funktionalismus‘, teleologische Konnotationen und redet lediglich von den *tatsächlichen* kausalen Rollen von Symboltoken. Ein gewisser Zusammenhang mit teleologischen Verständnissen von Funktionen scheint dabei nur noch durch die gemeinsame Annahme der Möglichkeit unterschiedlicher Weisen der Realisierung von als funktional-identisch definierten Zuständen gegeben. Welche dieser explanativ unterschiedlichen, grundsätzlichen Bestimmungen des Funktionsbegriffs nun hinreichend oder angemessen ist, das ist – insbesondere auch im Hinblick auf eine naturalistische Theorie des Geistes (bzw. des Verhaltens) – nach wie vor strittig.

So erhebt auch Cummins‘ einflussreicher Versuch, den Funktionsbegriff ohne Rückgriff auf teleologische Erklärungsmuster zu analysieren, den Anspruch, gängigen wissenschaftlichen Verwendungen dieses Begriffs gerecht zu werden. Cummins versteht funktionale Erklärungen ganz im Sinne der oben angeführten Bestimmung Schiffers, also als Zuordnung kausaler Rollen in Kontexten von Black-Box-Problemen. Im Rahmen funktionaler Erklärungen Funktionen zuzuschreiben soll demnach bedeuten, einem Subsystem eine Rolle oder einen Beitrag (im Sinn einer Disposition) zur Erklärung der Fähigkeiten (*capacities*) eines übergeordneten Systems *qua* Erfüllung einer kausalen Rolle zuzuordnen¹⁶⁵. Entsprechende ‚funktionalen Analysen‘ sollen komplexere ‚Fähigkeiten‘ (bzw. kausalen Dispositionen) eines Systems durch einfachere ‚Fähigkeiten‘ (bzw. kausale Dispositionen) von Systembestandteilen im Sinne von Eigenschaftstheorien erklären.

Vertreter eines solchen dispositional-kausalistischen Verständnisses funktionaler Erklärungen halten der These der Nicht-Ursprünglichkeit der ‚Intentionalität‘ von Computern entgegen, dass die Unterscheidung zwischen ‚originärer‘ und ‚abgeleiteter Intentionalität‘ auf der Ebene funktionaler Erklärung keine Entsprechung finde. Entscheidend für das Vorliegen von ‚Intentionalität‘ sei, dass die internen Zustände des Systems bestimmte kausale Rollen einnehmen – unabhängig davon, wie diese physisch realisiert sind. Wenn ein bestimmter innerer Zustand einer Maschine die gleiche kausale Rolle im Hinblick auf die Verursachung eines bestimmten Verhaltens spielt, wie ein bestimmter innerer Zustand beim Menschen, gäbe es demnach keinen Grund, dabei zwischen ‚abgeleiteter‘ und ‚ursprünglicher Intentionalität‘ zu unterscheiden.

Gegen die Idee, den Funktionsbegriff dispositional, also über faktische kausale Rollen von Zuständen definieren zu wollen, sind auch innerhalb des naturalistischen Lagers verschiedene Einwände erhoben worden. Ein gängiges Argument läuft darauf

¹⁶⁵ R. Cummins: ‚Functional Analysis‘ (1975). S. auch: ders. (1983).

hinaus, dass mit einer solchen Art der Analyse Funktionen nicht von zufälligen oder ‚bloßen‘ Wirkungen unterschieden werden könnten. Beispielsweise erzeugt ein Otto-Motor normalerweise eine bestimmte Art von Geräuschen. Besteht nun die Funktion des Vergasers in dessen kausalem Beitrag zur Hervorbringung dieser Geräusche? Gehört es zur Funktion des Herzens, Geräusche zu produzieren? – Wohl kaum.

Man könnte diesem Einwand vielleicht entgegenhalten, er sei insofern nicht ganz fair, als beispielsweise Cummins seine Analyse des Funktionsbegriffs ja nur im Hinblick auf angemessene *wissenschaftliche* Verwendungen versteht, die sich jeweils durch den *explanativen Wert* einer entsprechenden funktionalen Erklärung rechtfertigen und relativ auf ein bestimmtes Erkenntnisinteresse sind¹⁶⁶. Doch offenbar unterliegen Verwendungen des Funktionsbegriffs auch unter dieser Einschränkung idealisierenden, also *normativen* Bedingungen, insofern funktional zu erklärende Systeme doch durch bestimmte Inputs, Outputs und Hintergrundbedingungen spezifiziert werden müssen. Otto-Motoren laufen beispielsweise nur bei einem bestimmten Luftdruck, ihr Input besteht aus einer bestimmten Menge von Sauerstoff und Benzin, ihr Output üblicherweise in Energie, die auf eine Antriebswelle übertragen wird. Außerdem kann, abhängig vom Erklärungsinteresse, ein und dasselbe ‚Etwas‘ gleichzeitig unterschiedliche Systeme im Sinne Cummins exemplifizieren – je nach dem, was als spezifische Outputs und als mögliche Inputs oder Hintergrundbedingungen festgelegt werden. Vielleicht möchte ein ‚Cummins-Funktionalist‘ auch dies als unproblematisch ansehen. Eine solche ‚instrumentalistische‘ Analyse des Funktionsbegriffs würde allerdings nicht mehr dem von der überwiegende Mehrheit der Vertreter des Computermodells des Geistes zu Grunde gelegten ‚realistischen‘ Verständnis funktionaler Rollen entsprechen.

Die entscheidende Schwierigkeit dispositionaler Analysen, auf die insbesondere Millikan hingewiesen hat, besteht meines Erachtens darin, zumindest für eine wichtige Klasse von wissenschaftlichen Verwendungen des Funktionsbegriffs nicht zwischen der *eigentlichen* Funktion und einer faktischen Abweichung oder Fehlfunktion unterscheiden zu können¹⁶⁷. Offenbar ist der Begriff der Funktion, im Gegensatz zu

¹⁶⁶ Dennoch können mögliche Defizienzen nicht-teleologischer Analysen von Funktionen, wie ich meine, auch durch einen Vergleich mit sinnvoll möglichen alltäglichen Verwendungsweisen dieses Begriffs aufgezeigt werden. Dies ist insofern eine legitime Strategie, als die Verwendung des Funktionsbegriffs innerhalb naturalistischer Theorien des Geistes ja nicht völlig den Anschluss an jene außerwissenschaftlichen Praktiken verlieren darf, in denen der Funktionsbegriff paradigmatischerweise zum Tragen kommt. Denn sonst wäre es besser, gleich einen neuen *terminus technicus* einzuführen. Dass dies nicht getan wird, hat doch wohl seinen Grund darin, dass eine Form ‚semantischer‘ Nähe zum vorwissenschaftlichen Verständnis gewollt ist.

¹⁶⁷ Millikans einschlägiger Terminus für ein natürlich-teleologisches Verständnis von Funktionen ist „*proper function*“ (dies.: *Language, Thought and Other Biological Categories*

dem der Disposition oder der Kausalität, ein wesentlich *normativer*, weil wir Dingen oder Vorgängen eine *Funktion auch dann* zuschreiben können, wenn sie diese gar *nicht realisieren*. Beispielsweise besteht die Funktion einer Waschmaschine darin, Wäsche zu waschen, auch wenn einige, wegen bestimmter Fehler, stattdessen Keller überfluten. Selbst wenn – durch einen riesigen Zufall – plötzlich alle Waschmaschinen Keller überfluteten, anstatt die Wäsche zu reinigen, würden wir sicherlich die eigentliche Funktion von Waschmaschinen nicht neu bestimmen¹⁶⁸. Und sogar wenn ein Ding *nie* in der Lage ist, eine bestimmte Funktion auszuführen, können wir sagen, worin seine *eigentliche* Funktion besteht. Dies aber ist im Rahmen einer dispositionalen Analyse offensichtlich unmöglich. Hinzu kommt, dass dispositionale bzw. an kausalen Rollen orientierte Ansätze, viele interessante Generalisierungen außer Acht lassen müssen. Man denke nur an die Biologie, wo doch häufig von speziesübergreifender Funktionsidentität die Rede, etwa im Hinblick auf die gemeinsame Funktion von faktisch sehr unterschiedlichen Realisierungen bestimmter Sinnesorgane oder Verhaltensweisen¹⁶⁹.

Eine moderne Alternative zu Funktionsanalysen im Sinne kausaler Rollen oder Dispositionen bilden ätiologische bzw. ‚historische‘ Ansätze. Eine einflussreiche frühe Version einer solchen Analyse stammt von Larry Wright. Wrights Vorschlag für eine Definition von Funktionen lautet wie folgt:

- „The function of *X* is *Z* means
 (a) *X* is there because it does *Z*
 (b) *Z* is a consequence (or result) of *X*’s being there“¹⁷⁰

Diese Analyse besticht zwar durch ihre Einfachheit und Allgemeinheit, scheint aber überinklusiv zu sein, wie ein bekanntes Beispiel verdeutlichen soll. Stellen wir uns einen Kieselstein vor, der einen größeren Stein in einem reißenden Bach stützt. Ohne dieses Stützen würde der Kiesel fortgespült. Müssten wir nach Wrights Analyse nicht sagen, dass es die Funktion des Kieselsteins ist, den größeren Brocken zu stützen, da es diese Eigenschaft des Kiesels ist, die erklärt, ‚warum er da ist‘?¹⁷¹ Mir scheint die *Prima-facie*-Plausibilität dieses Gegenbeispiels allerdings auf einer unzulässigen Verschiebung des Sinnes der Wendung ‚is there‘ zu beruhen. Denn ich lese Wrights

(1984) und ‚In Defense of Proper Funktions‘ (1989a); ähnlich: K. Neander: ‚The Teleological Notion of ‚Function‘‘ (1991a). Millikan beruft sich insbesondere auf tatsächliche Verständnisse funktionaler Erklärungen in der Biologie.

¹⁶⁸ Diese Beispiel stammt von Millikan (in: dies.: ‚Biofunctions: Two Paradigms‘ (1999)).

¹⁶⁹ Zu diesem Argument, s. Neander (1991).

¹⁷⁰ L. Wright: ‚Functions‘ (1973), S. 161.

¹⁷¹ Vergl. C. Boorse: ‚Wright on Functions‘ (1976).

Analyse im Sinne einer Existenzfeststellung, und nicht im Sinne einer (deiktischen) Ortsangabe. Die *Existenz* des Kieselstein ist aber, auch im genannten Beispiel, sicherlich nicht dadurch verursacht, dass er den größeren Stein abstützt.

Mit der Berufung auf biologische Reproduktions- und natürliche Selektionsmechanismen hat die ätiologische Analyse des Funktionsbegriffs in der Folge, vor allem durch die Arbeiten Millikans und Neanders, einen stark vom Neodarwinismus geprägten Charakter erhalten¹⁷². Millikan und Neander zufolge ermöglicht der ‚selektionistische‘ Ansatz eine *naturalistische Definition* eines normativen, für teleologische Erklärungsformen relevanten, Begriffs von Funktion (bzw. ‚eigentlicher Funktion‘) geben:

„Having a proper function is a matter of having been ‚designed to‘ or of being ‚supposed to‘ (impersonal) perform a certain function. The task of the theory of proper functions is to define this sense of ‚designed to‘ or ‚supposed to‘ in naturalist, nonnormative, and nonmysterious terms.“ (Millikan 1984, S. 17)

In den aus teleofunktionalistischer Perspektive grundlegenden (wenngleich nicht auf solche beschränkten) *biologischen* Verwendungen des Funktionsbegriff steht dabei vor allem das evolutionstheoretische Konzept der (gegenüber Konkurrenten) *besseren Anpassung* von Organismen an bestehende Umweltbedingungen aufgrund bestimmter Eigenschaften, bzw. Fähigkeiten und eines dadurch gegebenen *Vorteils* bei der biologischen *Reproduktion* im Zentrum. Bestimmte Dinge oder Eigenschaften sind demnach dadurch funktional, dass sie ‚zu einem bestimmten Zweck‘ oder ‚auf ein bestimmte Ziel hin‘ einen Beitrag leisten (also, im gegebenen biologischen Fall, fürs Überleben und damit auch für die Möglichkeit der Reproduktion der Gene) und deshalb durch den *Prozess* einer *Auslese* der *Geeignetsten* – *qua* des faktischen Überlebens und der Fortpflanzung unter gegebenen Bedingungen – ‚ausgewählt‘ werden. Ein so verstandener Funktionsbegriff und das entsprechende teleologische bzw. intentionale Vokabular soll schließlich anhand einer neo-darwinistischen Erklärung durch Mechanismen willkürlicher Mutation und Selektion naturalistisch, also im Rahmen kausaler Prozesse oder Mechanismen, analysiert werden können. So definiert etwa Neander den biologischen Funktionsbegriff folgendermaßen:

„It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and

¹⁷² Millikan (1984), (1989) und (1999); K. Neander: ‚Functions as Selected Effects: The Conceptual Analyst's Defense‘ (1991b); vergl. auch: E. Sober: *The Nature of Selection* (1984).

which caused the genotype, of which X is the phenotypic expression, to be selected by natural selection.“¹⁷³

Mit dem Versuch der Anwendung eines evolutionistisch definierten Funktionsbegriffs auf ‚intentionale‘ Phänomene, wie propositionale Einstellungen oder sprachliche Bedeutung, werden wir uns insbesondere im Rahmen der ausführlichen Auseinandersetzung mit der Position Millikans noch im Einzelnen beschäftigen. Die Details sind für unsere hier angestellten allgemeinen Überlegungen zum Funktionsbegriff noch entbehrlich. Jedenfalls können Theoretiker, die der Ansicht sind, dass für eine naturalistische Erklärung propositionaler Einstellungen (bzw. Gehalte) ein *teleologischer* Funktionsbegriff im hier skizzierten Sinne geboten ist, auf zweierlei Weise mit dem Einwand der Nichtursprünglichkeit von menschlich *erzeugter* ‚Intelligenz‘ oder ‚Intentionalität‘ umgehen. Sie können entweder der These *zustimmen* und darauf verweisen, dass zu ‚originärer Intentionalität‘ eben eine *bestimmte kausale Geschichte* gehört; – eine teleologische Erklärung des Verhaltens oder der Eigenschaften ‚originär intentionaler Systeme‘ müsste also in einer kausal-genealogischen Begrifflichkeit, unter Annahme bestimmter selektiver Mechanismen, erfolgen¹⁷⁴. Oder sie können, wie Dennett¹⁷⁵, diese These zurückweisen und dafür argumentieren, dass *alle* Formen von Syntaktischen Maschinen, deren ‚intentionales Profil‘ ein gewisses Komplexitätsniveau erreicht – also auch Artefakte – nur unter Berücksichtigung ihres Verhaltens im Rahmen gegebener Umweltbedingungen und Selektionsmechanismen *angemessen interpretiert* werden können. In dieser Hinsicht bestünde kein Unterschied zwischen Artefakten und Gehirnen. Auch letztere seien am Ende nichts anderes als Syntaktische Maschinen, wenngleich ungeheuer komplexe. Auf der Ebene der *Software*-Beschreibung entsprächen damit auch *natürliche* Intentionale Systeme interpretierten Automatischen Formalen Systemen, die sich mit bestimmten Umweltbedingungen ‚herumschlagen‘ müssen und dabei aus ‚ihren Fehlern lernen können‘. Der Unterschied zwischen artifizieller und natürlich entstandener ‚Intentionalität‘ wäre jedoch, wie Dennett meint, keiner von *ursprünglich* und *abgeleitet*, da, wenn man überhaupt so reden wolle, auch menschliche ‚Intentionalität‘ als von der von ‚Mutter Natur‘ abgeleiteten verstanden werden müsse¹⁷⁶.

¹⁷³ K. Neander (1991), S. 174.

¹⁷⁴ C. McGinn: *Mental Content* (1989), S. 154 f.

¹⁷⁵ D. Dennett: ‚Evolution, Error, and Intentionality‘ (1987a)

¹⁷⁶ Diese These wird allerdings nur im Zusammenhang mit Dennetts ‚Anti-Realismus‘, bzw. ‚Instrumentalismus‘ wirklich verständlich. Das werden wir aber erst unten, im Kontext der Erörterung von Dennetts Variante des Interpretationismus, in angemessener Weise diskutieren können.

Auch gegenüber teleologisch-evolutionistischen Analysen des Funktionsbegriffs wurde – anhand der üblichen Methode des Beibringens scheinbar falsifizierender Beispiele – der Einwand erhoben, dass diese sich nicht mit der gängigen Verwendung des Funktionsbegriffs decke, da es Funktionen gebe, deren Existenz nicht durch selektive Mechanismen erklärbar sei¹⁷⁷. Dagegen hat sich auch Millikan damit verteidigt, dass sie keine begriffliche Analyse, sondern eine technische Definition eines Begriffs von Funktion anstrebe, die sich an einer entsprechenden wissenschaftlichen Erklärungspraxis orientiere¹⁷⁸. Wenn allerdings diese Zurückweisung des Anspruchs auf eine *allgemeine* begriffliche Analyse des Funktionsbegriffs im Falle Millikans recht sein kann, ist sie im Falle Cummins nur billig. Man kann die jeweils andere Position jedenfalls nicht mehr mit beliebigen Gegenbeispielen im Sinne abweichender Verwendungen des Funktionsbegriffs kontern.

Millikans Theorie ist eine funktionalistische, denn sie ist ‚Realistin‘ und beansprucht eine naturalistische Erklärung propositionaler Einstellungen, bzw. Gehalte als solcher. Darauf werden wir zurückkommen. An dieser Stelle soll es zunächst um die Frage gehen, ob im Rahmen neo-darwinistischer Theoriebildung tatsächlich eine vollständige Naturalisierung eines, als explanativ berechtigt unterstellten, teleologischen Vokabulars gelingen kann. Denn zunächst einmal hat man ja mit der Inanspruchnahme eines teleologischen Funktionsbegriffs, wie es scheint, eine nicht-mechanistische, antromorphe, bzw. intentionale Redeweise akzeptiert, was nach Ansicht mancher Theoretiker nicht mit dem modernen naturwissenschaftlichen Bild der Natur und Formen möglicher Erklärung in Einklang zu bringen sei. Die übliche Kritik teleologischer Verständnisse funktionaler Erklärungen stößt sich vor allem daran, dass solche Erklärungen erstens die Möglichkeit der Verursachung durch Zukünftiges unterstellten und deshalb mit gängigen mechanistischen Erklärungen inkompatibel seien, und dass zweitens wegen der in ihnen in Anschlag gebrachten Optimierungsannahmen, das Prinzip empirischer Falsifizierbarkeit nicht gewährleistet sei. Ich halte beide Einwände allerdings für falsch: Teleologische Erklärungen sind überhaupt *keine* Kausalerklärungen und daher auch keine solchen, die zukünftige Ursachen zulassen müssten. Das heißt nicht, das, etwa im Falle homöostatischer

¹⁷⁷ Darüber hinaus ist auch strittig, ob natürliche Selektion nicht nur die Häufigkeit erklären kann, mit der eine Eigenschaft in einer bestimmten Population auftritt, sondern, wie Millikan und Neander meinen, auch deren jeweilige Instantiierung durch Individuen (siehe hierzu etwa: D. M. Walsh: ‚The Scope of Selection: Sober an Neander on what Natural Selection Explains‘ (1998). Offensichtlich gibt es im Zusammenhang mit evolutionstheoretischen Erklärungen noch einigen Klärungsbedarf.

¹⁷⁸ Millikan räumte ein, dass es sinnvolle Anwendungen des cumminsschen Funktionsbegriffs sogar in der Biologie geben kann, wenn auch nicht in dem Umfang, wie von Anhängern einer entsprechenden Analyse behauptet (s. Millikan (1999)).

Prozesse, die zugrundeliegenden Mechanismen irgendwie mysteriös wären. – Ganz im Gegenteil: Wenn wir unter teleologischen Erklärungen von Ereignissen *ergebnisorientierte* Erklärungen verstehen, können wir einen entsprechenden *Prozess*, wie etwa die durch eine Umweltveränderung (z.B. Ansteigen der Temperatur) hervorgerufene *kompensatorische* Aktivität (Schwitzen) zur Erreichung eines bestimmten Ergebnisses (Aufrechterhaltung einer konstanten Körpertemperatur) vollständig *kausal* beschreiben – und zwar in der üblichen zeitlichen Reihenfolge von Verursachungsbeziehungen¹⁷⁹. Dennoch hätten wir es nicht mit einer kausalen Erklärung zu tun. Die Antwort, die eine teleologische Erklärung auf die Frage nach dem ‚Warum‘ eines Ereignisses gibt, hat einen gegenüber einer kausalen Erklärung *eigenständigen explanativen Sinn*. Und solche Erklärungen sind, obgleich ein normativer Funktionsbegriff ins Spiel kommt, auch empirisch fallibel. Denn wenn wir in Fällen der Abweichung von behaupteten homöostatischen Zusammenhängen eine Dysfunktionalität unterstellen, muss diese in eigenständiger Weise (empirisch fallibel) erklärbar sein.

Ob bestimmte Bereiche unseres Naturverständnisses einer teleologischen Erklärung und diese wiederum einer intentionalen Begrifflichkeit bedürfen oder nicht, ist Gegenstand einer Debatte, die natürlich nicht erst seit dem Aufkommen des Funktionalismus geführt wird. Sie hat aber in der Auseinandersetzung um den *explanativen Charakter evolutionistischer Erklärungsmodelle*, besonders seit *Darwin*, zusätzlich an Brisanz gewonnen. So liefert nach einem häufig vertretenen Standpunkt gerade Darwins „(...) influential theory of natural selection (...) purely causal, nonteleological descriptions of and explanations for phenomena previously supposed to be the products of design and purpose“¹⁸⁰. In dieser Einschätzung wird zunächst einfach vorausgesetzt, dass teleologische Erklärungen auf eine intentionale Begrifflichkeit zurückgreifen müssen. Dies ist aber, gemäß dem oben bereits erwähnten Vorschlag von Collins, teleologische Erklärungen kompensatorischer Ereignisse, wie etwa das Schwitzen, als ergebnisorientiert („*outcome orientated*“) zu begreifen – und damit die Rede von Zielen oder Zwecken zu vermeiden –, alles andere als selbstverständlich. Damit ist allerdings noch nicht gesagt, dass auch evolutionistische Erklärungen teleologische von der *Art* sind, die ohne eine intentionale Begrifflichkeit auskommen könnten. Angesichts der Darwinschen Terminologie von (natürlicher) „Auswahl“ (bzw. „Auslese“), „Anpassung“ und einem „*Survival of the fittest*“ drängt sich *prima facie* sogar eher der Verdacht auf, dass dies gerade nicht der Fall ist.

¹⁷⁹ Collins (1984). Collins ist zudem der Ansicht, dass, erstens, teleologische Erklärungen nicht auf eine intentionale Begrifflichkeit von Zielen oder Zwecken angewiesen sind, und dass, zweitens, Handlungserklärungen als teleologische zu verstehen seien.

¹⁸⁰ T. Beauchamp/A. Rosenberg: *Hume and the Problem of Causation* (1981); S. 317.

„Auswahl“ (ob natürlich oder nicht) ist fraglos ein intentional ‚aufgeladener‘ Begriff, dessen explanative Pointe es ohne Verlust kausalistisch zu kompensieren gälte. Ähnliches gilt für „Anpassung“. Und wenn man dementsprechend versuchte, dass zumindest teleologische Moment, welches im Begriff „*fittest*“ (bzw. bestangepassten oder geeignetsten¹⁸¹) steckt, kausalistisch aufzulösen und durch eine Identifikation mit dem Faktum des Überlebens zu tilgen, scheint ein Zirkularitätsproblem zu entstehen, denn dann würde man die Tatsache, dass eine Spezies oder ein Organismus existiert, durch die Tatsache erklären wollen, dass er die Eigenschaft hat, überlebt zu haben. Allerdings scheinen mir Zirkularitätseinwände, verstanden als *generelle* Kritik an neo-darwinistischen Erklärungsmodellen¹⁸² deren explanativen Komplexitätsniveau ebenso wenig gerecht zu werden, wie der damit zusammenhängende Einwand der Nicht-Falsifizierbarkeit (der ja bereits von Popper erhoben wurde). Evolutionistische Erklärungen können ja im Einzelfalle empirisch durchaus überprüfbare Angaben über den *statistischen* (also relativen) Überlebens- oder Fortpflanzungserfolg von bestimmten Anpassungsleistungen von Organismen, Populationen oder Arten *in eng spezifizierten biologischen Nischen* machen¹⁸³.

Dies einzuräumen, ändert jedoch nichts daran, dass Schlüsselbegriffe darwinistischer Terminologie eine gleichsam ‚inner-natürliche‘, aber *nachträgliche Urteilsbildung über das Erreichen eines möglichen Ziels* unterstellen. Und dieses Faktum wird auch nicht durch eine Kenntnis molekulargenetischer Mechanismen biologischer Reproduktion aus der Welt geschafft. Denn wie auch immer diese aussehen mögen, sie sind für eine evolutionistische Erklärung, die uns verständlich machen will, *warum* ein Organismus, eine Population oder eine Art mit dieser oder jener genetischen Beschaffenheit unter bestimmten Umweltbedingungen – relativ zu anderen – erfolgreich war (oder nicht) oder warum eine Art bestimmte (teils außergewöhnliche, vielleicht zunächst als eher hinderlich erscheinende) Eigenschaften oder Verhaltensweisen aufzuweisen hat, nur dann relevant, wenn sie in irgendeinen Zusammenhang mit einer ‚Vorteilhaft-für-Terminologie‘ gebracht werden können. In diesem Sinne argumentiert Keil:

„Darwins Metaphern sind (...) durch keine noch so verfeinerte Erklärung der angeblichen ‚Mechanismen‘ der Evolution eliminierbar, denn die Erforschung

¹⁸¹ Womit natürlich die Frage aufkommt: „*Besser* angepasst (bzw. geeignet) im Hinblick auf *was* (oder gar: nach wessen Maßgabe)?“

¹⁸² Vergl. hierzu: S. Gould/R. Lewontin: ‚The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptionist Programme‘ (1978).

¹⁸³ Inwieweit solche Prognosen tatsächlich überprüfbar wären, hängt auch von der biologischen Reproduktionsgeschwindigkeit der untersuchten Organismen ab. Diese kann z. B. bei Bakterien bekanntlich eine sehr hohe sein.

kausaler Prozesse hat mit der interpretativen Leistung der teleologischen *Beurteilung* der Prozesse als *Anpassungsleistungen* gar nichts zu tun“. (Keil 1993, S. 118)

Keil rückt mit seiner Rede von der „interpretativen Leistung“ evolutionistische Erklärungen scheinbar in die Nähe hermeneutischer Verstehensmethoden. So weit müssen wir nicht gehen, denn wir haben es immer noch mit natürlichen Prozessen, nicht mit verantwortbaren Subjekten zu tun. Immerhin scheint aber einleuchtend, dass, wollte man gänzlich ohne intentionale Terminologie auskommen, man damit auch auf eine wichtige explanative ‚Pointe‘ des evolutionistischen Modells verzichtete. So sieht es auch Dennett, der allerdings, genau wie Keil, in die Gefahr gerät, die kategoriale Differenz zwischen dem Erklären eines natürlichen Geschehens und dem Verstehen von und Interagieren mit verantwortbaren Subjekten zu verwischen, indem er die Berücksichtigung der teleologische Komponente der Evolutionstheorie der Einnahme eines ‚*Intentional Stance*‘ gegenüber Personen gleichsetzt. Auch über die biologische Evolution *müssen* wir, nach Dennett, aus Gründen explanativer Nützlichkeit so reden, *als ob* „Mutter Natur“ Eigenschaften ‚bewertet‘ und so gleichsam ‚die Fäden zieht‘:

„Certainly we can describe all processes of natural selection without appeal to such intentional language, but at enormous cost of cumbersomeness, lack of generality, and unwanted detail. We would miss the pattern that was there, the pattern that permits prediction and supports counterfactuals“. (Dennett 1987b, S. 316 ff)

Ich denke, Dennetts Bemerkung untertreibt sogar noch immer die Bedeutung intentionaler, bzw. teleologischer Redeweisen: Selbst wenn wir eine vollständige *physikalische* Beschreibung der Welt hätten, gewännen wir, auch bezüglich des Bereichs des Geschehnishaften, durch (wahre) biologische, medizinische, soziologische und andere teleofunktionale, bzw. intentionale Beschreibungen, eine ganze Menge an qualitativ andersartigem möglichen Wissen über die (bzw. Verständnis der) Welt hinzu. Und da eine gewaltige Menge unseres Alltagswissen nicht aus einer bloßen Kenntnis physikalischer Gesetzmäßigkeiten bestehen dürfte, würde jemand mit einem rein physikalischen Wissen auch kaum all zu weit kommen. So wäre beispielsweise kein Arzt aufgrund eines noch so detaillierten, aber rein physikalischen Wissens über die Abläufe in meinem Körper in der Lage, mir im Krankheitsfalle zu helfen. Denn ein Begriff biologisch *normaler Funktionsweise*, anhand dessen er eine entsprechende *Dysfunktionalität* (Krankheit) als solche überhaupt *identifizieren* könnte, stünde ihm ja *nicht zur Verfügung*.

In Übereinstimmung mit Dennett – und im Gegensatz zu Keil – verstehe ich das Argument einer explanativen Nicht-Substituierbarkeit allerdings nicht als ein *per se* anti-naturalistisches. Wir beschreiben hier, wenngleich in einer teleologischen oder intentionalen Terminologie, Vorgänge, die vollständig im Rahmen kausaler Zusammenhänge verstanden werden *können*. Zu sagen, dass es keine explanative Substitution teleologischer und keine begriffliche Reduktion intentionaler Beschreibungen durch kausale gibt, heißt nicht, dass entsprechend typidentische Ereignisse nicht *sowohl* in kausalen, *als auch* in teleologischen Erklärungen eine Rolle spielen können. Was ich hingegen im Hinblick auf *propositionale Einstellungen* behaupte, ist, dass sie *überhaupt keine natürlichen Zustände oder Ereignisse* sind.

Bevor ich dafür in den folgenden Kapiteln ausführlich argumentieren werde, will ich zunächst nochmal die oben schon angedeutete Ansicht bekräftigen, dass dies in ähnlicher Weise auch für *Subjekte* (im Sinne entscheidungsfähiger Akteure – paradigmatischerweise *Personen*) gilt, da diese zumindest *mehr* sind, als Zustände, Ereignisse oder Dinge in Raum und Zeit. Im *Homunkuli*-Modell einer sukzessiven Depotenzierung der ‚Subjekte‘ wird eine stufenweise Reduktion auf mechanistische Vorgänge letztlich eher vorgegaukelt, als tatsächlich eingelöst. Die Rede von schrittweise dümmer werdenden *Homunculi* kann nämlich nicht darüber hinwegtäuschen, dass man dabei *entweder* unterstellen muss, man hätte es jeweils noch immer mit *Akteuren* zu tun, die *Entscheidungen* treffen und daher berechtigterweise zum ‚Gegenstand‘ intentionaler und ‚intentionaler‘ Zuschreibungen werden können. *Oder* man nimmt diese Redeweise nicht wirklich ernst und will, terminologisch verklausuliert, *eigentlich* über mechanische Vorgänge sprechen. – Wie man es auch dreht und wendet, ein intermediärer Zwischenschritt zwischen der begrifflichen Sphäre der *Verantwortbarkeit* (von vernünftigen Subjekten) und der des *Geschehnishaften* (physischer Mechanismen) steht im Rahmen unserer begrifflichen Möglichkeiten offensichtlich nicht zur Verfügung. Wir müssen, selbst beim ‚dümmsten‘ Binär-Modul, *entweder* von Ja-Nein-*Entscheidungsinstanzen*, bzw. in irgendeiner Art (und wenn auch noch so schwach gemeinter) kognitiv-intentionaler Begrifflichkeit reden *oder* aber von den *mechanistisch-kausalen Eigenschaften* der zugrundeliegenden Physis. Insofern ein ‚blinder‘ Mechanismus, wie er etwa auf der Ebene von Schaltkreisen einer Computerhardware beschrieben wird, nicht wirklich zwischen (wenn auch nur zwei) Alternativen *wählen* kann, ist er auch kein Subjekt von Intentionalität bzw. ‚Intentionalität‘. Metaphorische Beschreibungen, bei denen so getan wird, als hätte man es *quasi* mit ‚kleinen Leutchen‘ zu tun, die uns prinzipiell ähneln und nur in kognitiver Hinsicht einen beschränkteren Horizont haben, mögen als Form des Redens über die *Funktionsweise von Computern* (oder auch in anderen Kontexten) mitunter erhellend

sein. Warum solche anthromorphe Metaphorik überhaupt, also nicht nur im Hinblick auf Artefakte, sondern für bestimmte Ebenen unseres Naturverständnisses generell, verständniserweiternd oder gar unverzichtbar zu sein scheint, ist ein Thema, das sicher eine eigenständige Untersuchung wert wäre. Wie dem auch sei, entsprechende Redeweisen sind jedenfalls klar *diesseits* ‚intentional‘-intentionaler Begrifflichkeit, wenn auch nur als eine gegenüber einem *Primärgebrauch*, der sich paradigmatischerweise an (sprach- und handlungskompetenten) Personen orientiert, *derivative* – und *praktisch* nicht ernstgenommene – Form¹⁸⁴. Dies heißt, wie gesagt, nicht, dass es in der Welt nicht unklare, nicht-eindeutige Fälle gibt. Unsere grundlegenden, paradigmatischen Verständnisse von Bedeutung und Geist orientieren sich aber sicherlich nicht an solchen Grenzfällen, wie Kleinkindern, Menschen in psychischen Ausnahmesituationen oder Schwachsinnigen.

Die These, dass jegliche Form der Zuschreibung propositionaler Einstellungen als derivativ zu derjenigen Zuschreibungspraxis verstanden werden muss, deren *prototypische* Adressaten *sprachkompetente* Personen sind, soll insbesondere durch die spätere Erörterung der kognitiven Struktur der Propositionalität untermauert werden. Ich werde dafür argumentieren, dass der kognitive Gehalt von Zuschreibungen propositionaler Einstellungen im primären, also uneingeschränkten Sinn, nur auf der Grundlage einer bestimmten Form (sozialer) kommunikativer Praxis verständlich gemacht werden kann, deren kognitives Niveau wesentlich durch die Funktion der Propositionalstruktur bestimmt ist. Darüber hinaus muss diese Praxis, wie ich denke, eine *originär lebensweltliche* sein, dass heißt, ihre Teilnehmer müssen fühlende, endliche Wesen mit Bedürfnissen, Neigungen und Leidensfähigkeit sein. Ersteres schließt offenbar Tiere aus; zumindest letzteres, Maschinen.

Im Folgenden werde ich mich aber zunächst mit funktionalistischen Theorien auseinandersetzen, die versuchen, propositionale Einstellungen durch inferentielle Beziehungen eindeutig zu definieren und diese mit kausalen Beziehungen identifizieren wollen.

¹⁸⁴So können wir unseren Computer natürlich anschnauzen, wenn er nicht so funktioniert, wie wir uns das vorstellen oder ihn sogar aus dem Fenster werfen. In solchen Fällen (unreflektierter Wutausbrüche) tun wir so, als könnten wir ihn oder bestimmte seiner Programmelemente in irgendeinem Sinne verantwortlich machen. Hinsichtlich eines natürlichen Geschehens – und diese Einsicht kehrt meist auf kurz oder lang wieder ein – ist jegliche Schuldzuweisung jedoch völlig fehl am Platze. Und die Entscheidung für die eine oder die andere Redeweise ist mit einem grundlegenden *Perspektivenwechsel* verbunden und läßt sich nicht in einen graduellen, allmählichen Übergang zwischen, nur in ihrem Extrem inkompatiblen, Beschreibungen auflösen.

4 Nicht-teleologischer Funktionalismus und kausale Rollen

Wie wir sahen, geht es bei den *reduktiven* Ambitionen der am CMG orientierten Varianten des modernen Materialismus zunächst darum, zu zeigen, wie ‚intentionale Gesetze‘ durch solche implementiert werden könnten, die auf einer computationalen Ebene funktionaler Erklärung angesiedelt sind. Der eigentliche reduktive Schritt, also eine physisch-mikrostrukturelle Implementierung von ‚formal-kalkülmäßig‘ beschriebenen funktionalen Rollen, soll dann dadurch möglich sein, dass diese Beschreibungen als ‚rein syntaktisch‘ verstanden werden, wobei mit ‚syntaktisch‘ ‚*second order physical properties*‘ gemeint sind. Der entscheidende Gedanke ist, dass die *Kausaleigenschaften* physischer Realisierungen von ‚intentionalen Zuständen‘ oder ‚Ereignissen‘ (insbesondere die äußere physische Form und die raum-zeitliche Anordnung von Symboltoken‘) so beschaffen sind, dass sie deren *semantischen* (d.h. wahrheitsrelevanten) Eigenschaften *entsprechen*.

Als ‚Realisten‘, also Vertreter gehaltbasierter Ansätze, versuchen Reduktionisten, wie gesagt, ‚intentionale Eigenschaften‘ insofern ernst zu nehmen, als sie davon ausgehen, dass Gehalt (bzw. semantischen Eigenschaften) und die Modus propositionaler Einstellungen tatsächlich kausalwirksame physische Mechanismen identifizieren. Ich hatte schon angemerkt, dass nicht alle Anhänger des CMG auch eine gehaltbasierte, reduktionistische Theorie vertreten (wie etwa Dennett). Und umgekehrt sind nicht alle funktionalistischen Reduktionisten, wie insbesondere jene, die in der nachfolgenden Diskussion eine zentrale Rolle spielen werden (wie Loar und Lewis), Anhänger des CMG. Doch egal, ob man als funktionalistischer Reduktionist nun Anhänger des CMG oder ein Freund psycho-neuronaler Typenreduktion oder ist, ein dem Versuch einer computationalen oder neuronalen Implementierung ‚intentionaler Gesetze‘ für ‚realistische‘, also gehaltsbasierte Theorien des Geistes^{PE} notwendigerweise vorgeordneter Schritt muss darin bestehen, propositionale Einstellungen selbst in irgendeinem Sinne kausalistisch zu definieren, um sie überhaupt (kausal-)gesetzes- und damit reduktionstauglich zu machen.

Die im Folgenden zu diskutierenden reduktionistischen Bemühungen können wir als funktionalistische Theorien im *engeren Sinne* verstehen¹⁸⁵. Damit meine ich solche Theorien, nach denen propositionale Einstellungen, bzw. propositionale Gehalte und Einstellungsmodi, mit system- oder subjekt-*internen kausalen Rollen* identifiziert werden sollen. Funktionalistisch *im weiten Sinne* wären demnach all jene Theorien, die die Kausalität propositionaler Einstellungen über *extern* und/oder *historisch-*

¹⁸⁵ Meine Darstellung und Kritik funktionalistischer Theorien profitiert, neben denen Schiffers, vor allem auch von den diesbezüglichen Ausführungen Heckmanns (1994) und Stichs (1983).

teleologisch identifizierte funktionale Eigenschaften von Zuständen oder Ereignissen bestimmen wollen. Mit solchen Theorien werden wir uns dann im Anschluss auseinandersetzen.

Die grundsätzliche Idee, mentale Zustände überhaupt als funktional zu bestimmende Zustände zu verstehen, geht wohl auf Wilfried Sellars und Hilary Putnam zurück¹⁸⁶. Einflussreiche, spätere Beiträge zur theoretischen Konzeption funktionalistischer Theorien propositionaler Einstellungen sind dann vor allem von David Lewis und Brian Loar¹⁸⁷ ausgearbeitet worden. Während sich etwa Lewis streng an den Postulaten der ‚Alltagspsychologie‘ orientieren will, akzeptiert Loar auch deren *partielle* Revisionen. Funktionalistische Theorien müssen also auch unterschieden werden in solche, die sich lediglich auf die ‚alltagspsychologisch‘ gültigen Generalisierungen über ‚intentionale Eigenschaften‘ stützen wollen und in solche, die diese verbessern, präzisieren oder gar partiell eliminieren wollen, falls dies eine adäquate empirische Theorie der Verhaltensklärung, bzw. -Vorhersage erforderte.

Formal gesprochen, lautet die Grundthese des Funktionalismus (im engeren Sinne):

Jede propositionale Einstellung *PE* entspricht einer funktionalen Rolle *F*, so dass eine Einstellung, dass *p* = ein *Token* eines Zustandstyps zu sein, der *F* hat.

Ein Zustandstyp, der *F* hat, ist demnach ein funktionaler Zustandstyp, für den gilt, dass er durch ein funktionales Prädikat der Form

– instantiiert (ggf. für *X* zum Zeitpunkt *t*) die funktionale Rolle einer propositionalen Einstellung *PE*.

individuiert wird. Ob deren Reichweite auf Personen und Zeitpunkte eingeschränkt werden muss, ist, wie wir noch sehen werden, strittig.

Lewis hat diesen Gedanken in der Weise ausgearbeitet, dass er ‚Alltagspsychologie‘ als diejenige Theorie zu versteht, durch die mentalistische Terme, wie „Verspüren eines stechenden Schmerzes im linken Daumen“, „Wahrnehmen eines Krachens“, „Wünschen, dass es bald hell werde“, „Glauben, dass Paul schön ist“ oder „Sich

¹⁸⁶ W. Sellars: ‚Some Reflections on Language Games‘ (1954); Putnam (1960). Allerdings fügen sich weder Sellars und schon gar nicht der ‚neuere‘ Putnam umstandslos in den Rahmen eines naturalistischen Programms.

¹⁸⁷ D. Lewis: ‚An Argument for the Identity Theory‘ (1966); ‚How to define theoretical Terms‘ (1970); ‚Psychophysical and Theoretical Identifications‘ (1972b); sowie: Loar (1981).

entschließen, dass Tempo zu forcieren“ nicht nur extensional bestimmt, sondern *definiert* würden. In diesem Zusammenhang führt Lewis das Konzept einer *impliziten funktionalen Definition* ein, wobei er sich einen, ursprünglich im Kontext der Debatte um die Schwierigkeiten des empiristischen Bedeutungsverifikationismus entwickelten, wissenschaftstheoretischen Ansatz zunutze macht. Dort war man, nach dem offensichtlichen Scheitern des Versuchs, theoretische Terme lediglich durch Beobachtbares definieren zu wollen, dazu übergegangen, die Bedeutung solcher Begriffe als vor allem durch ihre Einbettung in eine empirische Theorie bestimmt zu verstehen. Und bezeichnenderweise erwies sich die Unmöglichkeit einer Übersetzung theoretischer Terme in eine reine Beobachtungsterminologie gerade auch für die Versuche Logischer-Empiristen, wie Carnap und Hempel, propositionale Einstellungen rein behavioral zu *definieren*, als besonders offensichtlich¹⁸⁸.

Lewis' common-sensischer Psycho-Funktionalismus versteht die mentalistischen Terme der ‚Alltagspsychologie‘ dementsprechend als theoretische Terme, die aus Hypothesen über die Existenz ‚mentaler Zustände‘ und ihre kausalen Rollen resultieren. Diese Rollen lassen sich in etwa wie Knoten in einem Netzwerk verstehen, welches auf komplexe Weise bestimmte Inputs mit bestimmten Outputs verknüpft. Implizite funktionale Definitionen sollen nun die zwischen Inputs und Outputs vermittelnden Knotenpunkte durch Aussagen über die für Verhaltensklärungen relevanten mentalen Eigenschaften typindividuierten, ohne das dabei die (physischen) Träger dieser Eigenschaften selbst (referentiell) benannt werden. Funktionalistische Theorien sind, nach Lewis, mit einer Strategie vergleichbar, bei der man, anhand bestimmter bekannter Tatsachen, ein Geschehen rekonstruiert, das sich zwischen unbekanntem Personen abgespielt hat, indem man die betreffenden Personen und ihre Interaktionen mittels Beschreibungen individuiert (also implizit funktional definiert), die sich aus den bekannten Tatsachen ergeben (wobei man zugleich ein allgemeines Wissen über Rationalitätsnormen, Implikationsverhältnisse von Eigenschaften u. ä. m. in Anschlag bringt).

Die funktionale Definition eines intentionalen Terms, wie etwa „Glauben, dass es brennt“, bestünde nach Lewis in einer Konjunktion aller ‚Platitüden‘, die die ‚Alltagspsychologie‘ zu diesem Term liefert. Lewis formuliert diese Überlegung zunächst hinsichtlich aller Arten von ‚mentalen Zuständen‘:

„Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. Add also all the platitudes to

¹⁸⁸ Diese Schwierigkeiten, auf die ich in der Einleitung kurz eingegangen bin, sind innerhalb der Debatte wohlbekannt und als solche auch weitgehend akzeptiert, weswegen ich sie hier nicht mehr ausführen werde.

the effect that on mental state falls under another – ‚toothache is a kind of pain‘, and the like....Include only platitudes which are common knowledge among us – everyone knows them, everyone knows that everyone else knows them, and so on. For the meanings of our words are common knowledge and the names of mental states derive their meaning from these platitudes“ (Lewis 1972, Sekt. 3)

Im Hinblick auf ‚intentionale Zustände‘ sollen sich die entsprechenden Platiniden als typische Verursachungsbeziehungen verstehen lassen. Entsprechende kontrafaktische Generalisierungen der ‚Alltagspsychologie‘ könnten etwa dann folgende Formen annehmen:

(G_{AP1}) Wenn jemand (bzw. etwas) sensorischen Stimulationen des Types S_1 und S_2 ausgesetzt ist, dann führt dies (*ceteris paribus* und mit der und der Wahrscheinlichkeit) dazu, dass er (sie/es) sich daraufhin in mentalen Zuständen des Types M_1 und M_2 befindet.

(G_{AP2}) Wenn sich jemand (bzw. etwas) in mentalen Zuständen des Types M_3 befindet, so führt dies (*ceteris paribus* etc.) dazu, dass er (sie/es) sich auch in Zuständen des Types M_4 befindet

(G_{AP3}) Wenn sich jemand (bzw. etwas) in mentalen Zuständen des Types M_5 und M_6 befindet und sensorischen Stimulationen des Types S_3 und S_4 ausgesetzt ist, dann führt dies (*ceteris paribus* etc.) dazu, dass er (sie/es) sich anschließend in mentalen Zuständen des Typs M_7 und M_8 befindet.

(G_{AP4}) Wenn jemand (bzw. etwas) sensorischen Stimulationen des Typs S_3 ausgesetzt ist sich in einem mentalen Zustand des Typs M_9 befindet, dann führt dies (*ceteris paribus* etc.) dazu, dass er (sie/es) sich daraufhin im mentalen Zustand M_{10} befindet und dadurch eine motorische Reaktion R_1 produziert.

Gegeben, wir hätten die möglichen Input- und Outputparameter (also was als Spezifizierungen von S , bzw. R in Frage kommt) bezüglich möglicher Subjekte solcher Zuschreibungen (in diesem Fall, Personen) bestimmt, dann ließen sich, so der Gedanke von Lewis, die ‚intentionalistischen‘ Terme *qua* einer Ersetzung durch ungebundene Variablen eliminieren, die in eine Konjunktion aller ‚alltagspsychologisch‘-generalisierenden Aussagen der oben dargestellten Form (als dem ‚Postulat der Alltagspsychologie‘) einzusetzen sind. Spezifische Typen ‚intentionaler Zustände‘ oder ‚Ereignisse‘ könnten dann jeweils durch eine Konjunktion all derjenigen alltagspsychologischen Generalisierungen definiert werden, in denen die für deren

spezifische funktionale Rolle stehende ungebundene Variable vorkommt. Solche Definitionen ‚intentionaler Terme‘ lassen die Art ihrer materialistischen Realisierung offen. Wenn nun ‚intentionale Zustands-‘ oder ‚Ereignistypen‘ jeweils genau einer der durch eine solche Konjunktion spezifizierten funktionalen Rollen entsprechen und es von Fall zu Fall möglich wäre, entsprechend formulierte Gesetze schließlich durch physische Mechanismen (etwa neuronaler Art) zu implementieren, wäre so etwas wie eine psycho-physische Reduktion erreicht. Ein funktionalistischer Materialist kann ja die Behauptung der Typenidentität auf eine Entsprechung ‚intentionaler Zustands-‘ oder ‚Ereignistypen‘ mit – in welcher Form auch immer physisch implementierbaren – funktionalen Rollen beschränken¹⁸⁹.

Allerdings müssen wir uns fragen, ob die implizit- (oder besser: intern-) funktional definierte theoretische Terme der so verstandenen Theorie der ‚Alltagspsychologie‘ nicht irgendwie mit nicht-theoretischen Termen verknüpft sein müssten, also solchen, die unabhängig von der Theorie verständlich zu machen sind. Auch wenn dies heutzutage häufig bezweifelt wird, scheint es mir einleuchtend, zu fordern, dass eine Theorie theorieunabhängig beschreibbare Fakten erklären können, bzw. an diesen überprüfbar sein muss; Fakten, die als solche nicht – oder doch nicht zur Gänze – erst durch die Theorie selbst etabliert werden. Außerdem erscheint es mehr als zweifelhaft, ob lediglich *interne* Beziehungen zwischen propositionalen Einstellungen eine hinreichende Beschränkung für ihre Individuierung darstellen können¹⁹⁰. Das bedeutet, dass zumindest einige der ‚alltagspsychologisch‘ durch theoretische Terme bezeichneten Zustands- oder Ereignistypen auch durch ihr kausales Verhältnis zu externen, bzw. nicht-funktional bestimmten Zustandstypen definiert sein müssten. Diese Rolle spielen für Lewis jene Generalisierungen, die, wie in den obigen Beispielen, sensorische Input- und/oder motorische Output-Zustände enthalten, da sensorische *Stimuli* und motorische Reaktionen ja scheinbar rein physisch (neuronal) beschreibbare Zustände (bzw. Ereignisse) darstellen.

Nun ist es aber mehr als fraglich, ob die ‚Alltagspsychologie‘, bzw. unsere Zuschreibungspraxis propositionaler Einstellungen, überhaupt über ein Klassifikationsschema für ‚nicht-intentional‘ bestimmte *Stimuli* bzw. motorische Reaktionen verfügt. Die oben beispielhaft vorgestellten Input-Output-Generalisierungen für eine zulässige Formulierung von Schlüssen nach dem Schema

¹⁸⁹ Lewis geht allerdings davon aus, dass sich über funktionale Definitionen auch eine psychophysische Typenidentität begründen lässt. Diese stärkere Behauptung ist aber für eine Diskussion seiner Variante von Funktionalismus offensichtlich nicht notwendig.

¹⁹⁰ Um nur ein Beispiel von Loar zu zitieren: „Suppose two believes resemble each other except that ‚red‘ occurs in one and ‚orange‘ in the other. It seems most unlikely that there are any purely internal (non-observational) interpersonally ascribable constraints, that distinguish such beliefs.“ (Loar (1981), S. 65).

des praktischen Syllogismus zu halten, ist doch wohl eher einer szientistischen Voreingenommenheit geschuldet: Sind wir in der Alltagspraxis tatsächlich in der Lage, einen gesetzesartigen Zusammenhang zwischen sensorischen Stimulationen, bzw. motorischen Reaktionen einerseits und ‚intentionalen Zuständen‘ andererseits herzustellen?

Um die Möglichkeit psycho-physischer Generalisierungen mit theorieneutralen Inputs und Outputs plausibel zu machen, wählen Funktionalisten meist Empfindungsbeispiele, wie etwa Schmerzarten. Und sicherlich lassen sich die physischen Grundlagen von Empfindungen noch am ehesten neuronal lokalisieren. Wir können es auch dahingestellt sein lassen, ob es, beispielsweise für Schmerzen, tatsächlich entsprechende, *individuen-unspezifische* Input-Output-Generalisierungen geben könnte. Unser Thema sind jedoch ‚intentionale Zustände‘. Und hier drängt sich allemal die Frage auf, wie denn einzelne propositionale Einstellungen, wie beispielsweise die Überzeugung, dass Gerhard Schröder die Bundestagswahl gewonnen hat, mit bestimmten sensorischen Stimulus- oder motorischen Reaktionstypen zusammenhängen sollen (ähnliches gilt selbstverständlich auch für volitive Formen propositionaler Einstellungen).

Dass wir selbst als gewiefte ‚Alltagspsychologen‘ hierauf kaum eine plausible Antwort parat haben werden, könnte Lewis vielleicht mit dem Hinweis kontern, dass es ja nicht darum gehe, *jeden* ‚intentionalen Zustandstyp‘ (und jeden *separat*) theorieextern bestimmen zu können. Viele oder gar die meisten theoretischen Terme sollen ja intern, also durch ihre formal-, material- und praktisch-inferentiellen Beziehungen zu anderen solchen Termen definierbar sein (dies soll, aus funktionalistischer Sicht, ja gerade den wesentlichen Fortschritt gegenüber den Unzulänglichkeiten des *Behaviorismus* ausmachen). Wenn aber zumindest einige der ‚internen Zustände‘ auch über ihr Verhältnis zu rein physisch beschriebenen Input- und/oder Output-Zuständen bestimmbar wären, wäre ein ‚intentionalistischer Zirkel‘ funktionaler Definitionen vermieden.

Aber auch für solche Einstellungen, die man als ‚input-‘ oder ‚outputnah‘ bezeichnen könnte, sieht es nicht besser aus. Wie sollen wir uns die entsprechenden externen Generalisierungen eigentlich konkret vorstellen? Etwa im Sinne von „Druckabhängiger Schmerz im Zentrum der linken Fußsohle führt, *ceteris paribus*, zu der Überzeugung, dass sich ein Steinchen im linken Schuh befindet“; oder: „Wärmeempfindung + intensive Gelb-weiß-Wahrnehmung von der und der Form führt, *ceteris paribus*, zu der Überzeugung, dass die Sonne scheint (oder man sich in unmittelbarer Nähe einer Glühbirne befindet)“; oder wie: „Der Wunsch, den Abzug zu ziehen, führt, *ceteris paribus*, zu einer Muskelkontraktion des Zeigefingers“? – Wohl

kaum: Über solche ‚Generalisierungen‘ verfügt weder ein Alltagsweltler, noch ein Neurologe.

Auf der Outputseite scheinen Klassifikationen von *Körperbewegungen* vielleicht mitunter der Anforderung zu entsprechen, zwar einerseits ‚nicht-intentional‘ aber dennoch möglicher Bestandteil eines Alltagswissens zu sein. Doch es ist eine inzwischen ziemlich unstrittige Tatsache, dass Typen von Körperbewegungen keine *Handlungstypen* individuieren. Beispielweise kann ein und dieselbe Bewegung meines Arms die physische Grundlage sehr unterschiedlicher Handlungstypen bilden, wie etwa, jemanden zu grüßen, ein Taxi heranzuwinken oder eine Lockerungsübung zu absolvieren. Umgekehrt kann ein und derselbe Handlungstyp durch sehr unterschiedliche Bewegungen ausgeführt werden. Selbst wenn man bestimmte basal individuierte Handlungen berücksichtigt, die nur auf der Grundlage *einer* bestimmten Körperbewegung ausgeführt werden können, wie etwa das Krümmen des Zeigefingers, kann dies nicht als Zustandsbeschreibung der gewünschten Form gelten. Denn hinsichtlich einer solchen Körperbewegungen wäre ein krampfartiges Zusammenziehen des Fingers nicht von einem motivierten ‚Handlungs-Output‘ zu unterscheiden. Nur dann aber würde sie überhaupt für die Beurteilung durch Rationalitätsprinzipien nach dem Muster des praktischen Syllogismus in Frage kommen. Da also auch ein kausalistisch gedeutetes Schema des praktischen Schließens über Handlungen und nicht über motorische Outputs generalisieren muss, können letztere nicht Teil der gewünschten (theorieunabhängigen) externen Generalisierungen der ‚Alltagspsychologie‘ sein.

Ähnliches gilt auch für Versuche einer sprachbehavioralen *Output*-Bestimmungen, etwa im Sinne entsprechender Dispositionen zur Produktion von Geräuschfolgen. Denn solche Ereignisse müssten ja mindestens auf der Ebene von Wortfolgen beschrieben werden, um sie mit einem Mindestmaß an Plausibilität unter ‚alltagspsychologische‘ Generalisierungen fallen zu lassen. Doch dann hätten wir, wie auch der Naturalist Fodor einräumt, wiederum nicht die gewünschten, theorieunabhängig beschreibbaren ‚Output-Zustände‘:

„Word is a *psychological* category. (It is indeed *irreducibel* psychological, so far as anybody knows; there are for example, no acoustic properties that all and only tokens of the same word type must share. In fact, surprisingly, there are no acoustic properties that all and only fully intelligible tokens of the same word type must share. Which is why our best technology is currently unable to build a type writer that you can dictate to.)“ (Fodor 1987, S. 8.)

Andererseits lassen sich aber auch Handlungstypen nicht für *externe* Outputgeneralisierungen nutzbar machen. Denn *Handlungen* unterscheiden sich ja,

auch für Funktionalisten, von ‚nicht-intentionalen‘ Vorgängen dadurch, dass sie nur im Zusammenhang mit Überzeugungen, Wünschen und Absichten zugeschrieben werden können. Damit sind Handlungsbeschreibungen aber keine (im Sinne einer ‚Alltagspsychologie‘) theorie-unabhängigen Terme.

Nun kann ein Vertreter des Funktionalismus so vorgehen, dass er den Versuch einer Formulierung externer Generalisierungen auf die Inputseite beschränkt. Diesen Weg beschreitet Loar. Loars partiell revisionistischer Funktionalismus lockert die Bindung einer funktionalistischen Theorie des Geistes an die ‚Alltagspsychologie‘, indem er auch theoretische Begriffe und Generalisierungen zulässt, die nicht Teil ‚alltagspsychologischen‘ Wissens sein müssen. Zukünftige Forschungen über Beschaffenheit des Gehirns und des Nervensystems könnten, nach seiner Ansicht, ‚nicht-intentionale‘ Zustände klassifizieren, welche die für eine verbesserte Verhaltensklärung und -Prognose erforderlichen Generalisierungen liefern.

Die Grundidee auch von Loars Version des Funktionalismus ist, dass es vor allem die inferentielle Rolle propositionaler Einstellungen (und nicht ihr repräsentationaler Charakter) ist, welche deren Modus und Gehalt bestimmen. Propositionale Einstellungen sollen für die Erklärung von Verhalten deswegen Relevanz haben, weil die, in einem weiteren Sinne, logischen Beziehungen zwischen Einstellungstypen, die die ‚Alltagspsychologie‘ unterstellt, den kausalen Beziehungen zwischen den entsprechenden physischen Realisierungen von Einstellungstoken entsprechen. Auch für Loar widerspiegeln also die inferentiellen Beziehungen zwischen Einstellungstypen, zumindest im Großen und Ganzen, die tatsächlichen kausalen Gesetze, unter welche Typen ihrer physischen Realisierung fallen. Die zugrundeliegende Intuition ist, dass es für das Verhalten eines Wesens ja beispielsweise einen Unterschied macht, ob es glaubt, dass sich ein gefährliches Objekt nähert, oder ob es glaubt, dass es harmlos ist. Wenn es zu der Überzeugung kommt, dass das sich nähernde Objekt ein Fressfeind ist, kann es daraus schließen, dass es gefährlich ist und praktische Reaktionsmöglichkeiten daraufhin abwägen, ob sie die größte Wahrscheinlichkeit auf Risikovermeidung bei geringst möglichem Aufwand bedeuten würden. Solche Abwägungen ließen sich dann spieltheoretisch analysieren.

Die Regeln des formalen und des praktischen Schließens (‚Rationalitätsprinzipien‘) und die material-inferentielle Beziehungen (‚Bedeutungsrelationen‘) bilden nach Loar die *internen* Beschränkungen einer Theorie der funktionalen Organisation eines ‚intentionalen‘ Wesens. Die notwendigen *externen* Beschränkungen sollen durch die ‚Beobachtungsüberzeugungen‘ ins Spiel kommen. Diese würden durch eine gesetzesartig beschreibbare (kausale) Beziehung zwischen ‚intentionalen‘ Subjekten und raumzeitlich-physischen Objekten hervorgerufen. Loars externe

Inputgeneralisierungen kann man so verstehen, dass ein entsprechendes Subjekt unter bestimmten Umständen (normalerweise) zu einer bestimmten ‚Beobachtungsüberzeugung‘ gebracht wird. Wenn also das Subjekt S zu einem Zeitpunkt t in einer Beobachtungsbeziehung B zu einer Tatsache, dass p steht, dann glaubt S , dass p ¹⁹¹. Allerdings unterliegen solche Generalisierungen (wie durch das in Klammern gesetzte „normal“ bereits angedeutet wird) noch einigen Einschränkungen. Denn erstens ist es natürlich möglich, den – sozusagen vor den Augen liegenden – Tatsachen ‚die Aufmerksamkeit zu verweigern; und zweitens könnte man skeptischerweise geneigt sein, der Evidenz der Sinne nicht zu trauen. Dies bringt Loar zu folgender Formulierung der Form von externen Inputgeneralisierungen:

(IG_e) Wenn p wahr ist, $B(p, t, S)$, S zu t gegenüber der Tatsache, dass p , aufmerksam und offen ist, dann glaubt S zu t , dass p .

Man kann sich natürlich fragen, ob diese Formulierung vollständig ist. Vor allem wird keine Anforderung an die epistemischen Bedingungen gestellt, unter denen S eine ‚Beobachtungsüberzeugung‘ bilden wird. Dies scheint jedoch notwendig, um die Möglichkeit des Irrtums und der Täuschung berücksichtigen zu können. Mit einer Qualifizierung von epistemischen Bedingungen (z.B. als adäquat, normal oder gar ideal) käme jedoch eine *normative* Komponente ins Spiel, die sich nicht ohne weiteres naturalistisch ‚reparieren‘ lässt. Darauf werde ich noch zurückkommen. ‚Beobachtungsüberzeugungen‘ sollen jedenfalls durch *De-dicto*-Zuschreibungen in ‚direkten Wahrnehmungssituationen‘, im Sinne einer Angabe von Wahrheitsbedingungen *mit Zeit-Subjekt-Index*, individuiert werden¹⁹². Loar will so einen beobachtbaren, kausalen Einfluss physischer Dinge oder Ereignisse auf *perspektivische* Subjekte als externe Beschränkung funktionalistischer Theorien plausibel machen, ohne in die Gefahr zu geraten, dem Subjekt widersprüchliche Überzeugungen zuschreiben zu müssen (was ja die internen Beschränkungen nicht zulassen).

Die komplizierten Details von Loars Konzeption müssen hier nicht interessieren; für eine Diskussion ihrer grundlegenden Schwierigkeiten sind sie nicht notwendig. Immerhin vermeidet es Loars Vorschlag, die externe Beschränkung der Theorie über eine, für die Individuierung von propositionalen Gehalten, bzw. Einstellungen offensichtlich ungeeignete, reduktionistische Rede von Reizen oder Motorik erreichen zu wollen. Damit vertritt er eigentlich bereits eine *externalistische* Position, deren

¹⁹¹ Loar (1981), S. 67 f.

¹⁹² Loar (1981), S. 97 f. Loar spricht hier von „beliefs about perceived particulars“.

repräsentationaler Charakter sich allerdings auf einen, wie ich meine, schwerlich abzugrenzenden Bereich von ‚wahrnehmungsnahen‘ Überzeugungen beschränken soll. Auf Fragen, die sich im Zusammenhang mit externalistischen Positionen des Naturalismus im Allgemeinen stellen, werde ich an verschiedenen Stellen noch ausführlich eingehen. Zunächst sollen hier aber noch einige grundlegende Probleme gehaltbasierter, ‚realistischer‘ Theorien angeführt werden, die auch nicht-externalistische Theorien betreffen.

Ein häufig erkanntes Problem jeder Form des Funktionalismus ist das sogenannte *Anzahlproblem*¹⁹³. Es besteht darin, dass wir den meisten Menschen mit gutem Recht eine Unmenge an Überzeugungen zuschreiben könnten, darunter auch solche, die diese sich weder jemals bewusst gemacht haben, noch jemals irgendeine Rolle für ihre Handlungspraxis gespielt haben mögen. So bin ich beispielsweise davon überzeugt, dass Schnee weiß und Gras grün ist. Ich bin aber auch davon überzeugt, dass ich größer als einen Meter bin, und ebenso, dass ich kleiner als fünf Meter bin. Auch glaube ich, dass Honig eine andere Konsistenz hat, als Whiskey und dass 45872 größer ist als 45869. All dies würde ich bejahen, wenn man mich jetzt fragen würde, ob ich es für wahr halte, auch wenn ich mir einige dieser Überzeugungen dabei zum ersten Mal bewusst machte. Wenn Überzeugungen aber funktionale Zustände sein sollen und funktionale Zustände ihrerseits als physische Zustände realisiert sein müssen, dann stellt sich die Frage, ob es plausibel ist, anzunehmen, dass die potentiell unendlich hohe Zahl an Überzeugungen, die man mir zu Recht zuschreiben könnte, tatsächlich als physische Zustände in mir realisiert sein können. Eine übliche Weise, auf dieses Problem zu reagieren, besteht darin, zwischen *impliziten* und *expliziten* (oder ‚aktuellen‘ und ‚virtuellen‘) ‚intentionalen Zuständen‘ oder ‚Ereignissen‘ zu unterscheiden¹⁹⁴. Doch wie auch immer diese Unterscheidung von Funktionalisten ausbuchstabiert werden mag, es ist am Ende fraglich, ob sie klare Kriterien einer entsprechenden Einteilung liefern kann. Nach Heckmann würde mit der Unterscheidung von impliziten und expliziten Einstellungen zudem der grundsätzliche Anspruch aufgegeben, *alle* zuschreibbaren propositionalen Einstellungen als physische Instantiierungen funktionaler Rollen zu verstehen¹⁹⁵.

Ein wesentlicher Kritikpunkt Schiffers am Funktionalismus betrifft den *ceteris-paribus*-Charakter der ‚Gesetze‘ einer ‚intentionalen Psychologie‘ und entspricht in etwa dem oben gegen Churchlands ‚Handlungsgesetz‘ vorgebrachten Analytizitätseinwand. Da entsprechende kontrafaktische Generalisierungen nicht nur von Funktionalisten als verhaltensklärend angenommen werden, trifft diese Kritik

¹⁹³ S. etwa Heckmann (1994), S. 124. f.

¹⁹⁴ Für eine Erörterung, siehe z. B. Lycan (1988), Kap. 3.

¹⁹⁵ Heckmann (1994), S. 126.

auch andere Formen des Naturalismus. Nehmen wir einfache Beispiele solcher Generalisierungen, wie etwa:

Wenn X glaubt, dass (p nur wenn q) und p wünscht und..., dann wünscht X , dass q
oder

Wenn X glaubt, dass ((wenn p , dann q) und p) und..., dann glaubt X , dass q .

Wie wir bereits gesehen haben, wären ‚Intentionale Generalisierungen‘ ohne die (durch die Pünktchen symbolisierte) Offenheitsklausel sicher falsch. Jemand mag sich sehr wünschen, reich zu sein, glauben, dass man reich wird, wenn man eine Bank ausraubt und dennoch nicht wünschen, eine Bank auszurauben. Die Offenheitsklausel jeweils ausfüllen zu wollen, würde aber ähnliche Schwierigkeiten mit sich bringen, wie sie bei behavioristischen Reduktionsversuchen propositionaler Einstellungen auftraten und in der Einleitung bereits dargestellt wurden: Wir könnten nämlich eine Menge an einschränkenden, bzw. relativierenden weiteren Überzeugungen oder übertrumpfenden, stärkeren Wünschen aufzählen. Außerdem müssten wir auch Möglichkeiten wie Selbsttäuschung, intellektuelle Trägheit, geistige Erkrankungen oder Drogeneinfluss ausschließen.

Ähnliches gilt auch für Inputgeneralisierungen: Wie ließe sich etwa eine ‚alltagspsychologische‘ Inputgeneralisierung, wie:

Wenn sich direkt vor X ein roter Klotz befindet und..., dann wird X glauben, dass sich vor ihm ein roter Klotz befindet.

vervollständigen? Nun, X darf zumindest nicht blind oder farbenblind sein, die Beleuchtungsverhältnisse sollten normal sein; X muss alt genug sein, um zu wissen, was ein Klotz ist; sie darf nicht geblendet werden, ein Tuch vor die Augen gebunden haben, abgelenkt oder geistig verwirrt sein oder unter Drogeneinfluss stehen usw. Auch hier ist ein Ende der Liste nicht absehbar¹⁹⁶. Im Falle der Inputgeneralisierungen kann man die *Ceteris-paribus*-Klausel als Unterstellung üblicher *Normalitätsbedingungen* verstehen. Die unmögliche Aufgabe, eine (mutmaßlich) nicht-endliche Liste zu vervollständigen, entspräche in diesen Fällen der Aufgabe, *normative* Bedingungen, wie etwa: „ X ist ein kompetenter Sprecher“ oder „Die Beleuchtungs- oder die Wahrnehmungsverhältnisse sind normal“ naturalistisch zu reformulieren, bzw. zu reduzieren.

¹⁹⁶ Schiffer (1987), S. 30 f.

Fodor (1988) hält diesem Einwand entgegen, dass mit der gleichen Berechtigung die kontrafaktischen Generalisierungen vieler Spezialwissenschaften, wie etwa der Ökonomie oder der Geologie, kritisiert werden könnten, denn auch sie sind mit einer *ceteris paribus* Einschränkung belastet. Es liege uns aber normalerweise fern, wegen dieser Eigenschaft den eigenständigen Erklärungswert und die Nützlichkeit solcher Wissenschaften und ihrer Gesetze in Frage zu stellen. Die einzige Wissenschaft mit strikten Gesetzen (also ohne Offenheitsklausel) sei die Physik.

Wir können Fodor darin zustimmen, dass uns dies nicht daran hindert, gesetzesartige (möglicherweise statistische, bzw. probabilistische) Erklärungen auch in anderen wissenschaftlichen Disziplinen als der Physik vorzunehmen. Damit ist allerdings noch nicht entschieden, ob auch Erklärungs- (bzw. Verstehens-) Praktiken, deren Gegenstand Handlungen – also nicht Verhalten oder Verhaltensmuster – sind, überhaupt an naturwissenschaftliche Erklärungsweisen assimiliert werden können. Und dies ist ja die Frage, um die es gerade geht, wenn wir die Möglichkeit einer ‚intentionalen Psychologie‘ diskutieren.

Fodors Einwand liefert aber einen Hinweis darauf, dass der entscheidende Aspekt von Schiffers Kritik nicht den *ceteris-paribus*-Charakter ‚psychologischer Gesetze‘ betrifft, sondern die augenscheinliche Analytizität solcher ‚Gesetze‘, *falls diese sich tatsächlich vervollständigen ließen*. Schiffer veranschaulicht seine Version dieses Einwandes an einem Beispiel: Ava wich zurück, als sie sah, dass sich ein Auto näherte. Die entsprechende Generalisierung würde etwa lauten: Wenn eine Person ein bestimmtes Resultat vermeiden will, bzw. zu vermeiden wünscht (hier wohl, von einem Auto überfahren zu werden), glaubt dass sie dies durch eine Handlung eines bestimmten Typs erreicht (also durch Zurückweichen auf den Bürgersteig), glaubt, dass sie in der Lage ist eine Handlung dieses Types auszuführen, nicht glaubt, dass es eine bessere oder gleich gute Möglichkeit gibt, das unerwünschte Resultat zu vermeiden (usw., dann) ... wird sie eine Handlung des betreffenden Typs ausführen. Dazu bemerkt Schiffer:

„(...) to the extent that we can fill (this „generalisation“) out, to that extent it begins to look more and more expressive of truth constitutive of our propositional-attitude concepts, and thus less and less like a contingent causal law.“ (1987, S. 148)

Schiffers Beispiel entspricht hier wieder der Form des Praktischen Syllogismus, doch der Analytizitätseinwand betrifft auch Inputgeneralisierungen bzw. ‚funktionalistische Gesetze‘ ganz allgemein. Solche ‚Gesetze‘ sind nicht falsifizierbar, weil die für sie notwendigerweise in Anschlag zu bringenden Optimierungsannahmen, etwa auch in

Form von normalen Beobachtungsbedingungen bei Inputgeneralisierungen, gleichsam selbstimmunisierend wirken: Immer wenn wir sich die behauptete Wirkung nicht einstellt, müssten wir davon ausgehen, dass das ‚Gesetz‘ nicht anzuwenden war, weil die entsprechenden Bedingungen nicht gegeben waren.

Nun könnte man möglicherweise versucht sein, mit den Achseln zu zucken und zu behaupten, dass nicht nur *alle ceteris-paribus*-Gesetze, sondern darüber hinaus auch die strikten Gesetze einen konventionellen, bzw. analytischen Charakter haben. Zumindest Naturalisten können so jedoch nicht argumentieren, denn eine wesentliche Voraussetzung ihrer Kritik am Analytizitätseinwand war ja, dass *alle* Generalisierungen einen *empirischen Sinn* haben (wobei man sich üblicherweise auf Quines einschlägige Kritik an einer klaren Trennbarkeit analytischer und synthetischer Aussagen beruft). Nun hat insbesondere auch Fodor Einwänden im Sinne des Schifferschen immer wieder entgegengehalten, dass diese auf alle, also auch die naturwissenschaftlichen Gesetze anwendbar wäre, aber falsch seien, da sie auf einer Verwechslung beruhen:

„What seems to be *analyticity*, (conventionality, etc.) of these (den physikalischen / W.-J. C.) laws is just their *centrality* misperceived; the doctrine that „ $F=MA$ “ (as it might be) is quasi-analytic postulates a semantic fact where there is only an epistemic one. But if this is true of the putative analyticity of the basic laws of mechanics, why shouldn't it also be true of the putative analyticity of the basic decision-theoretic laws that relate beliefs and desires to one another and to actions.“ (1988, S. 184).

Dieses Argument hatte ja in ähnlicher Form auch schon Churchland vorgebracht. Ich halte es, aus den bereits genannten Gründen, für nicht überzeugend. Die *Zentralität* bestimmter physikalischer Gesetze ist eben nicht dasselbe, wie die *Analytizität* ‚intentionaler‘ Generalisierungen, denn erstere, aber nicht letztere, erfüllen, trotz ihrer Zentralität, zwei für Kausalgesetze wesentliche Bedingungen, nämlich ihre prinzipielle Falsifizierbarkeit und die Möglichkeit einer unabhängigen Feststellung des Eintretens desjenigen Ereignisses, welches die Ursache bildet (wenn auch nur innerhalb eines spezifischen theoretischen Rahmens). Diese Einwände betreffen aber keineswegs alle Formen von *empirischen ceteris-paribus*-Generalisierungen, sondern nur Handlungsrationalisierungen und andere Formen ‚intentionaler Gesetze‘, wie etwa Wahrnehmungsschlüsse (‚Inputgeneralisierungen‘). Und in Bezug auf diese, weist der Analytizitätseinwand, nach meinem Dafürhalten, darauf hin, dass wir es hier gar nicht mit empirischen Verallgemeinerungen zu tun haben, sondern um *normative Stellungnahmen zur richtigen* (vernünftigen) *Urteilsbildung* einerseits oder *transzendente* (Vernunfts-) Prinzipien des Verstehens andererseits handelt. Diese, oben

bereits im Zusammenhang mit der Diskussion der Positionen Glüers und Brandoms entwickelte These, wird, wie ich hoffe, durch die in der Folge noch zu diskutierenden Schwierigkeiten von Theorien funktionaler Rollen zusätzlich an Plausibilität gewinnen.

Dafür ist zunächst auch ein von naturalistischer Seite gegen ‚realistische‘ Versionen funktionalistischer Theorien vorgebrachter Einwand aufschlussreich, der dem ‚neo-behavioristischen‘ Lager (insbesondere Quine und Dennett) zuzuordnen ist. Dieser Einwand beruht auf der These der ‚Unter-‘ oder ‚Unbestimmtheit‘ des Gehaltes von ‚intentionalen Zuständen‘ und ‚Ereignissen‘. Gemäß diesem Einwand ist nämlich deren *eindeutige* Bestimmung durch funktionale Rollen angesichts der verfügbaren Daten (beobachtbares Verhalten und Umweltbedingungen) und den internen Beschränkungen der Theorie (Rationalitätsprinzipien), gar nicht möglich. Genau dies aber wird von Loar explizit behauptet:

„ (...) A proposition can index a belief’s systematic role by having a *unique* place in a system of logical and ‚meaning‘ relations and extra-logical relations (between observational conditions and external conditions), for that system of intensional relations can mirror counterfactual relations among underlying states and between some of them and external conditions.“ (Loar 1981, S. 64)
(Hervorhebung von mir /W.-J. C)

Angesichts der Tatsache, dass Loars ‚Wahrnehmungsergebnisse‘ als externe Beschränkungen der Theorie nur für einige der möglichen propositionalen Einstellungen unmittelbar relevant sein können, stellt sich, aus einer Beobachterperspektive, die Frage nach einer *eindeutigen* Bestimmbarkeit funktionaler Rollen besonders im Hinblick auf ‚empirie-ferne‘ theoretische oder abstrakte Einstellungen. Solche Einstellungen sollen ja *unmittelbar* nur durch eine Subsumtion unter solche Arten funktionaler Generalisierungen bestimmt sein, die internen Beschränkungen im Sinne von ‚Rationalitätsprinzipien‘ (etwa der Formalen Logik) genügen. Aus der *Perspektive empirischer Theoriebildung* aber scheint es gute Gründe dafür zu geben, dass für die Zuordnung funktionaler Rollen zu ‚intentionalen Zuständen‘ oder ‚Ereignissen‘ prinzipiell das gleiche gilt, wie das, was Quine bezüglich der Möglichkeit einer eindeutigen Zuordnung (‚Interpretation‘) sprachlicher Bedeutung (bzw. Referenz) zu einem beobachtbaren Sprachverhalten im Falle der Übersetzung einer völlig unbekanntem Sprache in die eigene behauptet hatte: Auch im Hinblick auf durch funktionale Rollen individuierte ‚intentionale Zustände‘ oder ‚Ereignisse‘, wird es mehrere unterschiedliche, bzw. miteinander inkompatible Zuordnungsschemata geben, die sowohl den internen, als auch den externen Beschränkungen der Theorie

genügen¹⁹⁷. Es scheint somit *prinzipiell* unmöglich zu sein, propositionale Einstellungen über funktionale Rollen als *intrinsisch*, also unabhängig von der Wahl eines Zuordnungsschemas *eindeutig bestimmt* verstehen zu können. Da die internen und externen Beschränkungen nicht *per se* ein einzelnes, explanativ adäquates Zuordnungsschema für funktionale Rollen festlegen, wäre die Wahl eines der möglichen Schemata schließlich eine Frage pragmatischer, vielleicht sogar ästhetischer, Beurteilungskriterien. Dieses Resultat ist natürlich mit einem ‚intentionalen Realismus‘ unvereinbar

Nun setzt die These von der Unter- oder Unbestimmtheit der Bedeutung (bzw. des ‚Intentionalen‘ ganz allgemein) offenbar einen *Holismus* voraus, nach dem die Bedeutung jedes einzelnen Satzes oder der Gehalt jeder einzelnen Einstellung nicht unabhängig ist von der Bedeutung anderer Sätze, bzw. dem Gehalt anderer Einstellungen. Zumindest in eingeschränkter Form wird genau dies vom Funktionalismus auch behauptet, da die funktionale Definition einer propositionalen Einstellung ja in einer Konjunktion (mindestens) all derjenigen Generalisierungen bestehen soll, die deren unmittelbare Kausalbeziehungen zu anderen Einstellungen erfassen sollen. Es scheint mir allerdings zweifelhaft, ob Funktionalisten tatsächlich bei einem solch ‚schwachen Holismus‘ stehen bleiben können, denn diejenigen ‚intentionalen Zustände‘ und ‚Ereignisse‘, die in einem unmittelbaren Zusammenhang mit einem bestimmten solchen ‚Zustand‘ oder ‚Ereignis‘ stehen sollen, wären ja ihrerseits durch ihr Verhältnis zu anderen solcher ‚Zustände‘ oder ‚Ereignisse‘ bestimmt usw. Ein ‚Intentionales System‘ kann seine Wünsche oder Überzeugungen natürlich punktuell ändern, was, zumindest für uns, ja durchaus nichts Ungewöhnliches, sondern eher der Regelfall ist. Wenn aber nur *einer* der für die Bestimmung eines funktional individuierten ‚intentionalen Zustandstyps‘ (IZ_{f1}) relevanten, seinerseits funktional bestimmten Zustandstypen (also z.B. (IZ_{f2})) seine Rolle dadurch ändert, dass er nicht mehr unter eine derjenigen Generalisierungen fällt, die dessen Kausalverhältnis zu wiederum anderen Zustandstypen (also z.B. (IZ_{f3})) erfassen, dann kann damit auch die Generalisierbarkeit des kausalen Verhältnisses zwischen (IZ_{f1}) und (IZ_{f2}) hinfällig werden. In diesem Fall aber wäre der für einen Zeitpunkt t_0 bestimmte ursprüngliche Zustandstyp (IZ_{f1}) nicht mehr mit dem entsprechenden, für t_1 bestimmten Zustandstyp (IZ_{f1neu}) identisch. – Schlimmer noch:

¹⁹⁷ Dieses Argument setzt allerdings ein Verständnis von Bedeutung, bzw. Bedeutungsgleichheit voraus, das ich nicht teile. Daher spielt es hier lediglich die Rolle, eine *naturalismus-intern* berechtigt erscheinende Schwierigkeit zu benennen, weswegen ich auch auf eine ausführlichere Darlegung an dieser Stelle verzichte. Wir werden uns im Rahmen der Diskussion der Positionen Quines und Davidsons (und Dennetts) noch grundsätzlich mit der Unbestimmtheitsthese und dem ihr zugrundeliegenden Verständnis von Bedeutung auseinandersetzen.

Konsequenterweise müssten sehr viele andere, vielleicht sogar alle funktionalen Zustände oder Ereignisse, dann ebenfalls einer Rollenänderung unterliegen, da es keine natürliche Grenze für jenen ‚Dominoeffekt‘ gibt, wenn man davon ausgeht, dass kein Bereich der Sprache semantisch gleichsam autark ist. Ein solcher Holismus ließe es, angesichts der Uneinheitlichkeit von kognitiven und volitiven Einstellungen sowohl bei verschiedenen Individuen, als auch bei einzelnen Individuen zu unterschiedlichen Zeitpunkten, sehr unplausibel erscheinen, dass es eine *allgemeine* funktionalistische Theorie ‚Intentionaler Systeme‘ geben kann. Es wäre dann nämlich extrem unwahrscheinlich, dass es überhaupt zeit- und individuenübergreifende funktionale *Rollenidentitäten* gibt.

Eine mögliche Reaktion auf dieses Problem seitens eines Funktionalisten könnte darin bestehen, den Holismus irgendwie zu begrenzen, etwa indem eine Theorie funktionaler Rollen einige der möglichen Veränderungen innerhalb von ‚Überzeugungs-Wunsch-Systemen‘ als *relevant* für die Individuierung von bestimmten ‚intentionalen Zuständen‘ oder ‚Ereignissen‘ durch funktionale Rollen verstanden werden und andere nicht. Wie aber ließe sich diese Relevanz begründen? Eine mögliche Berufung auf so etwas wie eine begriffliche Analyse wird ja von den meisten Naturalisten unter Berufung auf Quines Kritik an der Unterscheidbarkeit begrifflicher und empirischer Fragen, bzw. des Analytischen und Apriorischen vom Synthetischen und Aposteriorischen, abgelehnt. Doch wie auch immer man zu dieser Kritik steht, es wird deutlich, dass völlig unklar ist, worauf sich eine allgemeine Theorie funktionaler Rollen überhaupt stützen könnte. Eine Berufung auf ‚Alltagspsychologie‘, etwa im Sinne eines *faktischen Konsenses* über Wahres und inferentielle Zusammenhänge, dürfte schnell an Grenzen gelangen. Kann es dann eine Frage empirischer Forschung sein, die Lücken auszufüllen? Wie sollte das vor sich gehen? Hier deutet sich ein für den Funktionalisten unerfreuliches Dilemma an: Will er eine empirische – und keine normative – Theorie, darf er sich natürlich nicht an den *eigenen Urteilen* über inferentielle Zusammenhänge orientieren. Stützt er sich aber auf die einzige Form von Empirie, die hier überhaupt denkbar ist, nämlich das *faktische Urteilsverhalten* ‚intentionaler‘ Subjekte, könnte er im Rahmen einer *funktionalen* Theorie bestenfalls zu *statistischen* Aussagen über die Wahrscheinlichkeit eines zu erwartenden Verhaltens gelangen. Wir hätten es jedenfalls nicht mehr mit einer mit Allgemeinheitsanspruch auftretenden Theorie *der Natur propositionaler Einstellungen* im Sinne funktionaler Rollen zu tun. Dies möchte ich im Folgenden etwas ausführlicher erläutern.

Diskutieren wir zunächst mal die Unterstellung, wir könnten uns in weiten Teilen der Bestimmung funktionaler Rollen von propositionalen Einstellungen auf

unproblematische Generalisierungen einer ‚Alltagspsychologie‘ berufen. Funktionalisten, wie Loar, scheinen hier von einer *definitiven Bestimmbarkeit* jener inferenziellen Beziehungen auszugehen, die die funktionalen Rollen propositionaler Einstellungen definieren sollen. Das Problem, welches sich dabei stellt, könnte man als das der *Wahl, bzw. der Auflistbarkeit* der (theorie-) externen und internen Beschränkungen von als allgemein verstandenen Theorien ‚Intentionaler Systeme‘ bezeichnet¹⁹⁸. So stellt sich zunächst die Frage, ob die von Loar angeführten internen und externen Beschränkungen funktionalistischer Theorien des Geistes die Bandbreite dessen erschöpfen, was wir unter Begründungs- oder Rechtfertigungsrelationen fassen können. Es wäre offensichtlich unzureichend, sich lediglich auf logische und praktisch-inferentielle (dem praktischen Syllogismus entsprechende) Generalisierungen zu beschränken. Doch auch die darüber hinaus von Loar angeführten „*meaning relations*“, im Sinne material-logischer Schlüsse, wie „Wenn etwas blau ist, so ist es farbig“ (was ja einer kontrafaktischen Verknüpfung aller Realisierungen der funktionalen Rolle des ‚Überzeugungszustandes‘ ‚Dies ist blau‘ mit all jenen des ‚Überzeugungszustandes‘ ‚Dies ist farbig‘ entsprechen soll), reichen, wie ich denke, nicht aus, der Komplexität wechselseitiger Abhängigkeiten und möglicher Abwägungen von propositionalen Einstellungen im Rahmen unserer kognitiven Kompetenz gerecht zu werden – auch wenn dadurch die internen Beschränkungen funktionalistischer Theorien bereits erheblich ausgeweitet werden. Offenbar gibt es doch darüber hinaus sehr *unterschiedliche Arten von Gründen*, die wir anführen könnten, um eine bestimmte Überzeugung zu rechtfertigen. Für die Vernünftigkeit des Habens oder Bildens von propositionalen Einstellungen sind oft Arten von Beziehungen zwischen diesen relevant, die nicht einfach in praktisch-syllogistischen, formal-logischen und material-logischen Schlusschemata aufgehen. Einfache Beispiele bilden etwa für richtig gehaltene Maximen, wie „Bei wichtigen Entscheidungen sollte man vor allem den eigenen Intuition zu vertrauen“ Einschätzungen, wie „Bei diesem (miesem) Wetter solltest Du nicht fahren“, einfache Lebensweisheiten, wie „Was sich neckt, das liebt sich“ oder alle möglichen Formen empirischer Feststellungen. Bei entsprechende Begründungen kann nämlich ein komplexes Netz an persönlichen Erfahrungen, Tatsachenunterstellungen, Wahrscheinlichkeitsabwägungen und Bewertungen zu Grunde liegen. Ob ich beispielsweise mit Gründen glaube, dass man jetzt zur ‚Tagesschau‘ einschalten kann, hängt auch davon ab, ob ich glaube, dass ein Fernseher in der Nähe ist, dass wir uns im Sendebereich der ‚Tagesschau‘ befinden, dass die Tagesschau normalerweise pünktlich kommt und beispielsweise nicht gerade ein Fußballländerspiel übertragen wird usw. Ähnliches gilt, in abgeschwächter Form, selbst

¹⁹⁸ Vergl. Heckmann (1994), S. 130 f.

für einfache, empirienahe, bzw. situative Überzeugungen (z.B., dass es draußen schneit) und zwar insofern, als wir – implizit und erforderlichenfalls begründungsbedürftig – *unterstellen*, die *Wahrnehmungsbedingungen* seien (mehr oder minder) *normal*. Des Weiteren sind Erwägungen, wie die, dass es mitunter sinnvoll ist, eine bisher wohlbegründete Überzeugung auch bei momentan scheinbar gegebenen Gegenevidenzen beizubehalten, auch in unseren vorthoretischen Praktiken üblich. Oft nehmen wir bei der Beurteilung eines Aussagegehaltes eine intuitive Globalabwägung des *pro* und *contra* im Lichte unserer allgemeinen Erfahrung dessen vor, worauf oder auf wessen Urteil man sich normalerweise verlassen kann; wir bedenken beispielsweise, was im Fall eines vorliegenden Aussagegehaltes besonders zu berücksichtigen sein könnte (z.B. wer ein bestimmtes Interesse an der Behauptung seiner Wahrheit oder Falschheit haben könnte), welche Art von Begründung normalerweise triftig ist u. ä. m.¹⁹⁹.

Alle diese Arten von möglichen Gründen, bzw. von Weisen begründeter Urteilsbildung stellen jedoch klarerweise eine für die Bestimmung der – im weitesten Sinne – inferentiellen Rollen von Überzeugungen (und anderer Einstellungen, da sich ähnliches ja offensichtlich auch für volitive propositionale Einstellungen, wie Wünsche oder Absichten zeigen lässt) notwendige Erweiterung der von Loar berücksichtigten Formen interner Beziehungen zwischen propositionalen Einstellungen dar. Die Vielzahl und Vielfalt möglicher Begründungsbeziehungen scheint unüberschaubar. So kritisiert auch Heckmann (a. a. O. S. 129 f.) den Funktionalismus dahingehend, dass es keine definitive Liste, bzw. keine vollständige disjunktive funktionale Generalisierung aller möglichen Gründe (bzw. entsprechender Schlüsse) dafür, eine bestimmte Überzeugung zu haben, bzw. dass es keinen *Algorithmus vernünftiger Überzeugungsbildung* geben könne. Allerdings scheint Heckmann dies lediglich auf die potentiell unendliche Variabilität der Anpassungsmöglichkeiten holistischer Systeme an externe Beschränkungen und die Heterogenität möglicher funktionaler Organisationen zurückzuführen. Damit wird er aber, wie ich meine, dem eigentlichen

¹⁹⁹ Wenn jemand beispielsweise glaubt, dass Delphine intelligente Lebewesen sind, so mag dies aus einer *komplexen Menge persönlicher Erfahrung* im Umgang mit diesen Tieren, in Kombination mit bestimmten, vernünftigen *Kriterien* für intelligentes Verhalten, entspringen. Sowohl die Erfahrungen, als auch die entsprechenden Kriterien ihrer Bewertung, hätten allerdings durchaus anders ausfallen können, und dennoch hätte man (gerechtfertigterweise) zu der gleichen Überzeugung gelangen können. Jemand könnte aber auch durch das Urteil eines für *kompetent* gehaltenen Bekannten oder durch einen Bericht im Fernsehen zu dieser Ansicht gekommen sein. Oder dadurch, dass er/sie jemanden, der als ‚Experte‘ für die japanische Fischfangindustrie arbeitet, das Gegenteil behaupten hört – Grund genug, wegen der *Unterstellung eindeutiger Interessenlagen*, davon auszugehen, dass er lügt. Die Aufzählung der Möglichkeiten einer entsprechenden Überzeugungsbildung ließe sich sicherlich fortsetzen, ohne dass eine klare Grenze in Sicht käme.

Problem noch nicht wirklich gerecht. Die genannten Schwierigkeiten des Funktionalismus liegen nämlich tiefer, sie gehen, wie ich denke, auf einen *Kategorienverstoß* zurück. Das Problem besteht nicht einfach nur darin, dass wir, in Anbetracht aller möglichen (und vielleicht noch unbekannt) Umstände, die Bedingungen vernünftiger Urteilsbildung nicht definitiv bestimmen können. Es besteht darüber hinaus darin, dass schon die Vorstellung von einem jeweils *an sich einzig richtigen* oder doch zumindest *vernünftigsten* Urteil, welche die funktionalistische Idee einer disjunktiven Definierbarkeit von Überzeugungen unterstellt – jedenfalls in vielen Fällen – *keinen Sinn* macht. Dies lässt sich bereits durch die pragmatistische Einsicht verdeutlichen, dass sich Äußerungen, wie „Du hast recht“, „Das ist wahr“, „Dies ist keine gute Begründung“, „Dies folgt aber nicht aus dem bisher Gesagten“ oder „Das zu glauben, halte ich für vernünftig“ nicht einfach als *Beschreibungen von Eigenschaften* von Dingen, Ereignissen, Zuständen oder Beziehungen (wie etwa von Sätzen, Äußerungen, Urteilen, Überzeugungen oder gar Schlüssen oder Propositionen) verstehen lassen. Solche Äußerungen haben, abgesehen von jenen oben diskutierten konstitutiven (bzw. transzendentalen) Fällen im Sinne Schnädelbachs und Glüers, immer auch einen *wertenden und normativen Charakter* – etwa im Sinne von Zustimmung oder Berechtigungserteilung. Mit ihnen werden, wenn ich recht habe, Formen von Verantwortlichkeit zugeschrieben oder übernommen. Auf jeden Fall sollte klar sein, dass für entsprechende Äußerungen keine *gesonderten empirischen Bestätigungsverfahren* zur Verfügung stehen, das heißt, solche, die über die empirischen Bestätigungsverfahren von jenen Sätzen oder Überzeugungen, auf die sie sich beziehen, hinausgehen. Und auch wenn es dennoch in einer gewissen, *genuinen* Weise sinnvoll sein kann, zu sagen, dass wir Begründungsbeziehungen auch (durch Einsicht) entdecken, wäre es irreführend, diesen Aspekt über zu betonen und sie als etwas schlicht *Vorzufindendes* verstehen. Dies hängt zum einen damit zusammen, dass die Frage, was eine Begründung oder eine gutes Argument für was ist, unter unterschiedlichen Aspekten unterschiedlich beurteilt werden kann. Und die Frage, welcher dieser Aspekte der maßgebliche sein soll, kann wiederum unter unterschiedlichen Aspekten unterschiedlich beurteilt werden usw. Am Ende hängt das, was wir für eine gute Begründung oder für das Vernünftigste zu glauben oder zu tun halten, häufig auch von subjektiven Präferenzen, bzw. Bewertungen des Guten ab (wie etwa im Fall der oben bereits genannten Einschätzung: „Bei diesem (miesem) Wetter solltest Du lieber nicht fahren“). Zum anderen sind Begründungsbeziehungen ja nicht sozusagen ‚freistehend‘, sondern bedingt durch *Verständnisse der involvierten Begriffe*. Diese Verständnisse sind nicht nur kontext- und praxisbezogen, sondern können auch einer dynamischen, zukunfts-offenen Entwicklung unterliegen (auf diesen Aspekt werde

ich im nächsten Abschnitt ausführlich zurückkommen). Zudem haben es explizite Bewertungen ganz generell an sich, nie reflexiv die Bedingungen ihrer Richtigkeit vollständig explizieren zu können, da die Auszeichnung der Explikationsversuche – als gut begründet – ihrerseits eine explizite evaluative Stellungnahme darstellt²⁰⁰.

Wenn nun ein Funktionalist die genannten Schwierigkeiten zu vermeiden sucht, gerät er unweigerlich auf die andere Seite des Dilemmas, nämlich die Infragestellung des Allgemeinheitsanspruches seiner Theorie. Funktionalistische Theorien waren ja ursprünglich mit dem Vorsatz angetreten, die Bedingungen einer universalen – also auch spezies-unspezifischen – materialistischen Theorie ‚intentionaler Zustände‘ oder ‚Ereignisse‘ angeben zu können; einer Theorie, welche die diesbezügliche Unplausibilität von psycho-physischen Typenidentitätstheorien vermeidet. Doch selbst wenn wir die im vorigen Absatz genannten grundsätzlichen Einwände mal außer acht lassen, muss sich doch auch ein eingefleischter Funktionalist die Frage stellen, wie funktionale Rollen bereits im scheinbar einfachen Fall, nämlich den Wahrnehmungsüberzeugungen, für *alle* Formen ‚Intentionaler Systeme‘ (eine Kennzeichnung, die nach Einschätzung naturalistischer Philosophen ja grundsätzlich nicht nur auf Menschen zutreffen können soll) *einheitlich* über eine Subsumtion unter entsprechende Inputgeneralisierungen bestimmt werden können sollen? Vielleicht ließe sich der Unterschied zwischen ‚Intentionalen Systemen‘ mit offensichtlich abweichenden kognitiven oder perzeptuellen Vermögen, also etwa zwischen einer Biene, einer Fledermaus, einem sehenden oder einem blinden Menschen, im Hinblick auf sehr unspezifische Überzeugungen, wie beispielsweise „Vor mir befindet sich ein Hindernis“, noch dadurch umgehen, dass man, ganz im Sinne Loars, die externen Beschränkungen nicht reduktionistisch, auf der Ebene sensorischer Stimulation, sondern durch eine kausale Beziehungen zu subjekt-externen *Gegenständen* oder *Tatsachen* beschreibt. Auf Theorien, die Gehalte grundsätzlich als durch ihre externen Ursachen bestimmt verstehen wollen, werden noch genauer eingehen. Doch zumindest hinsichtlich des universalistischen Anspruchs, würden man so noch nicht der Diskrepanz zwischen Bienen, Fledermäusen, Kleinkindern, zivilisationsfern lebenden Kulturformen und (normal gebildeten) Erwachsenen moderner Gesellschaften

²⁰⁰ Dieses Argument lässt sich, in spezifischer Weise auf Vernünftigkeit bezogen, vielleicht auch als ein transzendentes formulieren: Die zugrundeliegende Überlegung läuft darauf hinaus, dass jeder Versuch einer definitiven Explikation dessen, was rational oder vernünftig ist, seinerseits auf seine Vernünftigkeit (z. B. im Sinne der Angemessenheit oder Vollständigkeit) hin beurteilt werden müsste. Diese Beurteilung müsste sich aber ihrerseits auf Standards des Vernünftigen stützen, die selbst nicht zugleich Gegenstand der Beurteilung sein können. Denn sonst würde man als gültig voraussetzen müssen, wessen Geltung ja zunächst in Frage steht. Damit aber bleibt die vernunftgeleitete Urteilskraft einer *vollständigen* und sozusagen ‚*selbstredenden*‘ Explikation ihrer Bedingungen prinzipiell entrückt. Die Gültigkeit dieses Argumentes ist allerdings keine Voraussetzung der oben angeführten Einwände.

bezüglich deren diskriminatorischer Fähigkeiten gerecht; denn für die ersteren vier Formen ‚Intentionaler Systeme‘ würden viele, der für letztere selbstverständlichen ‚Inputgeneralisierungen‘ kaum zuschreibbar sein (wie etwa in Bezug auf die Wahrnehmung eines Kupplungspedals) – und auch umgekehrt (wie etwa in Bezug auf eine ‚Wahrnehmung‘ im Infrarotbereich. Es dürfte darüber hinaus schwerfallen, sogar jene inferentiellen *Konsequenzen* universell einheitlich zu beschreiben, die selbst den (vermeintlich) einfachsten Wahrnehmungsüberzeugungen zukommen (was für den einen Flucht! bedeuten mag, heißt für den anderen womöglich Futter!).

Zwar sind bestimmte *einfache diskriminatorische Kompetenzen*, wie etwa die Wahrnehmungsfähigkeit roter Gegenstände (die sich unter anderem in einer Disposition zur Äußerung der Lautfolge „Dies ist rot“ manifestieren kann), natürlich sowohl Kleinkindern, als auch Papageien zuschreibbar. Ersteren sollten wir jedoch nur in einem jeweils mehr oder minder eingeschränkten, letzteren in gar keinem Sinne ein *vollwertiges Verständnis* der Aussage „Dies ist rot“ zubilligen, da Kleinkinder kaum, bzw. Papageien gar nicht, über entsprechende, mit dieser Aussage verbundene inferentielle Kompetenzen verfügen. – Papageien *wissen* eben nicht, dass aus der Aussage „Dies ist rot“ die Aussage „Dies ist farbig“ oder „Dies ist nicht blau“ folgt. Dieses Wissen ist aber, zumindest nach unserem üblichen Verständnis, eine notwendige Bedingung dafür, eine entsprechende begriffliche Kompetenz, bzw. eine entsprechende Überzeugung *im primären Sinne*, berechtigterweise zuzuschreiben²⁰¹.

Loar sieht diese Schwierigkeiten und versucht sie zu umgehen, indem er die Reichweite seiner Theorie auf die erwachsenen Mitglieder der Gesellschaftsform des Theoretikers beschränkt:

„We want a functional system somewhere in between universal human psychology and the personal evidential connections that define idiosyncratic conceptual roles (...). That there is such a middling level of funktional organisation is not far-fetched, for members of the same society surely acquire many common structures of dispositions to belief. (...) The *recognitional abilities* we can ascribe to virtually any adult member of our society are numerous (...).“
(Loar 1981, S. 67)

Der Preis einer solchen Einschränkung ist allerdings hoch, denn sie läuft darauf hinaus, den Anspruch auf eine *allgemeine* funktionale Theorie der ‚Intentionalität‘ aufzugeben. Außerdem führt dies in neue Schwierigkeiten: Es scheint nämlich *keine empirische*, sondern eine *normative* Frage zu sein, *wer* als erwachsenes Mitglied unserer Gesellschaft gelten kann, also für wen die Theorie gelten soll. Es spricht vieles dafür, dass sowohl die Grenze zwischen Kindern und Erwachsenen, als auch die zwischen

²⁰¹ So auch bei Stich (1983), Dennett (1981) und vor allem bei Brandom (1994).

unserer Gesellschaft und anderen, zumindest in der fraglichen Hinsicht, nicht sauber zu ziehen ist. Der Funktionalist müsste sich also auf Minimalkriterien begrifflicher Kompetenz *festlegen*, deren Erfüllung notwendige Bedingung für die Anwendbarkeit einer funktionalistischen Theorie ‚intentionaler Zustände‘ wäre. Doch dies wäre mit einem *realistischen* Verständnis propositionaler Einstellungen offensichtlich kaum verträglich.

Darüber hinaus bleibt auch unklar, wieso Loar glaubt, gerade bei dieser Einschränkung der Reichweite funktionalistischer Theorie stehen bleiben zu können und nicht auch die individuellen Idiosynkrasien jeweiliger Instantiierungen funktionaler Organisation von ‚Intentionalität‘ berücksichtigen zu müssen. Es dürfte klar sein, dass die individuellen Kenntnisse von „*evidential connections*“ (oder besser: *was* für eine Begründung für *was* gehalten wird) auch innerhalb der erwachsenen Mitglieder ‚unserer Gesellschaft‘ schwanken. Denn solche Unterschiede betreffen ja nicht nur, durch unterschiedliche Bildungsniveaus in unterschiedlichen Bereichen ungleich verteilte, *diskriminatorischen* Kompetenzen „gebildeter Erwachsener unserer Gesellschaftsform“, sondern, wie gesagt, vor allem auch deren uneinheitliche *Kenntnis* oder *Akzeptanz von möglichen gültigen Gründen, bzw. Begründungsrelationen*.

Das grundsätzliche Dilemma, mit dem sich funktionalistische Theorien funktionaler Rollen konfrontiert sehen, sieht demnach, zusammengefasst, folgendermaßen aus: Entweder orientiert sich die Theorie an den durch einzelne Individuen jeweils *faktisch realisierten* Rechtfertigungs- oder Begründungsprofilen – also der Gesamtheit dessen, was, nach *momentaner Ansicht eines Individuums*, Begründung für *was* ist, um so zu einer empirisch adäquaten Theorie über die tatsächlich Ursachen zu gelangen, mit denen ein individuelles Verhalten je erklärbar sein soll. Diese Alternative verspricht sicherlich noch am ehesten, dem naturalistischen *Credo* einer *Beschreibung* und *Erklärung* von eigentlich natürlichen Zuständen oder Vorgängen nahe zu kommen, aber dazu führen, dass der Allgemeinheitsanspruch funktionalistischer Definitionen von Einstellungstypen, wegen Unterschiedlichkeit faktisch realisierter, individueller Rechtfertigungs- und Begründungsprofile, aufgegeben werden müsste. Das heißt, wir hätten es überhaupt nicht mehr mit einer funktionalistischen Theorie propositionaler Einstellungen, sondern nur noch mit einer funktionalen Theorie zur Erklärung eines individuellen Verhaltens zu tun. Kurioserweise sieht sich der Funktionalismus also, auf einer abstrakteren Ebene, mit Schwierigkeiten konfrontiert, die analog sind zu jenen, an denen schon die kruden Typenidentitätstheorien gescheitert waren und bezüglich denen der Funktionalismus ja eigentlich Abhilfe versprach²⁰². Bestimmte propositionale

²⁰² So bemerkt auch Putnam: „Der Grund für die Einführung des Funktionalismus war ja gerade die Erkenntnis, dass wir keinen physikalischen Zustand ausfindig machen werden, in dem sich alle möglichen Glaubenden befinden müssen, um eine bestimmte Überzeugung oder

Einstellungen lassen sich offenbar genauso wenig in allgemeiner Weise mit bestimmten funktionalen Rollen identifizieren, wie mit bestimmten physischen Zuständen. Die jeweiligen Überzeugungen müsste ja gegebenenfalls für jedes Individuum eigens funktional bestimmt werden, um deren tatsächlichen kausalen Rolle im Hinblick auf eine mögliche Erklärung oder Prognose von Verhalten gerecht werden zu können. Auf jeden Fall könnte von einer *allgemeinen* funktionalen Definition von Einstellungstypen keine Rede mehr sein.

Wenn funktionalistische Theorien nun aber andererseits von den faktischen Differenzen individueller ‚Rechtfertigungsprofile‘ abstrahieren wollten, liefe dies, wie wir oben gesehen haben, darauf hinaus, *evaluative*, bzw. *normativ gehaltvolle Aussagen* darüber zu machen, *welche Begründungsbeziehungen* für Definitionen von funktionalen Rollen jeweils als *gültig* (und wesentlich) und als *welche ungültig* (oder zu vernachlässigen) anzusehen sein sollen²⁰³. Dies würde bedeuten, dass die Formulierung ‚alltagspsychologischer‘ Verallgemeinerungen nicht nach dem Muster von Kausalgesetzen, sondern eher im Sinne von (stellungnehmenden) *Vernunftspostulaten* zu deuten wären, die sagen, welche Einstellungen man unter bestimmten Umständen von Subjekten *erwarten sollte, wenn sie vernünftig* sind. Dies ist offenbar mit üblichen Verständnissen empirischer Verallgemeinerungen unverträglich.

So oder so können Funktionalistische Theorien demnach ihrem Anspruch auf eine allgemeine, erklärende Theorie des Geistes_{PE} nicht gerecht werden: Denn entweder können sie nur noch den *Einzelfall beschreiben*, oder sie müssen den Naturalismus dadurch preisgeben, dass sie *Gültigkeitsfragen thematisieren*. Im ersten Fall würden sie versuchen, den naturalistischen Anspruch auf eine kausale Erklär- und Vorhersagbarkeit von Verhalten aufrechtzuerhalten, in dem sie die *Zuschreibungen propositionaler Einstellungen*, unter Berücksichtigung je individueller kognitiver Kompetenzen, inferentieller Bewertungen und volitiver Präferenzordnungen, *auf Individuen und Zeitpunkte* (da sich ein individuelles ‚Rechtfertigungsprofil‘ ja ändern kann) *relativieren*. Dann würden sie aber kaum noch das bieten, was man von einer

dergleichen zu haben (...). Nun stellt sich jedoch heraus, dass das gleiche auch für kalkülmäßige (also funktionale / W.-J. C.) Zustände gilt. Und (endliche) Konjunktionen, Disjunktionen usw. aus physikalischen und kalkülmäßigen Zuständen werden auch nichts nützen. Physikalisch und kalkülmäßig gesprochen, gibt es einfach zu viele ‚Entwurfsmodelle‘ für physikalisch mögliche empfindungsfähige Wesen, als dass ein Funktionalismus vom Typ ‚je ein kalkülmäßiger Zustand pro propositionale Einstellung‘ zutreffen könnte.“ (1988/1991), S. 155.

²⁰³ Dies ist nicht dasselbe, wie die oben, im Zusammenhang mit dem Holismusproblem, angesprochene Frage nach einer eindeutigen Abgrenzbarkeit *zentraler* (apriorischer) interner Beziehungen zwischen Einstellungen, von mehr oder minder variablen, denn dort wurde die *allgemeine Gültigkeit* dieser Beziehungen vorausgesetzt.

reduktiven Erklärung *des* Geistes, unter naturalistischen Prämissen, eigentlich erwarten können sollte – nämlich eine *allgemeine Erklärung* dessen, was es heißt, dass physische Zustände oder Ereignisse (über ihre funktionale Rolle definierte) propositionale Einstellungen realisieren. Wenn eine entsprechende ‚Theorie‘, die ja eigentlich verallgemeinernde Aussagen über die funktionale Organisation ‚Intentionaler Systeme‘ liefern sollte, im Extremfall nur auf ein System zu einem bestimmten Zeitpunkt Anwendung finden könnte, genügte sie ohnehin nicht mehr den üblichen Standards wissenschaftlicher (verallgemeinernder) Aussagen und ihrer empirischen Überprüfbarkeit. Und es wäre geradezu absurd, von einer Erklärung eines Verhaltensereignisses durch eine funktionale Theorie reden zu wollen, wenn die theoretischen Terme, die ‚intentionale Zustands-‘ oder ‚Ereignistypen‘ individuieren sollen, nur ein einziges mal instantiiert wären; denn die entsprechenden ‚Gesetze‘ würden ja dementsprechend auch nur für diesen einen Fall gelten. Aber auch ohne die Annahme solcher Extremfälle dürfte es auf der Hand liegen, dass wir es jedenfalls nicht mehr mit einer *allgemeinen* erklärenden Theorie *der* ‚Intentionalität‘ zu tun hätten und ob ein von einer entsprechend eingeschränkten Theorie postulierter ‚intentionaler Zustandstyp‘ jemals einem anderen gleichen würde, wäre empirisch kaum nachzuweisen. Und verstünden sich funktionalistische Theorien deswegen so, dass sie sich dahingehend festlegen müssen, *welche* Verallgemeinerungen hinsichtlich der Rolle einer propositionalen Einstellung als *gültig* (bzw. als wahr) anzusehen sind und welche nicht, müsste der *metatheoretische* Anspruch auf eine *Erklärung des Sinns unserer Praxis* ‚intentionaler‘ Zuschreibungen im Sinne von Postulaten einer empirischen Theorie über die interne funktionale Organisation von Wesen, deren Verhalten *kausal* erklärbar sein soll, aufgegeben werden. Funktionalistische Theorien müssten dann *Stellung* bezüglich dessen *beziehen*, was *rational*, bzw. *vernünftig* ist – und dies erforderlichenfalls auch begründen können. Von einer materialistisch gedeuteten ‚Psychologie‘, als empirischer Wissenschaft der Verhaltensvorhersage oder -Erklärung (im Sinne einer modifizierten ‚Alltagspsychologie‘), könnte dann keine Rede mehr sein. Es würde dann eher naheliegen, ‚intentionale Generalisierungen‘ so zu deuten, dass sie entweder Aussagen darüber darstellen, was von vernünftigen Subjekten jeweils normalerweise zu erwarten wäre oder dass sie eine nachträgliche Rationalisierung von Verhalten ermöglichen, welche verständlich machen kann, um *welche Handlung* es sich gehandelt hat. Und diese Deutung unserer Praxis der Zuschreibung propositionaler Einstellungen halte ich auch für die richtige. Wir werden darauf im Zusammenhang mit interpretationistischen Theorien noch zurückkommen.

Der am meisten beachtete Typ von Einwänden gegen bestimmte Formen naturalistischer Theorien des Geistes_{PE}, vor allem auch gegen den Funktionalismus,

wurde bisher ausgespart. Auch dieser Typ Einwand betrifft keine sozusagen ‚technische‘ Schwierigkeit spezifisch funktionalistischer Theorien, sondern richtet sich gegen grundlegende semantische Verständnisse begrifflichen, bzw. propositionalen Gehaltes (weshalb es auch Sinn macht, ihn gesondert zu diskutieren). Diese Verständnisse sind vor allem durch die Voraussetzung begründet, dass die von einer ‚Intentionalen Psychologie‘ zu postulierenden propositionalen Einstellungen vollständig durch die *internen* Zustände ihrer Träger, bzw. in einem ‚engen‘ Sinne, bestimmt sind. Ich halte die Einwände, die gegen ein ‚enges‘ Verständnis propositionaler Einstellungen (und des Sinns allgemein) vorgebracht wurden, nicht nur für im Kern zutreffend (was nicht bedeutet, dass ich alle der jeweils unterstellten theoretischen Voraussetzungen der Kritiker teile), sondern auch für grundsätzlich bedeutsam. Sie deuten auf Unzulänglichkeiten hin, die, richtig verstanden, zugleich auch eine der grundlegenden Annahmen des IBS-Programms, nämlich die Reduzierbarkeit sprachlicher Bedeutung auf Sprecherintentionen, in Frage stellen. Bevor wir auf diese Einwände eingehen, muss aber der Begriff des ‚engen Gehaltes‘ etwas ausführlicher erläutert werden.

‚Enger Gehalt‘ und die Präsuppositionen empiristischer Theorien der Bedeutung

Mit „engem Gehalt“ ist, kurz gesagt, ein Verständnis der semantischen Eigenschaften propositionaler Einstellungen gemeint, welches sich ergibt, wenn ‚Psychologie‘ im Sinne eines ‚Methodischen Solipsismus‘²⁰⁴ betrieben werden soll. Nach Maßgabe eines ‚Methodischen Solipsismus‘ dürfen *subjekt-externe* Faktoren bei der Individuierung der Gehalte ‚intentionaler Einstellungen‘ keine Rolle spielen. Eine solche Voraussetzung

²⁰⁴ Fodor (1980). Fodor greift diese Bezeichnung von Putnam (ders.: ‚The meaning of „Meaning“‘ (1975)) auf. Der offensichtliche Bezug auf kartesianische oder phänomenologische Strategien, etwa im Sinne von Husserls ‚Welteinklammerung‘, bzw. ‚*epoché*‘, ist natürlich kein Zufall. Allerdings ist der Ausdruck „Methodischer Solipsismus“ für eine ‚intentionale Psychologie‘, als *empirischer* Wissenschaft, eher irreführend, da hier die ‚Welteinklammerung‘ nicht, wie in der *Transzendental-Phänomenologie*, bezüglich der eigenen Perspektive, sondern bezüglich der eines Dritten vollzogen werden soll. Außerdem ist das *Motiv nicht Geltungsfundierung*, sondern *explanative Adäquatheit* der Beschreibung der für die Verhaltensklärung kausal relevanten Faktoren.

Zum Begriff ‚engen Gehaltes‘ („*narrow content*“), siehe etwa: N. Block: ‚Advertisements for a Semantics for Psychology‘ (1986). Theorien dieses Typs werden mitunter auch als ‚internalistisch‘ (im Gegensatz zu ‚externalistisch‘) bezeichnet.

machen allerdings nicht nur Funktionalisten. John Searle etwa – ansonsten ein prominenter Kritiker des Funktionalismus – begründet sie folgendermaßen:

„If I were a brain in a vat I could have exactly the same mental states I have now; it's just that most of them would be false (...). The operation of the brain is causally sufficient for intentionality. It is the operation of the brain, and not the impact of the outside world that matters for the content of our mental states“ (Searle 1980, S. 452)

In einschlägigen Gedankenexperimenten vom ‚Gehirn im Tank‘ (‚*brain in a vat*‘) übernimmt der Wissenschaftler, der alle elektrischen Impulse kontrolliert, die ein in einer Nährlösung schwimmendes Gehirn übers Nervensystem von außen stimulieren können, die zeitgemäßere Fassung der Rolle des kartesischen Dämons. Der Grundgedanke ist allerdings der gleiche: Mein subjektives mentales Erleben würde sich im Fall einer umfassenden Manipulation, bzw. Täuschung durch nichts von demjenigen unterscheiden, welches ich habe (oder hätte), wenn meine Erlebnisse einer äußeren Welt korrespondier(t)en.

‚*Internalisten*‘ ziehen u. a. aus solchen Gehirn-im-Tank-Intuitionen die Konsequenz, dass die ‚intentionale Psychologie‘ ihr Geschäft ohne Rücksicht auf die dem intentionalen Subjekt äußeren Umstände betreiben müsse. Allerdings sind moderne Internalisten keine altbackenen Subjektiven Idealisten – sie sind moderne Materialisten: Zumindest das Gehirn, seine externen ‚Verdrahtungen‘ und der Manipulator müssten demnach physisch reale Dinge sein – nicht *alles* kann nur Vorstellung sein! Und anders, als etwa der an anti-skeptischer Geltungsfundierung interessierten, psychologismuskritisch, bzw. anti-naturalistisch orientierten Phänomenologie Husserls, liegen den internalistischen Theorien des Reduktiven Materialismus metaphysische Erwägungen zugrunde: Zuschreibungen propositionaler Einstellungen sollen ja, nach ‚realistischem‘ Verständnis, *kausalwirksame* Eigenschaften ‚Intentionaler Systeme‘ individuieren. Sie müssen also so verstanden werden, dass sie den tatsächlichen ‚kognitiven‘ Bedingungen genügen, die für das Subjekt entsprechender Zuschreibungen jeweils bestehen. Wenn ein Verhalten *qua* ‚intentionaler Gesetze‘ erklärt werden können soll, kommt es, so der Gedanke, darauf an, wie die Welt aus der Perspektive eines sich verhaltenden Subjektes *erscheint*, nicht wie sie tatsächlich *ist*. Und wenn die ‚intentionalen Eigenschaften‘ mentaler Repräsentationen über ihre *kausale* Rolle in Verhaltenserklärungen bestimmt sein sollen, müssten sie ja vollständig durch die jeweiligen internen physischen Zustände von ‚Intentionalen Systemen‘ festgelegt sein, denn dem ‚Intentionalen‘ kann Kausalwirksamkeit ja (nach der Supervenienzthese) nur *qua* seiner physischen Grundlagen zukommen. So erscheint auch der oben erläuterte Versuch, die Möglichkeit eines reduktiven Übergangs von semantischen zu

syntaktisch-formalen Eigenschaften als eine Implementierung ‚intentionaler Gesetze‘ auf einer computationalen Ebene erläutern zu wollen, überhaupt nur dann – wenigstens *prima facie* – verständlich, wenn der Gehalt ‚intentionaler Zustände‘ in einem ‚engen‘ Sinne verstanden wird. Denn die ‚symbolmanipulierenden‘ Prozesse von Computern sollen ja *formalen* Operationen (nach Algorithmen) entsprechen – Operationen also, die völlig unabhängig davon sind, wie die (‚äußere‘) Welt ist.

Wie aber könnte nun eine *Semantik* aussehen, die eine Beschreibung entsprechender, in einem ‚engen‘ Sinne verstandener, propositionaler Einstellungen ermöglicht? Für eine Auseinandersetzung mit dieser Frage ist es aufschlussreich, sich zunächst eine wichtige *gemeinsame Voraussetzung* zweier sonst eher gegensätzlicher, philosophischer Hauptströmungen des 20. Jahrhundert zu vergegenwärtigen, nämlich der *Phänomenologie* (bzw. der *Hermeneutik*) und des *Neo-Empirismus*. Diese gemeinsame Voraussetzung besteht in der These von der *Indirektheit jedes möglichen Weltbezugs*. Nach dieser These ist der Kontakt zwischen Subjekt und Welt nicht direkt, sondern die Welt ist Subjekten nur über eine Art *epistemisch relevanter Vermittlungsinstanzen* zugänglich²⁰⁵.

Bekanntermaßen ist das *klassische* Motiv von Indirektheitsthesen typischerweise ein *Cartesianisches*. Bedeutung oder Sinn kann danach nicht in einem direkten Bezug auf die Gegenstände der Welt (und ihre Eigenschaften) bestehen, da wir von diesen Dingen *kein gesichertes Wissen* haben. Grundlage dieser Überzeugung ist ein skeptischer Schluss, den man etwa so formulieren könnte: Kann man sich prinzipiell in jedem Einzelfall täuschen, dann kann man sich auch *in allem* prinzipiell täuschen. Was wir sicher wissen können, beträfe dann nur unser direktes ‚inneres‘ Erleben. Unbezweifelbares Wissen könnten wir nur hinsichtlich unserer eigenpsychischen Erlebnisse oder von unserem Bewusstseinsstrom haben. Der Unterschied zwischen tatsächlich bestehenden Sachverhalten und nur eingebildeten, müsste dann offensichtlich *Bewusstseinsintern* verständlich gemacht werden können. (So redet etwa

²⁰⁵ Diese intermediären Instanzen sind es auch, die Rorty ironisch als „Spiegel der Natur“ bezeichnet hat (Rorty (1979/1981)). Für eine aufschlussreiche Ausarbeitung der Gemeinsamkeit der These der Indirektheit des Weltbezugs bei der – von der Phänomenologie beeinflussten – Hermeneutik (vor allem bei Heidegger) und Teilen der (vor allem empiristisch geprägten) sprachanalytischen Tradition, siehe: C. Lafont: *Sprache und Welterschließung* (1994).

Eine weitere Gemeinsamkeit, die zwischen Teilen der empiristischen Tradition und der Phänomenologie (nicht aber der Hermeneutik!) besteht, ist das schon erörterte *Lockesche Sprachmodell*, das durch P. Grice seine moderne Fassung erhielt. Einige Vertreter des Neo-Empirismus wie etwa Moritz Schlick vertraten allerdings, unter dem Einfluss Wittgensteins, eher eine verifikationistische bzw. ‚Gebrauchs-Theorie der Bedeutung.

Husserl davon, dass „Sinnintentionen erfüllt“ werden, wenn ihnen Wirklichkeit zukommt.)

Die aus solchen oder ähnlichen Überlegungen sich ergebenden unangenehmen Konsequenzen waren immer wieder Gegenstand von Auseinandersetzungen auch innerhalb des empiristischen Lagers. Wie sollte man sich die Möglichkeit eines, wenngleich indirekten, Einflusses der ‚äußeren‘ Welt auf ein erkennendes Subjekt überhaupt vorstellen? Ein notorisches Problem empiristischer Varianten der Indirektheitstheorie besteht darin, dass die postulierten intermediären ‚epistemischen Vermittler‘ zwischen Subjekt und Welt (‚einfache Ideen‘, Eindrücke, Sinnesdaten u. ä. m.) – also das ‚unsynthetisierte Rohmaterial der Erkenntnis‘ – einerseits irgendwie mehr sein sollen, als Dinge mit rein physisch-kausal beschreibbaren Eigenschaften, um überhaupt in irgendeinem Sinne *kognitiv relevant* sein zu können. Aber zugleich sollen sie doch *unterhalb* der Ebene von *begrifflich* strukturierten Wahrnehmungen angesiedelt sein, und damit noch *nicht* zum synthetisierten, kognitiv-rationalen Repertoire von Subjekten, bzw. zum *Space of reason* (Sellars), gehören, um so überhaupt als etwas die (um es mit Kant auszudrücken) „Spontanität des Subjektes“ irgendwie ‚von außen her‘ Einschränkendes möglich zu sein²⁰⁶.

Da sich die Indirektheitstheorie bei Empiristen also meist mit einer erkenntnistheoretisch-reduktiven These verbindet, blieb der einzig direkte, also unvermittelte Bezug, den sprachliche Ausdrücke (und dementsprechend auch propositionale Einstellungen) haben können, traditionellerweise den Sinnesdaten (oder auch, wie etwa bei Russell, dem Ich) vorbehalten. Dass alle anderen Subjektausdrücke, die uns ‚oberflächengrammatisch‘ als Namen erscheinen, eigentlich, von ihrer ‚logischen Tiefenstruktur‘ her, als Prädikate, bzw. Beschreibungen, zu verstehen seien, war ja gerade die Pointe von Russells berühmter Theorie der Kennzeichnung. Und Quine geht schließlich so weit, die vollständige Eliminierbarkeit singulärer Termini zu behaupten²⁰⁷.

Darauf werden wir noch zurückkommen. An dieser Stelle geht es mir zunächst darum, den häufig (vor allem im Kontext des Empirismus) hergestellten Zusammenhang zwischen der These von der Indirektheit des Weltbezugs und der These einer *subjektiven Konstitution* von Sinn, Bedeutung und propositionalen Einstellungen hervorzuheben. Ein solcher Zusammenhang scheint besonders dann nahe zu liegen,

²⁰⁶ Unter naturalistischen Vorzeichen übernehmen bei Quine schließlich Nervenreizungen die Rolle ‚epistemischer Vermittler‘, die irgendwie den Übergang zwischen kausalem Naturgeschehen einerseits und theoretischer Naturerklärung andererseits im Sinne einer zwar noch physisch beschreibbaren, aber doch auch schon kognitiv relevanten Ausgangsbasis bilden sollen

²⁰⁷ W. v. Quine: ‚Was es gibt‘ (1948/1979).

wenn mit der These von der Indirektheit die Behauptung einhergeht, dass die tatsächlichen Bezugnahmemöglichkeiten von semantischen Namen an den einfachsten Wahrnehmungs- oder Empfindungsinstanzen von ‚intentionalen‘ Subjekten enden müssen. Wie wir bereits sahen, findet dies, zumindest in manchen Varianten des *Funktionalismus*, eine Entsprechung in den oben genannten Bestimmungen der Form möglicher externer Beschränkungen auf der Inputseite und der These, dass die ‚intentionalen‘ Eigenschaften von Subjekten zu deren *internen* physischen Eigenschaften (ihrer lokalen Mikrostruktur) supervenient sein sollen.

Die These der prinzipiellen Indirektheit jeden Weltbezugs spielt allerdings, wengleich nicht unter reduktiven Vorzeichen, auch in nicht-empiristischen (und nicht-naturalistischen) Theorien der Bedeutung und des Geistes eine prominente Rolle. Es ist Freges Verwendung des Begriffs „Sinn“, als der *Gegebenheitsweise der Referenz*²⁰⁸, der den Ausgangspunkt all jener modernen Theorien sprachlicher Bedeutung bildet, für die *alle* semantischen Namen rein *attributivisch*, bzw. *qua Kennzeichnungen* zu analysieren sind. Nach solchen Theorien besteht die Bedeutung eines Namens bekanntermaßen nicht im Verweis auf einen Träger, bzw. in seinem Bezug, sondern in dem *Bündel von Eigenschaften*, mit denen ein Gegenstand *deskriptiv* individuiert werden kann²⁰⁹.

Häufig nennt man, in Anschluss an Carnap²¹⁰, den Sinn auch die „Intension“, die Referenz dagegen die „Extension“, wobei die Intension dem *Bezug in allen möglichen* Welten, die Extension dem *Bezug in der wirklichen Welt* entsprechen soll. Treffender erscheint es mir, den Fregeschen ‚Sinn‘ als eine *Funktion* zu verstehen, welche den *Bezug* eines semantischen Namens oder die *Proposition* (den Aussagegehalt), die ein Satz oder eine Äußerung zum Ausdruck bringt, *relativ auf Äußerungskontexte* festlegt. Dies lässt sich an einem Beispiel verdeutlichen. Wenn etwa Anna und Max in verschiedenen Situationen jeweils einen extrem indexikalischen Satz, wie „Ich bin jetzt

²⁰⁸ Freges berühmte Beispiel ist „Abendstern“ und der „Morgenstern“, die beide einen verschieden ‚Sinn‘ haben (also etwa: „Der hellste Stern, den man nach Sonnenuntergang sehen kann“ und „Der hellste Stern, den man vor Sonnenaufgang sehen kann“), aber den gleichen Bezug (den Planeten Venus). Nach Frege entspricht der ‚Sinn‘ *semantischer Namen* einerseits einer bestimmten *Weise der Beschreibung ihres Referenten* (ihres Bezugs). Andererseits soll der Sinn von Satzbestandteilen grundsätzlich das sein, was zur Bestimmung des Wahrheitswertes, als dem Bezug von Sätzen, relevant ist. Der ‚Sinn‘ eines *Satzes* lässt sich dann (im Sinne einer sozusagen ‚konsistenzheischenden‘ Lesart von Freges berühmten Aufsatz ‚Der Gedanke‘ (1918) als die ‚Gegebenheitsweise der Proposition‘ (des „Gedankens“), das heißt, desjenigen, dem Wahrheit oder Falschheit zeitlos zukommt, verstehen (obwohl Frege dort den ‚Sinn‘ eines Satzes mitunter mit dem ‚Gedanken‘, den er zum Ausdruck bringt, zu *identifizieren* scheint).

²⁰⁹ Einer der prominentesten Vertreter der ‚Eigenschaftsbündel-‘, oder ‚Cluster-‘ Theorie semantischer Namen ist John Searle; s. Searle: ‚Proper Names‘ (1958) und (1983/1987).

²¹⁰ R. Carnap: *Meaning and Necessity* (1947).

hier“ äußern, dann würden sie in gewisser Weise dasselbe, in gewisser Weise aber verschiedenes sagen. Die Weise, in der sie dasselbe sagen, könnten wir als den *deskriptiven Gehalt* ihrer Äußerung verstehen und diesen mit „Diejenige Person, die eine entsprechende Äußerung macht, befindet sich zu dem Zeitpunkt, zu dem sie sich äußert, an dem Ort, an dem sie sich äußert“ wiedergeben. Sie entspräche dem, was hier mit „Sinn“ gemeint sein soll (wenngleich wir damit möglicherweise von Freges Sprachgebrauch abweichen), bzw. der Intension der genannten Äußerungen. Für diejenige Weise, in der beide etwas Verschiedenes sagen, müssen wir dagegen die wahrheitsrelevante Indexikalität der Äußerungen, bzw. die Referentialität der Ausdrücke berücksichtigen, was bedeutet, dass Anna und Max Aussagen mit unterschiedlichen Wahrheitsbedingungen machen. Dies entspräche der jeweils zum Ausdruck gebrachten Proposition bzw., der Extension der beiden Äußerungen.

Nun ist zwar sowohl die Verwendung des Terms „Sinn“, als auch die des Terms „enger Gehalt“ in der einschlägigen Literatur nicht immer einheitlich. Für ein Verständnis dessen, was, im Hinblick auf eine naturalistische Deutung des Sinnbegriffs, mit „engem Gehalt“ grundsätzlich gemeint ist, kommt es aber vor allem darauf an, dass entsprechende Zuschreibungen von propositionalen Einstellungen Verhaltenserklärungen ja gerade dadurch ermöglichen sollen, dass sie die *kausale Rolle* erfassen, die solche ‚Zustände‘ oder ‚Ereignisse‘ im Verhalten von Subjekten spielen. Solche Zuschreibungen müssten dementsprechend darauf ausgerichtet sein, die relevanten epistemischen, bzw. kognitiven Voraussetzungen der jeweiligen Subjekte im Kontext des zu erklärenden Verhaltens zu berücksichtigen. Es scheint also einzig ein, in Analogie zum Fregeschen „Sinn“ verstandener, durch Subjekte jeweils ‚realisierter‘ oder ‚erfasster Gehalt‘ zu sein, der für die Individuierung *verhaltenserklärender* ‚psychologische Zustände‘ in Frage kommt – und nicht die entsprechende Proposition. Diese Überlegung wird etwa von Perry auch in einer explizit fregeschen Terminologie formuliert²¹¹:

„We use senses (...) to individuate psychological states in explaining and predicting action. It is the sense entertained, and not the thought apprehended, that is tied to human action. When you and I entertain the sense of ‚a bear is about to attack me,‘ we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same sense entertained, same behaviour. When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently. I roll up in a ball, you run to get help.

²¹¹ J. Perry: ‚Frege on Demonstratives‘ (1977). Perry bezieht sich bei seinen Überlegungen vor allem auch auf Kaplans Unterscheidung von „Charakter“, „Gehalt“ und „Extension“, die dieser allerdings im Rahmen sprachphilosophischer Erwägungen entwickelt hat. Siehe etwa D. Kaplan: ‚Bob and Carol and Ted and Alice‘ (1973) und ders.: ‚Dthat‘ (1978).

Same thought apprehended, different sense entertained, different behaviour.“ (S. 494)

Nach Perry ist es der ‚Sinn‘ allein, der für *Verhaltensklärung* (oder -Prognose) gefragt ist. Propositionen ins Spiel zu bringen, entspricht einer „psychologisch unreinen“ Stufe der Interpretation, die Verhalten mit Dingen in der Welt in Beziehung setzt und zu dessen *Verstehen* notwendig ist.

Wie man sich bei diesem Vorschlag das Verhältnis von Verhaltensklärungen und Verhaltensverstehen vorzustellen hat, ist natürlich eine Frage für sich. Jedenfalls kann man das naturalistische Konzept des ‚engen Gehaltes‘ als den Versuch ansehen, den fregeschen Sinnbegriff für eine materialistisch verstandene ‚Psychologie‘ zu nutzen²¹²: ‚Enge Gehalte‘ sollen verhaltensklärende ‚intentionale Zustände‘ (oder ‚Ereignisse‘) individuieren, die sich unabhängig von subjekt- (bzw. system-) externen *Faktoren* beschreiben lassen müssen. Die Semantik, die einer funktionalistischen Beschreibung ‚intentionaler Zustände‘ (oder ‚Ereignisse‘) im ‚engen‘ Sinne entspricht, nennt man üblicherweise „Semantik begrifflicher Rollen“ (SBR).

SBR geht meist mit der Annahme einer Sprache des Geistes einher²¹³, also einer irgendwie ‚syntaktisch‘ realisierten ‚Sprache‘ in den Köpfen von Sprechern, der Bedeutung (bzw. Repräsentation), anders als den Zeichen in verwendeten Sprachen, intrinsisch zukomme. Bei der Deutung der Semantik singulärer Termini orientiert sich SBR an attributivischen Namenstheorien in der sogenannten ‚Frege-Russell-Tradition‘. Die nun im nächsten Abschnitt zu diskutierende, insbesondere durch Keith Donnellan, Saul Kripke und Hilary Putnam vorgebrachte, Kritik attributivischer Namenstheorien²¹⁴, hatte eine überaus lebhafte und bis heute nicht abgeschlossene Debatte zur Folge. Dabei möchte ich zunächst zeigen, dass attributivische Namenstheorien denjenigen Aspekten alltagssprachlicher Bedeutungsverständnisse nicht gerecht werden zu können, die in der einschlägigen Debatte häufig „*direkt-referentiell*“ genannt werden. Übertragen auf die Philosophie des Geistes, also bezogen auf die Frage nach den semantischen Eigenschaften propositionaler Einstellungen wird der Standpunkt der Kritiker dann meist als „*externalistische*“ Position bezeichnet.

Die Argumente externalistischer Positionen werden im Fortgang – an unterschiedlichen Stellen – mehrfach zum Zuge kommen, dabei allerdings in den Kontext einer Deutung gestellt, die sich eher an Überlegungen Strawsons und insbesondere Tugendhats, aber auch, in bereits genannten Hinsichten, an Wittgenstein

²¹² Dass dies keine ganz selbstverständliche Operation ist, ergibt sich schon daraus, dass ein wesentlicher Zug von Freges Philosophie ja bekanntermaßen dessen Psychologismuskritik ist.

²¹³ So z. B. auch bei Block (1986), nicht jedoch bei Loar (1981).

²¹⁴ Die natürlich nicht mit Strawsons berühmter Kritik an Russells wahrheitssemantischer Analyse von Existenzunterstellungen verwechselt werden darf.

und neueren Positionen mit einer grundsätzlich normativ-pragmatistischen Ausrichtung orientiert. Die zunächst zur Debatte stehende Kritik ‚enger‘ Verständnisse von Gehalt wird dabei vor allem darauf hinauslaufen, dass der spezifische Zusammenhang zwischen einem freigeschen Sinn und dem *propositionalen* Gehalt, dem er jeweils perspektiven-, bzw. kontextrelativ entspricht, aus der Zugangsweise eines ‚Methodischen Solipsismus‘ nicht verständlich gemacht werden kann. Meine damit einhergehenden konstruktiven Überlegungen zur *kommunikativen* und *dynamischen* Funktion der Propositionalstruktur sollen aber nicht nur die Unzulänglichkeit der SBR, sondern schließlich *aller* naturalistisch motivierten Versuche einer Erklärung von Sinn (bzw. ‚Gehalt‘) aufweisen.

Theorien direkter Bezugnahme und die externalistische Kritik am Konzept des ‚engen Gehaltes‘

Die vielbeachtete Debatte um die Kritik an Kennzeichnungstheorien semantischer Namen – oder besser: attributivischen Theorien identifizierender Termini – durch bedeutungstheoretische Ansätze, die üblicherweise unter dem *Label* ‚Theorien direkter Referenz‘ subsummiert werden, hatte einen wichtigen Vorläufer in der von Keith Donnellan eingeklagten Unterscheidung *attributivischer* und *referentieller* Verwendungen von Kennzeichnungen²¹⁵. Anhand von einfachen Beispielen, wie der möglichen Frage eines Partygastes; „Wer ist der Mann, der einen Martini trinkt?“, zeigt Donnellan, dass eine Identifizierung auch dann gelingen kann, wenn die mit der Kennzeichnung zugeschriebene Eigenschaft („trinkt einen Martini“) unzutreffend ist, aber referentiell verwendet wird. Wenn etwa der Partygast die genannte Frage einem der anderen Anwesenden stellt, so kann Klarheit bestehen, wer gemeint ist, auch wenn der Gemeinte keinen Martini, sondern einen Sherry trinkt. ‚Eigenschaftsbündeltheorien‘, bzw. reine Kennzeichnungstheorien identifizierender Termini werden dieser Möglichkeit offenbar nicht gerecht. Sie berücksichtigen einseitig die attributivischen Verwendungen, bei denen man etwas (mittels einer Kennzeichnung) *rein deskriptiv identifiziert*, indem man bestimmte *Eigenschaften* aufzählt und dabei denjenigen Gegenstand – falls er existiert – meint, auf welchen sie zutreffen, *welcher immer es sein mag*. So kann ich mit der Äußerung „Der Mörder von Schmitz ist wahnsinnig“ die Person meinen, die Schmitz ermordet hat, wer immer es gewesen sein mag. Ich kann aber auch, in referenzieller Verwendung, eine bestimmte

²¹⁵ K. Donnellan: ‚Reference and Definite Descriptions‘ (1966).

Person meinen, die ich für den Mörder von Schmitz halte. Eine intersubjektive Identifizierung kann dabei gelingen, selbst wenn der Gemeinte nicht tatsächlich der Mörder von Schmitz ist.

Die eigentliche Kern von Donnellans Unterscheidung ist, wie ich meine, folgender: Während Kennzeichnungen in attributivischer Verwendung eine *situationsunabhängige*, aber, wie ich sagen würde, lediglich *virtuelle Identifizierung* allein durch ihren deskriptiven Gehalt ermöglichen, können mit referentiell verwendeten Kennzeichnungen, die von den Verständigungsteilnehmern *für wahr gehalten* werden²¹⁶, Dinge *real*, aber nur *situationsabhängig* identifiziert werden (wobei der deskriptive Gehalt der Kennzeichnung weitgehend – wenn auch nicht vollständig – unzutreffend sein kann). Dies scheint mir an einer ‚oberflächengrammatisch‘ verschleierten Indexikalität der referentiellen *Verwendung* von Kennzeichnungen zu liegen. Entsprechende Identifizierungen bedürfen zwar deskriptiver Gehalte, können aber gleichwohl nur Erfolg haben, weil sie *zugleich* im Sinne eines *direkten Zeigens* auf einen Gegenstand fungieren (etwa im Sinne von „Dieser Mann dort, der den Martini trinkt“).

Donnellan formuliert seine Unterscheidung, anders als mancher der späteren ‚Externalisten‘, eher aus einer pragmatistischen Perspektive auf Möglichkeiten des Gelingens alltäglicher Verständigungsformen, an der auch wir uns später orientieren werden. Von größerem Einfluss auf die Philosophie des Geistes waren aber die Argumente Kripkes, Putnams und Burges.

Saul Kripke hat sich in *Name und Notwendigkeit* (1972/1981) insbesondere mit der Semantik von *Eigennamen* beschäftigt und für diese auf die Wichtigkeit der ‚*Taufsituation*‘ hervorgehoben. Demnach ist es ein wesentlicher Aspekt der Bedeutung von Eigennamen, dass sie jeweils diejenigen Einzeldinge ‚direkt‘ bezeichnen, die auf den Namen in einer bestimmten historischen Situation ‚getauft‘ wurden, bzw. denen der Name in intersubjektiv verbindlicher Form zugeordnet wurde (Kripke sieht sich dabei in der Tradition J. S. Mills). Worauf es für die Bedeutung eines Namens bei späteren Verwendungen ankomme, sei die *kausale Beziehung* zu dem Gegenstand, für den er steht. Diese Beziehung besteht in einer lückenlosen, bis zur Taufsituation zurückreichenden *Kommunikationskette*: Jemand, der etwa den Namen „Aristoteles“ verwendet, hat ihn von jemandem gehört, der ihn wiederum von jemandem gehört hat usw., bis schließlich jemand den Namen von jemand gehört hat, der der Taufsituation beiwohnte. Nach Kripke bezeichnet ein Name, wie „Aristoteles“, einen bestimmten Gegenstand (bzw. eine Person) ‚*starr*‘ (*rigid*), wobei es nicht darauf ankomme, was

²¹⁶ Es mag hier allerdings auch Fälle geben, in denen die intersubjektive Identifizierung selbst dann gelingt, wenn nicht alle oder sogar keiner der Beteiligten die Beschreibung für zutreffend hält.

von dieser Person ausgesagt werden mag, bzw. was auch immer wir über sie herausfinden mögen.

Kripke deutet die ‚Starrheit‘ als metaphysische Notwendigkeit im Sinne einer einzigartigen kausalen Verbindung zwischen der Verwendung des Namens und dem Gegenstand. Doch auch wenn man ihm hierin nicht folgen mag, so hat auch Kripke wichtige Defizite der Kennzeichnungs- und Eigenschaftsbündeltheorien – und damit auch einer jeden Theorie ‚engen Gehaltes‘ – aufgezeigt. Denn für das Verständnis der Semantik und der pragmatischen Funktionsweise von Namen spielen vor allem auch solche Faktoren eine wesentliche Rolle, die dem *Verwender* eines Namens *unbekannt*, bzw. dem Träger einer entsprechenden Überzeugung äußerlich sein können, also als *subjektextern* zu verstehen sind. Eine plausible Theorie der *Semantik von Zuschreibungen propositionaler Einstellungen* muss berücksichtigen, dass die Bedeutung²¹⁷ von Namen über das jeweilige Wissen hinausweisen kann, über das die ‚Objekte mentaler Zuschreibungen‘ (also die Subjekte zugeschriebener Einstellungen), aber auch die Zuschreibenden selbst, im Hinblick auf den Namensträger verfügen. Demzufolge müsste es sich bei Zuschreibungen propositionaler Einstellungen, zumindest in solchen Fällen, in denen Eigennamen eine Rolle spielen, um mehr handeln, als um Beschreibungen rein subjektinterner Zustände oder Ereignisse.

Die Besonderheit der Semantik von Namen im Sinne ihrer ‚Starrheit‘ oder Direktheit der Referenz wurde von Hilary Putnam vom Fall der Eigennamen auf sogenannte ‚Terme für natürliche Arten‘ (*natural kind terms*) ausgeweitet²¹⁸. Putnam entwickelte seine Konzeption von Bedeutung unter anderem als eine Reaktion auf jene relativistischen Konsequenzen, die nach seiner Ansicht aus den quineschen Argumenten für die holistische Natur sprachlicher Bedeutung gezogen werden

²¹⁷ Ich weiche hier von Kripkes Terminologie ab, da für ihn Namen überhaupt keine ‚Bedeutung‘ (Konnotation), sondern nur eine Denotation haben. Es scheint überzogen, aufgrund der angeführten Argumente behaupten zu wollen, ein Name habe überhaupt keine deskriptive Bedeutung (bzw. keinen Sinn), sondern nur eine kausal bedingte Referenz. Denn *irgendwelche* Eigenschaften müssen wir für den Träger eines Namens sicherlich konstant halten, um ihn, im Falle von Revisionen unseres Wissens über ihn, als dasjenige, was mit dem entsprechenden Namen getauft wurde, *reidentifizieren* zu können. Der Träger des Namens ‚Aristoteles‘ muss immerhin Eigenschaften gehabt haben, die es für uns retrospektiv, im Falle auch tiefgreifender Revisionen unseres Wissens über diesen Träger, verständlich machen würden, warum ihm fälschlicherweise bestimmte Eigenschaften zugeschrieben wurden. Wir könnten sicherlich unter keinen Umständen zu der Einsicht gelangen, dass ‚Aristoteles‘ beispielsweise eigentlich auf eine bestimmte Haarnadel referiert. Zur Bedeutung des Namens ‚Aristoteles‘ gehört also wohl immerhin der deskriptive Gehalt des Sortals ‚Mensch‘ (oder doch wenigstens ‚menschenähnlich‘, falls jemand hier das notorische Beispiel der vom Mars ferngesteuerten perfekten Roboter anführen wollte). Vergl. hierzu auch Searle (1983/1987), Kap. 9.

²¹⁸ Putnam (1975). Putnam hat seine Position allerdings später modifiziert; siehe v. a. ders.: (1988/1991) und ders.: ‚Meaning Holism‘ (1986).

müssten, wenn sich Bedeutung vollständig in fregeschem Sinn erschöpfen würde. Nach einer gängigen Interpretation Quines soll ja die Bedeutung *aller* Terme einer Theorie (dies heißt bei Quine zugleich: einer Sprache) von der Theorie (der Sprache) als Ganzer abhängig sein. Damit werden wir uns im letzten Teil noch im Einzelnen beschäftigen. Jedenfalls ist es nach Quine möglich, bei einem Scheitern der Voraussagen einer Theorie „vor dem Tribunal der Erfahrung“ sowohl die zentralen als auch die peripheren Aussagen der Theorie/Sprache in Frage zu stellen. Aber wie auch immer man sich – womöglich mit guten Gründen – entscheidet, die jeweils vorzunehmende Änderungen der Verifikationsbedingungen einzelner Sätze würde dann offenbar auch die Verifikationsbedingungen (und damit die Bedeutung) anderer, im Extremfall *aller* Sätze (bzw. Terme) einer Theorie in den Strudel des Bedeutungswandels hineinziehen.

Die Probleme, die ein solcher semantischer Holismus für Versuchen einer Definition propositionaler Einstellungen durch funktionale Rollen mit sich bringt, hatten wir ja schon thematisiert. Aus erkenntnis- oder wissenschaftstheoretischer Perspektive werfen holistische Positionen²¹⁹ aber außerdem die Frage auf, wie wir überhaupt durch neue Erkenntnisse bzw. durch ein Scheitern von Theorien *qua* Falsifikation ihrer empirischen Voraussagen etwas lernen könnten, wenn sich mit jeder neuen Erkenntnis (mit jedem wahren, falsifizierenden Satz) die Bedeutung *aller* Terme einer Theorie/Sprache ändern würden. Die Bedeutungsänderung beträfe ja sowohl die Peripherie, also diejenigen Aussagen, mit denen die empirischen Daten im Sinne wahrnehmbarer Eigenschaften beschrieben werden können sollen, als auch jene zentralen (bzw. abstrakten) Begriffe, die vor allem in den von der Theorie behaupteten Gesetzen eine Rolle spielen. Wie aber könnte dann noch davon die Rede sein, dass die *gleichen* Phänomene oder Ereignisse durch eine neue Theorie besser erklärt werden, wenn also Wissenschaftler nach einem ‚Theoriewechsel‘ gleichsam – wie etwa Kuhn sich ausgedrückt – ‚in einer anderen Welt leben‘? Worin könnten dann noch die für inter-theoretische Vergleiche notwendigen Identitätskriterien für zu erklärende Phänomene bestehen?

Um zumindest dem Problem einer bereits theoriegetränkten Bestätigungsbasis von Theorien zu entgehen, hatte Quine in *Word and Objekt* das Konzept der ‚Reizbedeutung‘ (*stimulus meaning*), als neutralem Fundament der Bestätigung unterschiedlicher Theorien eingeführt. Doch wie wir oben schon bemerkten, führt offensichtlich kein nachvollziehbarer Weg von physisch beschriebenen Reizen zu Überzeugungen, bzw. Aussagen²²⁰. Außerdem bliebe das Problem bezüglich der

²¹⁹ Zu denen ja beispielsweise auch die T. Kuhns zählt.

²²⁰ Und wenn Quine Überzeugungen und Aussagen von vornherein naturalistisch (bzw. behavioristisch), also beispielsweise lediglich als Dispositionen zur Produktion bestimmter Geräuschfolgen, verstehen würde, hätte er damit natürlich seine eigene Problemstellung

übrigen, zentraleren Begriffe einer Sprache weiter bestehen. Putnam schlägt nun vor, dass man diesen Schwierigkeiten entgeht, wenn man Quines Bedeutungskeptizismus nicht mitmacht und zwischen einem Holismus hinsichtlich der Bildung von Überzeugungen und einem Holismus hinsichtlich Bedeutungen unterscheidet. Während er Quine beim ersteren zustimmt, lehnt er letzteren ab. Es geht Putnam darum, „(...) to concede that Quine’s holistic account of belief fixation is correct, while maintaining that an ontology of ‚meanings‘ is still scientifically or philosophically necessary and useful“²²¹. Putnam hat deshalb in ‚Die Bedeutung von „Bedeutung“‘ eine Konzeption des Bedeutungsbegriffs entwickelt, die einen theorie-, bzw. sprachen-bedingten Relativismus vermeiden und *Theorienwandel als Wissensfortschritt* deutbar machen soll, indem sie – jedenfalls im Hinblick auf Terme natürlicher Arten, wie etwa „Wasser“ – die Möglichkeit einer *theorie- bzw. überzeugungs-übergreifenden Identität des sprachlichen Bezugs* eröffnet. So ist es nach Putnam für die Bedeutung des Terms „Wasser“ wesentlich, auf den *Stoff* Bezug zu nehmen, den wir mit dem Ausdruck „Wasser“ *paradigmatischerweise* in Verbindung bringen. Dieser Stoff hat, nach unserer jetzigen Kenntnis, die chemische Mikrostruktur H₂O. Doch was auch immer die Mikrostruktur des Stoffes „Wasser“ am Ende sein mag; es sind die gesetzesartigen Eigenschaften der Mikrostruktur *desjenigen Stoffes* (der natürlichen Art), auf den wir in der Mehrzahl unserer *paradigmatischen Beispiele für diesen Stoff* referieren, die diejenige natürliche Art bestimmen, die wir mit „Wasser“ meinen²²².

Putnam bedient sich eines berühmt gewordenen Gedankenexperimentes, um die Defizienz derjenigen Konzeptionen von Bedeutung aufzuzeigen, nach denen der ‚Sinn‘ *per se* den Bezug bestimmt. Nehmen wir einmal an, dass es auf einem anderen, der Erde sonst in allen Hinsichten gleichenden Planeten (nennen wir ihn „Zw(illings)erde“) einen Stoff gibt, den die Einheimischen „Wasser“ nennen und der, obwohl mikrostrukturell anders beschaffen als ‚unser‘ Wasser, kurioserweise alle diejenigen von dessen Eigenschaften teilt, die vor der Entdeckung seiner chemischen Mikrostruktur auf der Erde als für eine korrekte Klassifizierung eines Stoffes als

unterlaufen. Auf diesen Problemkomplex werde ich später, im Zusammenhang mit der Position Davidsons, noch zurückkommen.

²²¹ Putnam (1986), S. 280. Damit hat Putnam der Bedeutung natürlich ein gewisses Maß an Autonomie gegenüber propositionalen Einstellungen eingeräumt – was mit ‚intentionalistischen Semantiken‘ unvereinbar ist.

²²² Diese Darstellung entspricht der modifizierten Variante von Putnams Position (s. Putnam (1986)). In ‚Die Bedeutung von „Bedeutung“‘ war Putnam offenbar (ebenso wie Kripke) noch der Ansicht, dass unseren Begriffen für natürliche Arten die extensions-bestimmenden Eigenschaften (also bei Wasser, die chemische Mikrostruktur H₂O) *logisch notwendig* zukommen. Heute dagegen vertritt er die Position, dass wir prinzipiell nicht über (interessen-) unabhängige Kriterien für Substanz- oder Stoffidentität im Sinne notwendiger und hinreichender Bedingungen verfügen.

„Wasser“ maßgeblich galten. Gemäß den Voraussetzungen des Gedankenexperimentes sollen die Einheimischen von Zwerde physische Doppelgänger der menschlichen Erdbewohner sein, uns also in physischer Hinsicht vollkommen gleichen²²³. Die Vertreter der Auffassung, nach der der Sinn den Bezug vollständig bestimmt, bzw. Anhänger von Theorien ‚engen Gehaltes‘, wären hier gezwungen, sowohl sprachliche Bedeutungs- als auch Überzeugungsgleichheit zwischen den Bewohnern von Erde und Zwerde anzunehmen. Dies erscheint jedoch kontraintuitiv: Die Bedeutungen oder Überzeugungen, die wir unseren ‚Zwillingen‘ auf Zwerde im Zusammenhang mit Wasser zuschreiben müssten, entsprächen doch gerade nicht denen auf der Erde. Wo sich unser deutsches Wort „Wasser“, bzw. eine entsprechende Überzeugung, auf Wasser bezieht, würde der ‚gleiche sprachliche Symboltyp‘, bzw. der ‚gleiche Typ mentaler Repräsentationen‘²²⁴ von jenem Stoff auf Zwerde mit der Mikrostruktur XYZ handeln. Vereinfacht ausgedrückt, ist die Bedeutung des Ausdrucks „Wasser“ auf der Erde Wasser, auf Zwerde hingegen wäre sie Zwasser.

Putnams These geht damit über die von Funktionalisten eingeräumte Möglichkeit, dass in physischer Hinsicht unterschiedliche Subjekte (bzw. Systeme) gleiche Überzeugungen haben können hinaus. Denn sein Gedankenexperiment soll ja umgekehrt auch zeigen, dass auch in physischer Hinsicht *gleiche* Subjekte unterschiedliche Überzeugungen haben können. Und dies gilt, wie Putnam erst später klar geworden ist, nicht nur im Hinblick auf die kruderen Formen des Materialismus, sondern auch für den Funktionalismus, der das Kriterium für physische Gleichheit im

²²³ Putnams Gedankenexperiment hat, wie mitunter schon bemerkt wurde, gewisse Schwächen. Denn erstens müssten auf der ‚Zwillingserde‘ zumindest einige chemische und physikalische Gesetze anders sein, damit ein Stoff mit der Molekulareigenschaft XYZ die gleichen ‚höherstufigen‘ Eigenschaften des Wassers haben könnte. Und zweitens könnten unsere physischen Zwillinge schon aus dem Grund keine exakten Duplikate sein, weil sie nicht wie wir zu ca. 70 % aus Wasser im Sinn von H₂O, sondern im Sinne von XYZ, bestehen würden. Allerdings könnten sie uns in funktionaler Hinsicht gleichen, und nur darauf kommt es im vorliegenden Zusammenhang ja tatsächlich an.

²²⁴ Die Rede von ‚Typen von Symbolen‘ bzw. ‚mentalenen Repräsentationen‘ ist hier ziemlich lax und eigentlich unzulässig, da die Beschreibung des Gedankenexperimentes lediglich die *physische Gleichheit* der (auch subjektinternen) Gegebenheiten auf Erde und Zwerde postuliert, während es Bedeutungsidentitäten gerade bestreitet. Man könnte höchstens von gleichen physischen Typen, also etwa gleichen Geräuschfolgen, sprechen. Es sind jedoch, wie wir oben schon sahen, nicht physische Realisierungen symbolischer Token (ob sprachlich oder mental), die eine Bedeutung, bzw. semantische Eigenschaften haben können, sondern *Symbole*. Wenn nun aber, wie ich behaupte, sprachliche Symbole *immer auch* – z.B. im Sinne der Regeln ihres korrekten Gebrauchs – *durch ihre semantischen* und *nicht nur durch ihre physischen Eigenschaften individuiert* werden, könnte im vorliegenden Fall, da die Bedeutung laut Voraussetzung ja abweicht, von ‚gleichen Typen von Symbolen‘ oder ‚gleichen Typen mentaler Repräsentationen‘ eigentlich keine Rede sein.

oben erläuterten Sinne abschwächt und nur noch von einer über funktionale Rollen bestimmten Gleichheit im Sinne ‚engen Gehaltes‘ spricht²²⁵.

Tyler Burge hat die wohl weitreichendste Variante externalistischer Kritik an einer ‚Psychologie enger Gehalte‘ vorgetragen²²⁶. Seine Einwände betreffen nämlich nicht nur Begriffe natürlicher Arten. Und war Putnams externalistisches Motto ‚*meanings are not in the head*‘ ursprünglich eher eine wahrheitssemantische Konsequenz seiner Kritik attributivischer Namenstheorien, so betont Burge von vorneherein die *anti-individualistische, soziale* Ausrichtung seines Ansatzes. Entschiedener noch als Putnam, für den Experten (durch ihr Wissen um die mikrostrukturellen oder genetischen Eigenschaften des Referenten) die Rolle von Autoritäten hinsichtlich der Bedeutung von Begriffen für natürliche Arten spielen, bringt Burge den *Sprachgebrauch einer Gemeinschaft* von Sprechern als für die Bedeutung von Ausdrücken, also auch für Zuschreibungen propositionaler Einstellungen, wesentlich ins Spiel.

Eine wichtige Voraussetzung von Burges Argument ist die eigentlich naheliegende Annahme, dass sich *auch als kompetent geltende Sprecher* bezüglich der korrekten Verwendung, bzw. der Bedeutung einzelner Wörter *irren können*²²⁷. Burges bekanntes Beispiel ist ein Fall, bei dem sich jemand hinsichtlich der Bedeutung, bzw. der korrekten Verwendung des Begriffs „Arthritis“ irrt. Üblicherweise bezeichnet „Arthritis“ (im Deutschen) eine Krankheit, die mit Entzündungen der Gelenke einhergeht. Dies ist ein wesentlicher Aspekt der Bedeutung dieses Wortes, egal, ob jemand, der es verwendet, dies weiß oder nicht. Entsprechendes gilt für Überzeugungen, die mit Arthritis zu tun haben. Wenn jemand davon überzeugt wäre, er habe Arthritis im Oberschenkel, weil er meint, Arthritis sei eine Entzündung der Knochen, so könnten wir durchaus sagen, dass er *glaube, Arthritis* im Oberschenkel zu haben, weil er fälschlicherweise meint, Arthritis sei eine Entzündung der Knochen (anstatt der Gelenke). Dies ist aber überhaupt nur deshalb möglich, weil der Standard für das, was wir mit „Arthritis“ meinen (bzw. was der Gehalt einer Überzeugung ist,

²²⁵ Vergl. Putnam (1988/1991).

²²⁶ T. Burge: ‚Individualism and the Mental‘ (1979). Allerdings laufen auch schon Kripkes Überlegungen in ‚A Puzzle about Belief‘ (ders.: (1976)) in der Konsequenz in eine ganz ähnliche Richtung.

²²⁷ Dies sollte jedenfalls für nicht-einfache Begriffe, wie „Arthritis“ gelten. Begriffe, wie etwa „oben“ oder „gestern“ sollte jedoch jeder Sprecher, der als kompetent gilt, in normalen Verwendungssituationen irrtumsfrei beherrschen. Damit wird allerdings *nicht* behauptet, dass es eine *klare*, kriterial bestimmbare Grenze zwischen *irrtümlichen* Sprachverwendungen und *abweichendem* Sprachgebrauch gibt. Normalerweise reagieren kompetente Sprecher aber, in Fällen des Irrtums, auf einen entsprechenden Hinweis – im Gegensatz zu ‚Abweichlern‘ – mit *Selbstkorrektur*.

die mit Arthritis zu tun hat), durch den allgemeinen Sprachgebrauch bestimmt wird – und nicht durch individuelle Sprecherüberzeugungen (bzw. einen ‚engen Gehalt‘).

Wir können das putnamsche Gedankenexperiment mit in physischer Hinsicht gleichen Zwillingen in Burges Sinne variieren. Dazu müssen wir uns einen kompetenten Sprecher vorstellen, der zu einer Sprachgemeinschaft gehört, in der der Ausdruck „Arthritis“ sowohl für eine Entzündung der Gelenke, *als auch* der Knochen steht. Wir könnten diese Krankheit in *unserer* Sprache „Zarthritis“ nennen. Ein Mitglied jener anderen Sprachgemeinschaft, der behauptete: „Ich habe Arthritis im Oberschenkel“ und dies, anders als bei „Arthritis“ in unserem Sprachgebrauch, deshalb *berechtigterweise* für möglich hält, weil er glaubt, dass Zarthritis eine Entzündung der Gelenke *und* der Knochen sei, befände sich nicht *zwangsläufig* im Irrtum. Wenn dagegen ein Mitglied der deutschen Sprachgemeinschaft glaubte oder behauptete, er habe Arthritis im Oberschenkel, wüssten wir, dass er mit seiner Diagnose falsch liegen muss. Und dieser Unterschied besteht, obwohl sich der irrende Deutschsprecher und sein nicht irrendes *Zwillings-Pendant* hinsichtlich ihrer ‚eng‘ verstandenen Überzeugungen und der relevanten physischen Eigenschaften vollkommen gleichen mögen (also sowohl was ihre neuronalen Zustände, bzw. ihre interne funktionale Organisation betrifft, als auch im Hinblick auf die physischen Aspekte ihrer Äußerungen). Eine ‚methodisch-solipsistisch‘ orientierte Psychologie kann den erläuterten Unterschied des *propositionalen* Gehaltes, bzw. den Unterschied im Wahrheitswert, einfach nicht erfassen. Burges Beispiel zeigt, dass der Gehalt von Zuschreibungen propositionaler Einstellungen vor allem auch davon abhängt, welche Bedeutung die Wörter, mit denen die Objekt der Zuschreibung ihre Einstellung selbst zum Ausdruck bringen würden, in der relevanten Sprachgemeinschaft haben. Und dies betreffe offenbar nicht nur Terme für natürliche Arten.

Als den für eine Kritik internalistischer Varianten naturalistischer Theorien der Bedeutung und des Geistes ausschlaggebenden, gemeinsamen Gedanken aller hier genannten Einwände, von Donellan bis Burge, können wir jedenfalls festhalten, dass *wesentliche Aspekte des propositionalen Gehaltes von entsprechenden Einstellungen oder der Bedeutung von Äußerungen nicht durch die individuellen Überzeugungen ihrer Subjekte festgelegt sind*. Zumindest all diejenigen ‚intentionale‘ Einstellungen individuierenden Gehalte, die *Namen* enthalten, sind nicht in erster Linie durch die Überzeugungen ihrer Subjekte, sondern vor allem durch ihre *Wahrheitsbedingungen* bestimmt. Unterschiede gibt es nun aber insbesondere in der Betonung desjenigen Aspektes, der für eine Bestimmung der Wahrheitsbedingungen wesentlich hinzukommen soll. Während bei Kripke der historische Kontext im Vordergrund steht,

ist es bei Donellan der *situative*. In Putnams Überlegungen wiederum kommt, wenn ich recht sehe, beides zum Zuge.

Darüber hinaus legen Burges und Putnams Argumente wie ich meine auch nahe, dass es für unser Verständnis von Bedeutung konstitutiv ist, eine durch einen allgemeinen Sprachgebrauch instituierten Normativität zu unterstellen, die wir *Standardbedeutung* nennen können²²⁸. Dies aber stellt nicht nur Theorien ‚engen Gehaltes‘ und das IBS-Projekt in Frage, sondern alle Theorien, die Bedeutung eng an Sprecherintentionen koppeln. Dem wird ein bedeutungstheoretischer ‚Intentionalist‘ wohl entgegenhalten wollen, dass die Voraussetzungen des Burgeschen‘ Argumentes falsch seien. Wir hätten, bezogen auf sein Beispiel, eben gar keine Veranlassung einer Person überhaupt Überzeugungen zuzuschreiben, die von Arthritis handeln, wenn wir merken, dass sie den Ausdruck „Arthritis“ anders verwendet, als wir (oder andere), nämlich auf der Grundlage von Überzeugungen, die nicht mit unseren (oder denen anderer) übereinstimmen. Es sei eben die jeweilige Verwendung eines Ausdrucks auf der Grundlage der intentionalen Einstellungen eines Sprechers, die für dessen Bedeutung maßgeblich ist, nicht eine vermeintliche Standardbedeutung.

Wir werden auf die Frage des sozialen Charakters sprachlicher Bedeutung, insbesondere im Zusammenhang mit der Diskussion der Position Davidsons, noch ausführlich eingehen. An dieser Stelle möchte ich nur bemerken, dass ein solcher Einwand jedenfalls wichtige praktische bzw. koordinatorische Funktionen unterschlägt, welche die wechselseitige Unterstellung einer sozial etablierten Standardbedeutung für unsere soziale Praxis hat – oder doch zumindest häufig haben kann. Ein anschauliches Beispiel hierfür liefert eine einfache Anreicherung des Burgeschen‘ Beispiels, bei der wir einen durchaus alltäglich möglichen Fall annehmen: Jemand sucht wegen anhaltender Schmerzen seinen Hausarzt auf. Der diagnostiziert Arthritis, verschreibt ein Medikament und trägt seinem Patienten auf, sich dieses Medikament in der Apotheke zu besorgen. Der Arthritiskranke macht sich nun in dem Glauben, Arthritis sei eine Knochenentzündung, auf den Weg. In der Apotheke stellt er fest, dass er das Rezept verloren hat. Zu allem Unglück ist es Freitagabend und die Arztpraxis hat inzwischen geschlossen. Falls der Apotheker seinem Kunden traut, wird er diesem

²²⁸ Die normative Rolle der Gemeinschaft hinsichtlich Fragen der Bedeutung hat allerdings bekanntermaßen auch Kripke später im Zusammenhang mit einer skeptischen Interpretation von Wittgensteins sogenannten ‚Regelfolgeproblem‘ betont; s.: ders. (1982/1987). Anders als Kripke, halte ich es jedoch – aus bereits genannten Gründen – nicht für sehr glücklich, den sozialen und normativen Aspekt von Bedeutung so darzustellen, als bestimme die Gemeinschaft (mehrheitlich?), was die Bedeutung eines Ausdrucks ist. Meine Behauptung ist, dass zu unserem Bedeutungsverständnis wesentlich die allgemeine *Unterstellung* einer als gemeinsam verbindlichen Standardbedeutung gehört. Bezüglich der Frage, *wie* ein Ausdruck gemäß der unterstellten Standardbedeutung *vernünftigerweise* verwendet werden *sollte*, könnte theoretisch auch die Mehrheit der Mitglieder einer Sprachgemeinschaft irren.

Glauben schenken, wenn der sagt, sein Arzt habe ihm ein Mittel gegen Arthritis verschrieben und, um ihm ein unangenehmes Wochenende zu ersparen, das Mittel geben. Für die Wahl des richtigen Medikamentes, welches die Beschwerden des arthritiskranken Kunden lindern soll, kommt es offenbar nicht darauf an, dass der Apotheker berücksichtigt, was das Wort „Arthritis“ *im Munde des Sprechers* bedeuten könnte, in dem er herauszubekommen versucht, was der über Arthritis *glaubt* (und deswegen etwa eine empirische – beispielsweise funktionale – Theorie über dessen ‚intentionale Zustände‘ aufzustellen versucht), sondern darauf, dass sowohl der verschreibende *Arzt*, als auch der *Apotheker* der *richtigen Verwendung* des Ausdrucks „Arthritis“ *mächtig* sind und sie zudem *wechselseitig davon ausgehen, dass dies so ist*²²⁹. Dieses Beispiel lässt sich nicht nur auf unterschiedliche Fälle hin variieren, es vermeidet zudem problematische Zwillingswelt-Fiktionen.

Ein wenig vorgreifend, möchte ich als ein wesentliches Ergebnis der externalistischen Kritik insbesondere auch den Verweis auf die Wichtigkeit von *intersubjektiven Situationen direkter Deixis* verstanden wissen. In wechselseitiger Abhängigkeit mit dem Konzept der *Wiederholbarkeit* bilden solche Situationen nämlich nach meinem Dafürhalten die notwendige *Grundlage* jeder Möglichkeit von Identifikation und Reidentifikation, bzw. jeden Begriffs von Ding- oder Stoffidentität und damit zugleich von Eigenschaftsgleichheit – was bedeutet, dass sie eine Bedingung der Möglichkeit propositionalen Gehaltes sind. Darauf werde ich noch zurückkommen. Jedenfalls ist es für unser Verständnis von Namen für Dinge oder Stoffe wesentlich, dass es nicht nur auf die Kenntnis der Eigenschaften ankommt, von denen man jeweils gerade annimmt, dass sie diesen zukommen, sondern auch darauf, Dinge oder Stoffe auch dann als identisch bzw. gleich identifizieren und reidentifizieren zu können, wenn wir uns über ihre tatsächlichen Eigenschaften weitgehend im Irrtum befinden.

Als Reaktion auf die vorstehenden Kritiken an einer Konzeption von Gehalt als vollständig durch Intensionen bestimmt haben unter anderem Colin McGinn und Jerry Fodor²³⁰ vorgeschlagen, propositionale Einstellungen *sowohl* im ‚engen‘ als auch im

²²⁹ Dieses Beispiel wurde durch eine ganz ähnliche Rekonstruktion von Burges Argument bei C. Gauker angeregt (in: ders. (1994), Kap. 3).

Es wäre übrigens auch unangemessen, auf die vorstehende Überlegung in der Weise zu reagieren, dass man sagt, der Apotheker müsse eben nicht die Überzeugungen des Kunden, sondern die des Arztes berücksichtigen. Er mag über den Arzt nämlich gar nichts weiter wissen. Der Apotheker kann höchsten allgemein davon ausgehen, dass ein Arzt die richtigen wesentlichen Überzeugungen über Arthritis hat und die Standardbedeutung von „Arthritis“ kennt.

²³⁰ C. McGinn (1982); Fodor (1987)

‚weiten‘ Sinne zu individuieren, nämlich als zwei sich ergänzende Aspekte („*Dual Aspekt Theories*“). Eine wissenschaftliche Psychologie als Theorie der *Verhaltensvorhersage und -Erklärung* müsse es zwar, so Fodor, mit engen Gehalten zu tun haben, da sich die Klassifizierungen der Wissenschaften ausschließlich an kausalen Eigenschaften orientierten. Eine zugleich an Extensionalität, bzw. Wahrheitsbedingungen orientierte Individuierung der Gehalte könnte aber darüber hinaus, ähnlich wie bei Perrys Vorschlag, den alltagspsychologischen Intuitionen gerecht werden. Beidem zu genügen, sei kein Widerspruch: Zwillingswelten-Beispiele zerschnitten nicht den Zusammenhang zwischen engem Gehalt und Extension; sie relativierten ihn lediglich auf Kontexte:

„(...) My Twin’s ‚water‘-thoughts are intentionally identical to my water-thoughts; they have the same contents even though, since their contexts are de facto different, they differ, de facto, in their truth conditions. In effect what we have here, is an extensional criterion for ‚narrow‘ content. The ‚broad content‘ of a thought, by contrast, is what you can semantically evaluate; it’s what you get, when you specify a narrow content *and fix a context*.“ (a. a. O., S. 48)

Nach Fodors dualistischer Version besteht die entscheidende Aufgabe einer materialistischen Theorie des Geistes dann darin, in einer ‚nicht-intentionalen‘, nicht-semanticen Begrifflichkeit sagen zu können, worin die Interpretation der primitiven Symbole des ‚Mentalesischen‘ relativ zu einem Kontext bestünde. Der Reduktive Materialist müsste also versuchen, *subjektexterne* Faktoren zu finden, anhand derer jene Relativierung ‚enger Gehalte‘ auf Kontexte möglich wird, und zwar solche, die in irgendeiner Weise *kausal spezifizierbar* sind: Denn ‚weiter Gehalt‘, bzw. die Bedingungen semantischer Evaluierbarkeit ‚engen Gehaltes‘ im Hinblick auf einen Kontext, müssen ja ihrerseits naturalistisch reformuliert werden können²³¹.

Im Rahmen eines im weiteren Sinne funktionalistischen Vokabulars liegt es nahe, solche dualistischen Ansätze so anzulegen, dass die ‚engen‘ Eigenschaften ‚intentionaler Zustände‘ oder ‚Ereignisse‘ ihre *efferenten kausalen Rollen* hinsichtlich der unmittelbaren Verhaltenssteuerung betreffen, ihre ‚weiten‘ Eigenschaften dagegen ihre *afferenten kausalen Korrelationen* mit äußeren Umständen. Welcher Zusammenhang besteht dann aber zwischen diesen? Und was heißt es eigentlich, einen Kontext zu bestimmen? – Vor allem aber: Wie kann ‚weiter Gehalt‘, bzw. die Wahr- oder Falschheit einer propositionalen Einstellung, überhaupt eine *kausale Rolle* für das

²³¹ Die Alternative zu einer solchen ‚Externalisierung‘ naturalistischer Verständnisse von Gehalten, die der Eliminative Materialismus vorschlägt, besteht natürlich darin, die Relevanz der Semantik, bzw. der Alltagsintuitionen einer Zuschreibbarkeit propositionaler Einstellungen als für eine Individuierung möglicher kausalwirksamer Typen einer wissenschaftliche ‚Psychologie‘ der *Verhaltensvorhersage* irrelevant abzulehnen.

Verhalten desjenigen Individuums spielen, welches Gegenstand einer entsprechenden Zuschreibung ist?

Wir werden im nächsten Kapitel Theorien diskutieren, die versuchen, auf diese Fragen zu antworten – und zwar unabhängig davon, ob sie sich als ‚*dual-aspect-theories*‘ verstehen (lassen) oder nicht. Ich konzentriere mich dabei auf den sogenannten ‚*teleofunktionalistischen*‘ Ansatz, und zwar in Gestalt der zwei wohl am meisten beachteten seiner Vertreter, nämlich Dretske und Millikan²³². Beide wollen die gesuchten afferenten Kausalrelationen grundsätzlich in der Weise bestimmen, dass es die *Funktion* eines kognitiven Zustandes oder Ereignisses ist, die *externen Umstände so anzuzeigen*, dass das *richtige Verhalten* daraus folgt. Meine Kritik wird auch hier auf die These hinauslaufen, dass die Schwierigkeiten, in die – wie wir sehen werden – auch solche Theorien mit ihrem Versuch einer ‚*naturalistischen Externalisierung*‘ der Bestimmung von Gehalt geraten, einem *Kategorienfehler* geschuldet sind. Und ich möchte bereits so viel verraten, dass dieser Kategorienfehler, kurz gesagt, darin besteht, den *Wahrheitsbegriff adjektivisch*, als für eine (besondere) Eigenschaft von Zuständen oder Ereignissen stehend, miss zu verstehen.

5 Teleofunktionaler Externalismus: Dretske und Millikan

Die unter naturalistischen Prämissen naheliegende Antwort auf die vorstehend diskutierten Schwierigkeiten, ‚*intentionale Zustände*‘, bzw. den Gehalt von Überzeugungen in einem ‚engen‘ Sinne verständlich zu machen, besteht in der Idee der *kausalen Kovarianz* zwischen Token von Überzeugungen und den *äußeren* Umständen, die sie repräsentieren. Sie wurde zunächst durch D. Stampe (1977) in die neuere Diskussion eingebracht. Ihre prominentesten Vertreter innerhalb des reduktiv-materialistischen Lagers sind Dretske, Fodor, Millikan und A. Goldman.

²³² Fodor selbst hat, nachdem er zunächst mit dem Teleofunktionalismus geliebäugelt hatte (ders.: ‚*Semantics, Wisconsin Style*‘ (1984)), einen eigenen, nicht-teleologischen, bzw. nicht-evolutionistischen Ansatz vertreten (ders. (1990)). Zu Fodors Vorschlag und den Problemen, mit denen er sich konfrontiert sieht, siehe etwa Beckermann (1999), S. 337 f.

Sogenannte ‚kausale Theorien des Gehaltes‘ behaupten, vereinfachend ausgedrückt, dass mentale Repräsentationen (zuverlässiger Weise) durch diejenigen „semantisch relevanten Situationen“ verursacht werden, die sie repräsentieren²³³. Solche Theorien würden, so scheint es, unter naturalistischen Bedingungen mit der Kritik an engen Verständnissen von Gehalt und Bedeutung fertig, wenn sie Instantiierungen von Überzeugungen (etwa von Symbolen des ‚Mentalesischen‘) im Sinne einer *kausalen Kovarianz von Gehalten und externen Ursachen* unter der Bedingung ihrer kausalen Relevanz für Verhalten verständlich machen könnten. Jedenfalls wären etwa die jeweiligen, nur im ‚engen‘ Sinne gleichen, ‚Wasser-Überzeugungen‘ von Erdbewohnern und ihren Zwillingen auf Zwerde als im ‚weiten‘ Sinne verschieden dadurch begründet, dass die einen durch Token von Wasser, die andern aber durch Token von Zwasser verursacht werden.

Nun haben allerdings auch kausale Theorien des Gehaltes schon *prima facie* mit, von ihren Vertretern selbst eingestandenen, ernsthaften Schwierigkeiten zu kämpfen. Denn wenn sie darauf hinauslaufen, zu behaupten, dass etwa ein ‚mentalesisches Symbol‘ eine Eigenschaft dann repräsentiert, wenn – mit Fodors Worten – „it’s nomologically necessary that *all and only* instances of the property cause tokenings of the symbol“²³⁴, dann fragt es sich natürlich einerseits, ob wirklich *alle* Instantiierungen einer Eigenschaft tatsächlich Instantiierungen des entsprechenden ‚Symbols‘ (bzw. Symboltyps) verursachen müssen und ob andererseits *nur* Instantiierungen der ‚richtigen‘ Eigenschaft Instantiierungen des entsprechenden ‚Symbols‘ verursachen können. Wenn ja, wie könnten wir uns dann täuschen?

Ich werde mich im Folgenden wie gesagt auf die teleofunktionalen Ansätze Dretskes und Millikans konzentrieren, da sie mir auch als die am überzeugendsten ausgearbeiteten und theoretisch fruchtbarsten erscheinen²³⁵. Dretskes, vor allem aber Millikans Position lassen sich jedoch, wie wir noch sehen werden, nicht umstandslos als kausale Theorien des Gehaltes deuten.

Dretske

²³³ Für Formulierungen in diesem Sinne, vergl. etwa Fodor (1987), S. 99 f.

²³⁴ Fodor (1987), S. 100.

²³⁵ Für ähnlich geartete Ansätze, vergl. etwa C. McGinn (1989) und D. Papineau: *Philosophical Naturalism* (1993).

Fred Dretskes hat in seinem einflussreichen Buch *Knowledge and the Flow of Information* (1981) (in der Folge: *KFI*) einen großangelegten und eigenständigen Versuch unternommen, einen Weg aufzuzeigen, wie man einen ‚intentionalen Kuchen aus nicht-intentionalen Zutaten gebacken‘ bekommt. Dazu gehört für ihn insbesondere auch, die oben gestellte Frage zu beantworten, wie Irrtum im Rahmen einer naturalistischen Theorie möglich ist. Die Möglichkeit von Irrtum scheint nämlich ein Kriterium ‚echter Intentionalität‘ zu sein. Auf den Ergebnissen von *KFI* aufbauend, versucht Dretske dann, in *Explaining Behaviour*, insbesondere auch eine Antwort auf die oben angeschnittene, die jüngeren Debatte der *Philosophy of Mind* beherrschende, Frage nach der kausalen, bzw. explanativen Relevanz von Bedeutung und ‚weitem‘ Gehalt zu geben. Bei der zunächst notwendigen, vereinfachenden Darstellung der komplexen Überlegungen von *KFI* werde ich mich, auch terminologisch, auf die für die spätere Diskussion wichtigen Punkte beschränken.

Etwas konkreter ausgedrückt, geht es Dretskes in *KFI* darum, ‚*nicht-natürliche*‘ Bedeutung oder Repräsentationalität, bzw. semantische Gehalte propositionaler Einstellungen, auf der Grundlage ‚*natürlicher Bedeutung*‘ zu erklären. Realisierungen natürlicher Bedeutung (im Sinne von Grice) nennt Dretske ‚*Anzeiger*‘, bzw. ‚*Anzeichen*‘ („*indicator*“). Beispielsweise zeigt das Vorkommen von Spuren von Bärenatzen an, dass Bären da waren – dass Bären anwesend waren, ist die natürliche Bedeutung dieses Anzeichens. Es trägt einen entsprechenden Informationsgehalt allerdings nur, wenn *nur* Bären diese Spuren hinterlassen haben können. Natürliche Bedeutung soll sich dadurch auszeichnen, dass zwischen dem Anzeigen eines Sachverhaltes und dem angezeigten Sachverhalt ein Verhältnis *gesetzmäßiger Abhängigkeit* besteht: Ein Anzeichen zeigt nur dann etwas an, wenn die Beziehung eine eindeutige ist, das heißt, nichts kann im natürlichen Sinne bedeuten, dass *p*, wenn *p* nicht der Fall ist. Dagegen kann etwas offensichtlich auch dann *im nicht-natürlichen Sinne* Bedeutung haben (repräsentieren), wenn der entsprechende Sachverhalt nicht besteht.

Viele natürliche Anzeichen (z. B. bestimmte Wolkenformationen) tragen, so Dretske, einen Informationsgehalt, ohne dass dies ihre *Funktion* wäre. Ein physischer Zustand wird zu einem mentalen Zustand erst dadurch, dass es seine *Funktion* ist, *Informationen zu liefern und umzusetzen*. Eine naturalistische Analyse des Informationsbegriffs im Hinblick auf informationsverarbeitende Systeme mit mentalen Zuständen müsse allerdings die Frage ignorieren, *was* Information sei und statt dessen danach fragen, wie man die *Menge* an Information *quantifizieren* kann, die von einer Quelle zu einem Empfänger transportiert wird. Die mathematische Theorie der ‚*Kommunikation*‘, bzw. Information beantwortet dies üblicherweise im Sinne der

Einschränkung von Möglichkeiten: Jede Halbierung der gegebenen Möglichkeiten hinsichtlich dessen, was der Fall sein kann, entspricht einem *Bit* Information.

Information, im Sinne natürlicher Bedeutung, ist für Dretske der *objektive Grundbaustein* individuellen, propositionalen Geistes – objektiv, weil Information, wie er glaubt, nicht nur ein quantitativer, sondern auch ein *interpretationsunabhängiger* Begriff sei. Man kann sich natürlich fragen, ob ein solcher ‚objektiver‘ Informationsbegriff überhaupt plausibel ist. Ich will an dieser Stelle dazu nur bemerken, dass selbst dann, wenn man unterstellt, dass ein Informationsbegriff allein auf der Grundlage einer kausalen Relation zwischen Anzeigendem und Angezeigtem, also etwas ‚Objektivem‘, Interpretations-Unabhängigem, verständlich zu machen ist, es noch immer zumindest fragwürdig erscheint, dass, wie Dretske darüber hinaus behauptet, der *Begriff des Anzeichens* auch ohne die Unterstellung von Subjekten auskommen könne. Dass die Information, die ein Anzeichen trägt, irgendwie mehr ist, als eine gleichsam ‚nackte‘ Kausalrelation zwischen zwei Ereignissen, kann, wie mir scheint, doch eigentlich nur heißen, dass durch das kausal hervorgerufene Ereignis – bzw. durch einen durch das Ereignis bedingten Zustand – etwas *für etwas* angezeigt wird.

Doch wie dem auch sei, jedenfalls will auch Dretske Information nicht mit *semantischer* Bedeutung verwechselt wissen. Semantische Bedeutung versteht Dretske als eine *Funktion Information zu liefern und umzusetzen*; und sie sei, anders als Information, nicht quantifizierbar. Andererseits hätten sprachliche Symbole oder Überzeugungen zwar eine semantische Bedeutung (bzw. einen ‚engen‘ Gehalt), zur *Wahrheit* und zum *Wissen* brauche es aber wiederum Information:

„(...) signals may *have* a meaning but they *carry* information. What information a signal carries is what it is capable of ‚telling‘ us, telling us *truly*, about another state of affairs. Roughly speaking, information is that commodity capable of yielding knowledge, and what information a signal carries is what we can learn from it“ (Dretske 1981, S. 44)

So haben meine sprachlichen Äußerungen (bzw. von mir hervorgebrachten ‚Signale‘) etwa in Fällen, in denen ich lüge, zwar eine semantische Bedeutung, aber sie transportieren keine Information im Sinne Dretskes. Erst durch den *extrinsischen*, afferent-kausalen Vorgang eines Informationsflusses könnten aus Äußerungen wahre Aussagen, aus Überzeugungen Wissen oder aus Erfahrungen Wahrnehmungen werden. Anders als die Quantität jedoch, soll die ‚Qualität‘ der Information, also ihre Bestimmtheit, die Fälle nicht-natürlicher Bedeutung auszeichnet, auch vom Vorwissen (bzw. – nicht-intentional ausgedrückt – der physischen Vorstrukturiertheit) des Empfängers abhängen. Dabei werde die, im Moment der sinnlichen Stimulation

zunächst analoge und deshalb ‚diffuse‘, Information in einem weiterverarbeitenden *Prozess* digitalisiert und so präzisiert, wobei jeder dadurch ermöglichte Gewinn an Qualität durch einen Verlust an Quantität der ursprünglichen Informationsmenge ‚eingekauft‘ wird. Dieser Verarbeitungsprozess, „the successful conversion of information into (appropriate) digital form“ ist es, der, so Dretske, „constitutes the essence of cognitive activity“ (a. a. O., S. 142).

Dretske will allerdings nicht allen informationsverarbeitenden Systemen eine ‚Intentionalität höherer Ordnung‘, das heißt, *propositionale Einstellungen*, zubilligen. Denn wenn, so Dretske, zwei Eigenschaften *F* und *G* gesetzmäßig korrelieren, dann übertragen Signale, welche auf einer ‚niedrigen Stufe von Intentionalität‘ die Information übermitteln, dass *x* ein *F* ist, *ununterscheidbar* auch die Information, dass *x* ein *G* ist. Auf der Ebene propositionaler Einstellungen könne eine entsprechende Unterscheidung zwischen *F* und *G* jedoch möglich sein. Für höher stufige Intentionalität im Sinne Dretskes bedarf es also mehr, als der Digitalisierung von Information auf der Grundlage natürlicher Bedeutung. Vor allem aber ist die, für propositionale Einstellungen konstitutive, Möglichkeit des *Irrtums*, den Dretske ‚Fehlrepräsentation‘ („*misrepresentation*“) nennt, offenbar nicht umstandslos auf der Grundlage *gesetzmäßiger* Korrelationen zwischen einem ‚intentionalen Ereignis‘ und dessen Ursache verständlich zu machen.

Dretske selbst diskutiert dieses Problem in ‚Misrepresentation‘ (1986) ausführlich anhand eines anschaulichen Beispiels: Einige im Meer lebende Bakterien haben interne Magnete (Magnetosome), mit denen sie sich am Magnetfeld der Erde orientieren. Die Funktion dieses ‚Anzeigers‘ besteht offenbar darin, den Mikroben den Weg in tiefere, sauerstoffärmere Gewässer zu weisen, was für ihr Überleben notwendig ist. Da nun die Feldlinien des Magnetfeldes der nördlichen Hemisphäre genau umgekehrt zu denen der südlichen verlaufen, schwimmt ein südliches Bakterium, welches man in der nördlichen Hemisphäre aussetzt, genau in die falsche Richtung, also dem schädlichen, sauerstoffreichen Wasser an der Oberfläche entgegen. Eine dem entsprechende ‚Verwirrung‘ kann etwa auch ein in die Nähe der Bakterien gebrachter Stabmagnet stiften.

Nun könnte man entweder die Ansicht vertreten, es handle sich in solchen Fällen um Fehlrepräsentationen. Denn wenn man die Anzeigefunktion der Magnetosome so interpretiert, dass sie den Weg zu sauerstoffarmem Wasser weisen sollen, werden sie in den genannten Fällen eindeutig in die *falsche* Richtung deuten, also anzeigen, es gäbe dort sauerstoffarmes Wasser, wo dies gerade nicht der Fall ist. Man könnte aber offenbar genauso gut sagen, dass die Anzeigefunktion der Magnetosome in den betreffenden Bakterien lediglich darin besteht, die Richtung des Magnetfeldes

anzuzeigen. In diesem Fall wäre der entsprechende Mechanismus zwar *unzuverlässig* im Hinblick auf die Anzeige sauerstoffarmen Wassers. Es handelte sich jedoch nicht um eine Fehlrepräsentation, sondern eher – wenn man so will – um einen ‚Fehlschluss‘ von der (richtig angezeigten) Richtung des Magnetfeldes auf die Richtung des sauerstoffarmen Wassers.

Für den Dretske-Kritiker Dennett stellen solche Beispiele nur Spezialfälle eines allgemeinen Phänomens dar, nämlich der *grundsätzlichen Unbestimmtheit des ‚Intentionalen‘* (bzw. der ‚intentionalen Eigenschaften‘ eines entsprechend komplexen Systems), sowohl gegenüber den für ‚intentionale Erklärungen‘ relevanten behavioralen Fakten, als auch gegenüber den systeminternen physischen Fakten. Es mache eben keinen Sinn, so Dennett, danach zu fragen, worin der Gehalt eines ‚intentionalen Zustandes‘ oder ‚Ereignisses‘ *wirklich* besteht. Dennetts Lieblingsbeispiel ist der Fall eines Automaten für Erfrischungsgetränke, der in den USA so entworfen und produziert wurde, dass er auf den Einwurf eines Vierteldollarstückes mit dem Auswurf eines Getränkes reagiert²³⁶. Beim Einwurf eines Vierteldollars geht der Automat in einen funktionalen Zustand *Z* über, der ‚bedeutet‘: ‚Ich erhalte/akzeptiere jetzt einen echten Vierteldollar‘. Solche ‚Zwei-Bitser‘ könnten natürlich ‚Fehler machen‘, etwa wenn eine andere Münze mit den gleichen physisch relevanten Eigenschaften eingeworfen wird. Dann handelt es sich, nach Dennett, deshalb um einen Fehler, weil der Automat auf andere Inputs, als Vierteldollarstücke, nicht mit der Herausgabe eines Getränkes reagieren *soll*. Was dabei als Fehler gilt, sei jedoch als *relativ auf die Umgebung* zu verstehen, in welcher der Automat nach den *Absichten der Hersteller* seine Funktion erfüllen soll. Nehmen wir nun an, der Getränkeautomat würde, nach einigen ‚Dienstjahren‘, in ein anderes Land verkauft, in dem ein bestimmter Münz-Typ der dortigen Währungseinheit zufälligerweise die für die Maschine relevanten physischen Eigenschaften eines Vierteldollars hat. Dieser Münz-Typ hat dort auch genau jenen Wert, der dem Preis entspricht, den ein Erfrischungsgetränk aus dem Automaten kostet. *Unter diesen Umständen* müssten wir wohl sagen, dass der Automat ‚einen Fehler macht‘, wenn er ein *Vierteldollarstück* als Einwurf annimmt. Realisiert der entsprechende innere physische Zustand des Automaten deshalb einen *anderen* funktionalen Zustand (*Z₁*)? Für Dennett jedenfalls ist es einfach sinnlos, herausbekommen zu wollen, was der innere Zustand des Automaten *wirklich bedeutet*, also unabhängig von den Absichten derer, die den Automaten gebaut haben oder unabhängig von der Umgebung, in der er aufgestellt wurde. Und seine zusätzliche Pointe ist, dass wir in dieser Hinsicht überhaupt keinen Grund hätten, einen

²³⁶ Dennett (1987a), S. 290 f.

prinzipiellen Unterschied zwischen Artefakten und Organismen, zwischen dem Getränkeautomaten und den Bakterien oder uns selbst zu machen.

Eine etwas andere Form, in der das Problem der Möglichkeit von Fehlrepräsentationen aufgeworfen wurde, stammt von unserem alten Bekannten Jerry Fodor und wird als *Disjunktionsproblem* bezeichnet²³⁷. Das Disjunktionsproblem betrifft *prima facie* nicht nur die dretskesche, sondern alle Versionen kausalistisch-externalistischer Theorien des Gehaltes: Nehmen wir an, dass ein mentaler Zustand M_1 (etwa eine Überzeugung mit dem Gehalt „Da ist ein Pferd“) nicht nur durch A 's (Pferde) verursacht wird, sondern manchmal auch durch B 's (Kühe in der Dämmerung). Was macht nun einen durch ein B verursachten mentalen Zustand M_1 zu einer *Fehlrepräsentation* mit dem Gehalt p („Da ist ein Pferd“), anstatt zu einer Repräsentation mit dem Gehalt $p \vee q$ („Da ist ein Pferd *oder* eine Kuh in der Dämmerung“)? Auf der Grundlage von natürlicher Bedeutung allein scheint es prinzipiell keine Möglichkeit zu geben, zwischen einem Irrtum und einem entsprechend disjunktiv erweiterbaren Gehalt zu unterscheiden.

Dretske selbst war sich des Problems bereits in *KFI* bewusst. Er schlägt dort vor, dass wir uns zu seiner Klärung kognitiven Systemen mit einem gewissen Grad an Komplexität zuwenden müssen; Systemen, die über (mindestens) zwei *voneinander unabhängige Indikatoren* für bestimmte Zustände oder Ereignisse verfügen: Denken wir uns ein System S , welches über zwei in der Regel *verlässliche* Indikatoren I_1 und I_2 verfügt, die ihrerseits unabhängig voneinander Ursache einer bestimmte Repräsentation R sein können. Dann, so Dretske, könne man die eigentliche, gehaltbestimmende Ursache U_1 einer Repräsentation R in S an dem Punkt lokalisieren, an dem sich die (zurückverfolgten) Linien des Informationsflusses der R verursachenden Indikatoren I_1 und I_2 außerhalb des Systems überschneiden. I_1 und I_2 reagieren – da unabhängig voneinander – zwar auf unterschiedliche Eigenschaften (E_1 und E_2), aber diese unterschiedlichen Eigenschaften sind dann eben beide Eigenschaften von U_1 . Auf der Grundlage dieser Überlegung scheint das Problem der *Möglichkeit* von Fehlrepräsentationen nun dadurch lösbar, dass man sagt, dass es einen Typ von Ereignissen – und damit eine mögliche Ursache U_2 – geben kann, der zwar die Eigenschaft E_1 , nicht aber die Eigenschaft E_2 hat. U_2 könnte dann R schon *qua* einer Verursachung von I_1 hervorrufen. Da R aber U_1 repräsentiert, weil es gleichzeitig von I_1 und I_2 verursacht wird, handele es sich um einen Fall von Fehlrepräsentation.

Damit sind die Schwierigkeiten aber offensichtlich noch keineswegs ausgeräumt. Denn wodurch ist eigentlich bestimmt, dass R wirklich U_1 repräsentiert? Selbst wenn wir irgendwie ausschließen könnten, dass R lediglich U_1 , und nicht U_1 *oder* U_2

²³⁷ Zuerst in Fodor (1984). Vergl. auch ders. (1987) und (1990).

‚bedeutet‘, wäre damit ja nicht auch schon gezeigt, dass R nicht I_1 oder I_2 ‚bedeuten‘ könnte. Wenn I_1 und I_2 unabhängig voneinander funktionieren, spricht ja *per se* nichts dafür, dass sie *immer gleichzeitig* R bewirken müssen. Dies könnte man doch offenbar nur sagen, wenn R nur dann U_1 ‚bedeuten‘ soll, wenn beide Indikatoren R gleichzeitig bewirken. Damit kämen jedoch normative oder teleologische Spezifizierungen der Bedeutung von R ins Spiel, die nun ihrerseits im Rahmen einer naturalistischen Begrifflichkeit zu erläutern wären.

Dretske versucht dies bereits in *KFI* anhand eines dispositional verstandenen Konzeptes einer *Lernphase*. Dadurch soll es möglich sein, zwischen der in der Lernphase etablierten Korrelation von mentalen Token eines bestimmten Typs mit äußeren Umständen und dem, was deren Instantiierung danach tatsächlich verursacht, zu unterscheiden. Für Dretske ist es plausibel anzunehmen, dass zumindest ‚einfache‘ (beobachtungsbezogene) Begriffe in einem Prozess geformt werden, bei dem bestimmte neuronale Zustände (Strukturen) eine informative *Funktion erwerben*, bezüglich der sich spätere Instantiierungen eines entsprechenden Typs als bestimmte Überzeugungen klassifizieren lassen. Die Ursache einer späteren Instantiierung einer Überzeugung kann dann auch von dem *durch den jeweiligen Lernprozess fixierten* Typ von Ursache abweichen. Wenn also M_1 -Zustände, nach einer Lernperiode, die sie zuverlässig (nur) mit Umständen vom Typ U_1 korreliert, dennoch durch Umstände vom Typ U_2 verursacht würden, so handele es sich um Fälle von Fehlrepräsentation. (Wie wir im einzelnen noch sehen werden, unterscheidet auch Dretske zwischen ‚originärer‘ und ‚abgeleiteter Intentionalität‘: Nur bei ‚originärer Intentionalität‘ nämlich, nicht jedoch im Falle von Artefakten, lasse sich sinnvoll von Fehlrepräsentationen reden, da ja letztere keine Lernphase durchliefen. Damit wäre zumindest Dennetts Beispiel des ‚wandernden‘ Getränkeautomaten irrelevant)

Bevor wir uns nun Dretskes spätere Ausarbeitung des Konzeptes der Lernphase in *Explaining Behaviour* genauer ansehen, möchte ich hier nur einwerfen, dass es in dieser Form dem Einwand ausgesetzt ist, dass sich eine klare Grenze zwischen einer Lernphase und dem Danach – jedenfalls unter den in *KFI* gemachten Voraussetzungen – kaum mit guten Gründen ziehen lässt. So stellt denn auch Fodor (1984) die rhetorische Frage, ob jemand in eine Trillerpfeife blase, wenn die Lernperiode vorbei ist. Dretske ist zwar der Ansicht, dass es keine scharfen Grenzen geben müsse (und auch nicht gäbe), wenn das zu erklärende Phänomen selbst unscharf sei. Wir könnten ja schließlich auch nicht genau bestimmen, wann jemand einen Begriff beherrscht und wann nicht. Doch unter dieser Voraussetzung würde möglicherweise – entgegen Dretskes eigenen ‚realistischen‘ Intentionen – auch die Differenz zwischen richtigen und falschen Repräsentationen unscharf.

Man kann zudem fragen, ob mit dem Konzept der Lernphase wirklich ein entscheidender Fortschritt im Hinblick auf das Disjunktionsproblem erreicht wird. Denn warum sollte dies nicht auch für die Lernphase gelten? So argumentiert Fodor, dass für die Frage, *welche* Beziehung zwischen einer Repräsentation und der Welt durch das Training während einer Lernphase herbeigeführt wird, auch *kontrafaktische* Beziehungen relevant seien. Denn *wären* Umstände des Typs U_2 *während der Lernphase* ins Blickfeld geraten, *hätten*, gemäß den Voraussetzungen, auch sie M_1 -Zustände hervorgerufen. Dieser Einwand ist aber, wie ich denke, nicht stichhaltig, wenn man die fragliche Beziehung retrospektiv als auf die Umstände beschränkt bestimmt, die während der Lernphase *tatsächlich* ins Blickfeld geraten sind.

Dennoch spricht manches dafür, dass der einzige, noch im Rahmen einer materialistischen, bzw. funktionalistischen Begrifflichkeit verbleibende, Weg aus diesen Schwierigkeiten in einer *normativen Qualifizierung* der Umstände der Verursachung mentaler Zustände oder Ereignisse liegt. Dies müsste natürlich als mit den Vorgaben des Naturalismus vereinbar verständlich gemacht werden können. Ein zeitweise von Fodor vertretener Vorschlag will richtige (wahre) von falschen Repräsentationen dadurch unterscheiden, dass erstere nur unter *biologisch idealen Bedingungen* zustande kommen²³⁸. Der Gehalt einer Repräsentation entspräche dann demjenigen, durch was sie unter entsprechend zu spezifizierenden Bedingungen verursacht würde. Wie Fodor jedoch später, in *Psychosemantics*, selbst bemerkt, ist es unmöglich, ideale Bedingungen für Beobachtungen *unabhängig vom Gehalt* eines entsprechenden Beobachtungsurteils festzulegen: Denn was etwa ideale Bedingungen zur Beobachtung sehr kleiner Objekte sein mögen, wären für eine Beobachtung sehr großer Objekte alles andere als ideal²³⁹. Der vielversprechendere Weg scheint deshalb zu sein, die *Funktionen* ‚intentionaler Zustände‘ oder ‚Ereignisse‘ *selbst* normativ, bzw. teleologisch zu bestimmen: Normativ in dem Sinne, dass eine bestimmte ‚Struktur‘ eine bestimmte Information liefern *soll*. Teleologisch etwa in dem Sinne, dass ein

²³⁸ Fodor (1984).

²³⁹ Weil Fodor nun nicht mehr an eine zirkelfreie normative Spezifizierbarkeit der Bedingungen für (wahre) Überzeugungen im Sinne von Idealisierungen, bzw. eines evolutionär-teleologischen Verständnisses von Funktionen, glauben mag, macht er in (1990a) einen neuen Vorschlag, wie das Disjunktionsproblem zu lösen sei. Demnach repräsentiert ein Zustand r die Eigenschaft F (bspw.: ein Pferd zu sein) und nicht die Eigenschaft F oder G (resp.: ein Pferd oder eine Kuh in der Dämmerung zu sein), wenn die Kausalbeziehung zwischen G 's und r -Zuständen in *asymmetrischer Weise* von der Kausalbeziehung zwischen F 's und r -Zuständen abhängt. Dies wäre dann der Fall, wenn F 's auch dann r -Zustände verursachen würden, falls G 's dies nicht täten, aber umgekehrt G 's r -Zustände nicht verursachen würden, falls F 's dies nicht täten. Zur Kritik an dieser Konzeption Fodors, siehe Beckermann (1999), S. 341 f. und P. Boghossian: ‚Naturalizing Content‘ (1991).

mentaler Zustand, im Falle einer Fehlrepräsentation, nicht dafür ‚gemacht‘, bzw. es nicht ihr *Zweck* ist, unter solchen Umständen verursacht zu werden.

Diesen Weg schlägt Dretske in *Explaining Behaviour* dann auch konsequent ein. Nun hatte ich ja oben bereits in allgemeiner Form dafür argumentiert, dass sich intentionale oder teleologische Begriffe nicht ohne Verlust gerade desjenigen explanativen Wertes, den man durch sie erreicht, kausalistisch substituieren lassen. Für eine moderate Form von Materialismus kann es aber hinreichen, dass teleologische Erklärungen keine über-physischen Mechanismen oder Akteure postulieren müssen, sondern *auf der Grundlage* normaler physischer Gesetzmäßigkeiten möglich sind, bzw. in ein naturwissenschaftliches Weltbild *integrierbar* sind. Eine Frage, die sich stellen wird, ist, ob und inwiefern noch von einem ‚Reduktionismus‘ die Rede sein kann.

Ein erster, wichtiger Schritt in der argumentativen Struktur von *Explaining Behaviour* besteht darin, den ‚pragmatischen‘, bzw. den Output-Aspekt der Bedeutung, also die Rolle, die eine Repräsentation für das *Verhalten* eines Organismus spielt, explanativ stärker zur Geltung kommen zu lassen. Vereinfacht gesagt, soll eine neuronale Struktur in einem (ontogenetischen) Lernprozess eine Funktion erhalten, *weil* durch sie ein kausaler Zusammenhang zwischen ihrem (natürlichen) Informationsgehalt und einem für den Organismus vorteilhaften Verhalten besteht. Den Lernprozess beschreibt Dretske mehr oder minder behavioristisch: Ein innerer Mechanismus, der kontingenterweise ein bestimmtes äußeres Ereignis mit einem behavioralen Output korreliert, würde durch extrinsische, reizverstärkende Effekte verstetigt. Wichtig ist für Dretske aber, dass die teleologisch verstandenen Funktionen eine besondere, *nicht-reduzierbare explanative Rolle* in der Erklärung von Verhalten einnehmen. Neben dem Problem der Möglichkeit von Fehlrepräsentation, beunruhigt ihn nun nämlich insbesondere auch der, oben bereits ins Spiel gebrachte, Einwand des *Epiphänomenalismus*, nach dem ‚intentionale Eigenschaften‘ gegenüber den zugrundeliegenden physischen Eigenschaften *kausal* – und demnach eben scheinbar auch *explanativ* – *überflüssig* sind.

Das Problem des Epiphänomenalismus wird von manchen Theoretikern als ein (wirkliches oder nur vermeintliches) Dilemma gedeutet, welches dem Reduktiven Materialismus erst durch den Externalismus beschert wurde²⁴⁰: Einerseits würden internalistische Theorien, welche die kausalen Rollen von Gehalten als zu den internen physischen Zuständen von Subjekten supervenient verstehen könnten, unseren semantischen Intuitionen nicht gerecht, während es andererseits rätselhaft erscheinen kann, wie die durch diese Intuitionen geforderten externen Faktoren eine kausale Rolle im Verhalten der betreffenden Subjekte spielen können sollen. Die Drohung des

²⁴⁰ So etwa bei McGinn (1989).

Epiphänomenalismus betrifft prinzipiell jede ‚Intentionale Psychologie‘ mit realistischem Selbstverständnis, also jede Theorie, die semantische Eigenschaften (von propositionalen Einstellungen) als koextensiv mit kausalen Eigenschaften verstehen will²⁴¹. Es stellt sich nämlich *nicht nur* die Frage, wie semantische oder ‚intentionale‘ Eigenschaften in Verhaltenserklärungen zum Zuge kommen können sollen, wenn diese Eigenschaften nicht zu den tatsächlich kausalwirksamen Eigenschaften *supervenient* sind, das heißt, wenn etwa der Gehalt einer Überzeugung nicht allein von den *inneren physischen* Zuständen des Subjektes dieser Zustände determiniert ist (wie die externalistischen Argumente ja offenbar zeigen können und worauf ‚*Double-Aspekt-Theories*‘ eine ‚versöhnende‘ Antwort zu finden glauben). – Schlimmer noch: Es lässt sich nämlich, wie Dretske sieht, *ganz grundsätzlich*, also völlig unabhängig von externalistischen Erwägungen, fragen, welche Rolle Bedeutungshaftigkeit (im Sinne von Gehalten) bei der Verhaltenserklärung überhaupt spielen kann, wenn *allein* die Kausalität der *physischen Eigenschaften* eines Systems für dessen Verhalten ausschlaggebend ist. Eine der wesentlichen Zielsetzungen von *Explaining Behaviour* ist es daher, einen Weg aufzuzeigen, wie man Bedeutung ‚explanatorischen Biss‘ verleihen kann.

Dretskes Lösungsvorschlag setzt an der, im Zusammenhang mit der Analyse kausaler Erklärungen bereits vorgestellten, *Unterscheidung zwischen auslösenden und strukturierenden Ursachen* an: Das Auftreten eines Ereignis *B* kann, unter einer entsprechenden gesetzesartigen Verallgemeinerungen, durch das Auftreten des Ereignisses *A* als seiner *auslösenden* Ursache erklärt werden. Wir können jedoch auch nach der *Ursache des Verursachens* fragen, also danach, wie *A*’s Verursachen von *B* erklärt werden kann. Dretskes These ist nun, dass es eigentlich letzteres, nämlich die *strukturierende* Ursache ist, auf die es bei einer Erklärung von Verhalten ankommt, *wenn wir so etwas wie Überzeugungen oder Wünsche ins Spiel bringen*. Was erklärt werden soll, ist, warum ein bestimmter *Prozess* (*das Verursachen* von Körperbewegungen) auftrat, nicht das *Produkt* von Verhalten, also Körperbewegungen. Bei Verhalten, das wir durch *Gründe* erklären, redet Dretske auch von *Handeln*. Jemand handelt in einer bestimmten Weise, *weil* er dies glaubte und jenes wollte.

„On this view of behaviour and its explanation, it turns out that biologically indistinguishable organisms can not only be in psychologically different states, this psychological difference can help to explain their respective *bodily* behaviours. They are both moving their arm – moving it, in fact, in exactly the same way. Yet they may have quite different reasons for moving their arms in this way: the one person is waving goodbye, the other brushing away a fly. This means, in turn, that insofar as otherwise identical behaviours can qualify as

²⁴¹ Vergl. etwa: J. Kim: ‚Epiphenomenal and Supervenient Causation‘ (1984).

different *actions* because the movements are produced with different intentions (...) biologically identical organisms can *act* in much different ways. These results are achieved by identifying an organism's behaviour, not with the surface changes – including bodily movements – that are internally produced, but with the process in which such changes are brought about.“²⁴²

Eine wichtige Konsequenz dieses Verständnisses der Rolle, die *Gründe* in Verhaltenserklärungen spielen sollen, ist, dass sie als *Teil des Verhaltens*(-prozesses) verstanden werden und *nicht*, wie heute bei vielen analytisch gesinnten Philosophen üblich, als dessen *Ursache*. Sie bilden die strukturierende Ursache der *anderen Komponente* von Verhalten, den Körperbewegungen. Und nur wenn bei Verhaltenserklärungen die strukturierende Ursache von Körperbewegungen im Sinne von Gründen für ein Verhalten (bzw. Handeln) ins Spiel kommt, kann nach Dretske die explanative Rolle von Bedeutung genuin sein.

Damit hat sich Dretske auf eine, jedenfalls im Hinblick auf den *explanativen Charakter* ‚psychologischer Erklärungen‘, moderat anti-reduktionistische Position festgelegt. Und wie er auch selbst betont, würden Fodor, Lewis und andere der Besonderheit ‚intentionaler‘ Erklärungen gerade *nicht* gerecht, wenn sie nach Möglichkeiten einer Korrelationen zwischen neuronalen oder ‚syntaktischen‘ und semantischen Eigenschaften suchen. Solche reduktiven Korrelationen taugten lediglich zur *Vorhersage* von Verhalten, nicht jedoch zu dessen *Erklärung*, da sie die auslösenden, etwa neuronalen, Ursachen betreffen und nicht die strukturierenden. Denn man könne die auslösende Ursache eines Verhaltensprozesses kennen – und diesen somit schließlich vorhersagen, ohne deshalb zu wissen, *warum* der Verhaltensprozess ausgelöst wurde.

Andererseits versteht sich aber auch Dretske insofern als Reduktiver Materialist, als er davon ausgeht, dass es ‚psycho-physische Eigenschaftskorrelationen‘ in irgendeiner Form gibt (wenngleich nicht unbedingt als einfache Typenidentitäten) und dass allen Formen der Erklärung ‚mechanistische‘ Erklärungen auf der Grundlage physischer Eigenschaften zugrunde liegen müssen. Diesem augenscheinlichen Widerspruch glaubt er durch eine *Erweiterung der Reduktionsbasis* ‚psychischer Eigenschaften‘ um *extrinsische* physische Eigenschaften begegnen zu können: Bedeutungen seien eben als eine spezielle Art physischer Fakten zu verstehen, als Fakten, die nicht nur auf ‚intrinsischen‘ biologischen (etwa neuronalen) Eigenschaften, sondern darüber hinaus auf ein weites Netz von Beziehungen zwischen einem Organismus und der Welt inklusive seiner Geschichte supervenieren²⁴³. Mit dieser Erweiterung des

²⁴² F. Dretske: ‚Does Meaning Matter?‘ (1990), S. 12.

²⁴³ F. Dretske: ‚Dretske's Replies‘ (1991), S. 211.

Supervenienzbegriffs ist allerdings – abgesehen von den obigen, etwas vagen Andeutungen bezüglich einer Lernphase – noch nichts Genaueres darüber gesagt, *wie* die genannten extrinsischen Fakten mit intrinsischen Fakten zusammenhängen können sollen, bzw. *wie* die externen oder bereits zurückliegenden (,historischen‘) Kausalbeziehungen schließlich subjektintern kausale Wirksamkeit im jeweiligen Hier und Jetzt (im Sinne des Auslösens von Körperbewegungen) entfalten könnten.

Jedenfalls versteht Dretske ,psychologische Erklärungen‘, die nach strukturierenden Ursachen fragen, als nicht durch biologische (oder sonstige ,mechanistischen‘) Erklärungen substituierbar. Die methodologische Eigenständigkeit einer ,kognitiven Psychologie‘ verdanke sich der Tatsache, dass die explanative Rolle von Gründen eine *genuine* ist. Eine psychologische Verhaltensklärung als eine Erklärung *qua Gründe* zu verstehen, heißt für Dretske, die Verursachung solcher Körperbewegungen, die wir als Handlungen interpretieren, *dadurch* erklären zu können, dass ein mentaler Zustand die *Funktion* hat, *das* zu repräsentieren, was er repräsentiert (und ein bestimmter, rationaler Zusammenhang mit Wünschen oder Absichten besteht – was ich hier vernachlässige):

„An indicator element (...) becomes a representation by having part of *what* it indicates (...) promoted to an explanatorily relevant fact about itself. A belief is merely an indicator whose natural meaning has been converted into a form of non-natural meaning by being given a job in the explanation of behaviour. What you believe is relevant to what you do because beliefs are precisely those internal structures that have acquired control over output, and hence become relevant to the explanation of system behaviour, in virtue of what they, when performing satisfactorily, indicate about external conditions.“ (Dretske 1988, S. 84)

Eine Repräsentation wird also – so verstehe ich Dretske – zum Element einer Verhaltens- bzw. Handlungserklärung, zu einem Grund, weil *was* sie repräsentiert (Teil der) Ursache dafür ist, dass sie (unter bestimmten Umständen) zur Ursache von bestimmten Körperbewegungen werden kann. Die Frage, *warum* eine Körperbewegung durch einen mentalen Zustand verursacht wurde, führt nicht zu auslösenden Stimuli, sondern zur jeweiligen *Funktionsweise* mentaler Zustände – und damit schließlich zu den Umständen des Zustandekommens der entsprechenden Verursachungsbeziehung. Eine naturalistische Theorie des Geistes müsse also *historische Fakten* bemühen, um zu erklären, *warum* ein System sich so verhält, wie es sich verhält. Und sie tue dies, indem sie erklärt, *wie* ein mentaler Zustand eine Anzeige- oder Repräsentationsfunktion *aufgrund dessen*, *was* er anzeigt oder repräsentiert, erhalten hat. Dazu müsse ein Beobachter insbesondere herausfinden, wie der Verursachungszusammenhang zwischen mentalen Zuständen und motorischen Outputs zuallererst etabliert wurde.

Bevor wir mit der Erläuterung dieses Gedankens fortfahren, möchte ich kurz einhaken, um bereits hier einen schwerwiegenden Verdacht zu äußern: dass nämlich der ganze Trick von Dretskes naturalistischer ‚Versöhnungsstrategie‘ in *Explaining Behaviour* auf einer ähnlichen Art mehrdeutiger Verwendung des Wortes „Grund“ beruhen könnte, wie wir sie oben bereits bezüglich Fodors Verwendung des Ausdrucks „Syntax“ konstatiert hatten. Denn Gründe für Handlungen sind, nach unserem üblichen Verständnis, etwas, das ein Tun im Rahmen eines sinn-konstitutiven (und wie ich sagen würde: *normativen*) Möglichkeitsraumes rechtfertigt oder ‚rationalisiert‘, also verständlich macht; und sie müssen durch das *Subjekt einer Handlung* gegebenenfalls *selbst* zur Rechtfertigung oder Erklärung seines Tun *vorgebracht werden können*. An dieses Verständnis scheint Dretske zunächst anzuschließen, wenn er einräumt, dass Gründe nicht einfach als Ursachen von Körperbewegungen zu verstehen und deshalb ‚psychologische Erklärungen‘ genuin seien. Doch am Ende sollen Gründe offenbar *nur* aus der *Perspektive eines Beobachters*, der das Verhalten eines Organismus erklären will, als solche fungieren und, nach Dretskes Analyse, den *Ursachen einer Ursache* von Körperbewegungen entsprechen.

Diesen Punkt sollten wir jedenfalls im Hinterkopf behalten, wenn wir uns im Folgenden noch etwas genauer ansehen, wann und warum wir, nach Dretske, ein bestimmtes Verhalten überhaupt durch Gründe erklären können (bzw. müssen). Es soll nämlich von der *Art des Zustandekommens* der verhaltensrelevanten Verursachungszusammenhänge abhängen, worin eine Verhaltensklärung bestehen muss. Ziemlich klar scheint dies in Fällen der Erklärung des Verhaltens von Artefakten. Die teleologischen Erklärungen, die wir etwa bezüglich der Funktionsweise eines Thermostats in Anschlag bringen, sind nach Dretske genuin, lassen sich aber auf unsere Absichten und ihre technische Umsetzung zurückführen. Die auslösende Ursache, also etwa das Fallen der Raumtemperatur, führt zur Tätigkeit eines Heizbrenners, *weil wir* dem Thermostat diese *Funktion gegeben* haben. (Dem Thermostat kann jedoch eine ‚semantische Bedeutung‘ nur gegeben werden, weil ihm schon eine *natürliche*, kausale *Anzeigeeigenschaft* zukommt.)

Dretske versteht die ‚Intentionalität‘ von Artefakten also als *derivativ* zur der ihrer Hersteller. Da dementsprechend die Erklärung der strukturierenden Ursachen ihres Verhaltens bereits intentional (im doppelten Sinne von semantisch und absichtlich) ‚infiziert‘ sei, könne sie nicht zu einer naturalistischen Erklärung der kausalen Wirksamkeit von Bedeutung beitragen. Ein erster Schritt auf dem Weg, uns als die ‚Quelle von Intentionalität‘ aus der Erklärung herauszubekommen, wäre dagegen getan, wenn man sich die Entwicklung von Mechanismen, die der Verhaltenskontrolle bei Pflanzen und Tieren dienen, ansehe. Denn hier könnten wir durch den Prozess der

natürlichen Auslese erklären, wie ein bestimmter Anzeigezustand *A* zur Ursache eines bestimmten Verhaltens *V* wurde: Wenn es nämlich für einen Organismus *O* im Überlebenskampf gegenüber Konkurrenten irgendwie *vorteilhaft* war, wenn er unter den Umständen *U*, die durch *A* angezeigt werden, *V* produzierte, lag damit schließlich auch die Wahrscheinlichkeit der biologischen Reproduktion von *O* höher. Durch diesen ‚Selektionsmechanismus‘ erhielt dann *A* die *Funktion*, durch *U*’s verursacht zu werden und seinerseits *V* zu verursachen. Es sei die Natur, „in ihrer eigenen, absichtslosen Art“²⁴⁴, die in diesem Erklärungsmodell *quasi* die Rolle des ‚Designers‘ übernehme.

Allerdings glaubt Dretske nicht, dass solche evolutionstheoretischen, bzw. selektionistischen Erklärungsansätze Überzeugungen und Wünschen bereits eine genuine, nicht-überflüssige Rolle als Gründe für Handlungen einräumen können. Er ist nämlich mit Cummins, Sober und anderen der Auffassung, dass *individuelles Verhalten* durch selektionistische Verallgemeinerungen nicht erklärt werden kann. Dies ginge, wie Cummins bemerkt, genauso wenig, wie Peters Vorliebe für Rothaarige erklären kann, warum dessen Freundin Doris rote Haare hat. (Wenn überhaupt, dann sei es andersherum: Ihre Rothaarigkeit erkläre, warum Peter sie ausgewählt hat.²⁴⁵) Das *instinktgesteuerte* Fluchtverhalten eines *bestimmten* Tieres, also die Tatsache, dass ein bestimmter Zustand seines Nervensystems eine Änderung der Bewegungsrichtung bewirkt, wird nämlich, so Dretske, durch dessen genetische Veranlagung erklärt, nicht durch den selektiven Vorteil seiner Vorgänger. Was der jeweilige neuronale Zustand des Tieres anzeigt, spiele dabei keine Rolle, denn was immer er anzeigt, sei, *bei genetisch determinierten Verhaltensmustern*, für die Erklärung des jeweils durch ihn bewirkten motorischen Outputs ohne Belang.

Diese Behauptung kommt nun allerdings, wie ich meine, doch etwas überraschend. Bemerkenswert finde ich daran bereits, dass Dretske auch in diesem Fall eine für einen Naturalisten erstaunlich scharfe Grenze, nämlich zwischen instinktgesteuertem und nicht-(mehr-) instinktgesteuertem Verhalten, ziehen zu können glaubt. Und es fragt sich, warum wir überhaupt sinnvoll – und bisher nicht substituierbar – in ‚intentionaler‘ Form auch über instinktgesteuertes (individuelles) Tierverhalten reden können, wenn es bei dessen Erklärung auf die ‚Inhalte‘ gar nicht ankommt. Jedenfalls ist die explanatorische Differenz, die sich an der von ihm gezogenen Grenze festmachen lassen soll, für Dretske argumentativ so zentral, dass wir uns ihre Begründung noch etwas genauer ansehen sollten.

Die Differenz von Erklärungsweisen, auf die sich Dretske beruft, hat Elliot Sober (im Anschluss an Lewontin) als den Unterschied zwischen ‚*selectional explanations*‘

²⁴⁴ Dretske(1988), S. 92 (meine Übersetzung).

²⁴⁵ Cummins (1975).

und ‚*developmental explanations*‘ bezeichnet²⁴⁶. So könnten wir einerseits ‚entwicklungsmäßig‘ (bzw. ontogenetisch) erklären, warum alle Kinder in einem Klassenraum auf dem Niveau der dritten Klasse lesen, in dem wir erklären, warum jedes einzelne der Kinder in dem Raum entsprechend lesen kann. Wir könnten dies andererseits aber auch ‚selektional‘ erklären, indem wir erklären, warum nur Kinder, die auf dem Leseniveau von Drittklässlern sind, überhaupt in dem Raum sitzen. Letztere Erklärung sage uns allerdings nichts darüber, warum *bestimmte* Kinder auf dem Level von Drittklässlern lesen. Auf den uns interessierenden Fall von Verhaltensklärung übertragen, heißt dies für Dretske, dass Instantiierungen instinktgesteuerten Verhaltens von Tieren ‚entwicklungsmäßig‘ durch strukturierende Ursachen – insbesondere durch genetische Anlagen – erklärt werden können, ohne das deren ‚historische Eigenschaften‘ für die Erklärung eine Rolle spielen. ‚Selektional‘ könnten wir dagegen erklären, warum es im Hinblick auf Tiere einer bestimmten Art heute *vorwiegend* solche gibt, die sich in dieser und jener Weise verhalten, warum also ein bestimmter neuronaler Zustand die Funktion hat, eine spezifische motorische Reaktion zu bewirken²⁴⁷. Damit kognitive Zustände aber – als Gründe – für die Erklärung *individuellen Verhaltens* relevant werden, müssten sie mehr sein, als Repräsentationen, nämlich *Überzeugungen*, also kognitive Zustände, bei denen der *Inhalt* explanativ ins Spiel komme:

„ (...) to qualify as a belief it is not enough to *be* an internal representation (a map) that is among the causes of the output, something that helps us steer. *The fact that it is a map*, the fact that it *says* something about external conditions, must be relevantly engaged in the way it steers us through these conditions. What is required, in addition, (...) is that the structures indicator properties figure in the explanation of its causal properties, that what it *says* (about external affairs) helps to explain what it *does* (in the production of the output). That is what is missing in the case of reflexes, tropisms, and other instinctive behaviours. Meaning, though it is there, is not relevantly *engaged* in the production of the output. The system doesn't do what it does, *C* doesn't cause *M*, *because* of what *C* (or anything else) means or indicates about external conditions. Though *C* has meaning of the relevant kind, this is not the meaning it has *to* or *for* the animal in which it occurs. That, basically, is why genetically determined behaviours are not explicable in terms of actor's reasons. That is why they are not *actions*. What (if anything) one wants, believes, and intends is irrelevant to what one does.“ (Dretske 1988, S. 94)

²⁴⁶ Sober (1984); R. Lewontin: ‚Darwin's revolution‘ (1983).

²⁴⁷ Dieser Punkt ist in der Debatte allerdings strittig geblieben; s. etwa K. Neander. ‚What does Natural Selection Explain? A Correction to Sober‘ (1988).

Für die Erklärung eines instinktgesteuerten oder reflexartigen, bzw. ‚tropistischen‘ Verhaltens seien ‚Gründe‘ irrelevant, da wir es zwar mit Repräsentationen, nicht jedoch mit Überzeugungen zu tun hätten. Nur in Fällen, wo der Gehalt einer Repräsentation eine explanative Rolle spielt, werde gehandelt, also etwas *mit* Gründen (Überzeugungen, Wünsen usw.) getan. Nur dann sei ‚Intentionalität‘ *genuin und zugleich nicht-derivativ*, werde *individuelles* Verhalten durch strukturelle Ursachen erklärbar. ‚Gründe‘ – im Sinne Dretskes – könne es aber nur dort geben, wo es *individuelles Lernen* gibt: Durch individuelles Lernen²⁴⁸ erhalten neuronale Strukturen *Kraft der Information*, die sie übermitteln (und fürs Verhalten nutzbar machen), die kausale Rolle, die wir ihnen in Verhaltensklärungen zuschreiben.

Der entscheidende Punkt ist also für Dretske, dass tropistisches Verhalten *nicht durch Lernen modifiziert* werden kann. Denn nur im Fall des individuellen Lernen-Könnens haben wir es mit Systemen zu tun, deren ‚Kontrollstruktur‘ *im Rahmen ihrer individuellen Geschichte* durch die Art von Abhängigkeitsbeziehungen geformt worden sind, die zwischen internen und externen Bedingungen bestehen sollen:

„The place to look for these cases are places where individual learning is occurring, places where internal states acquire control duties or change their effect on motor output as a result of their relation to the circumstances on which the success of this output depends“ (ebd. S. 95)

Ganz im Sinne gängiger evolutionstheoretischer (aber auch pragmatistischer) Ansätze geht auch Dretske davon aus, dass die Entwicklung diskriminatorischer Fähigkeiten normalerweise nur so weit vorangetrieben wird, wie es zur Befriedigung der jeweiligen Bedürfnisse von Organismen unter ökonomischen Gesichtspunkten notwendig ist²⁴⁹. Externe Indikatoren, bzw. Repräsentationen würden überhaupt nur dadurch als Ursachen für einen motorischen Output ‚rekrutiert‘, könnten nur deshalb ‚Gründe‘ sein, weil sie durch Motive und Wünsche, bzw. – behavioristisch analysiert – durch die Rezeptivität eines Organismus relativ zu einem Ergebnis des Verhaltens begleitet werden. Da assoziatives Lernen auf positiven und negativen Verstärkungen beruhe, bedürfe es entsprechender ‚Detektorzustände‘, um die (Rück-)Wirkungen, welche die Reaktionen (bzw. das Verhalten) eines Organismus für diesen selbst haben, ‚evaluieren‘ zu können. Detektorzustände, wie etwa Schmerzempfindungen, hätten also die *Funktion*, körpereigene Zustände positiv oder negativ zu ‚bewerten‘. Dretske

²⁴⁸ Lernen versteht Dretske als den Erwerb diskriminatorischer Fähigkeiten durch instrumentelle Konditionierung.

²⁴⁹ Man mag allerdings Zweifel haben, ob sich auch die redliche ‚Suche nach der Wahrheit‘, die uns dazu bringen kann, überkommene Klassifizierungen in Frage zu stellen, umstandslos unter solche Prinzipien ordnen lässt.

analysiert dann entsprechende *motivationale* Zustände (etwa, Schmerzen vermeiden zu wollen) analog zu repräsentationalen Zuständen: Sie werden zu einem ‚Grund‘, zu einer (Komponente einer) strukturellen Ursache eines motorischen Outputs, *weil* dieser Output ein bestimmtes (‚gewünschtes‘) Resultat hervorbringt. Die Details werde ich mir hier sparen. Jedenfalls glaubt Dretske, dass sich genuine mentale Verursachung nur auf der Grundlage des *Zusammenwirkens* von Detektorzuständen (bzw. Wünschen, Motiven) und Indikatorzuständen (bzw. Überzeugungen) verständlich machen lasse.

Dretske diskutiert noch weitere Komplikationen, wie die interne Integration von Überzeugungen in komplexeren Kontroll-Systeme, in denen, neben Wünschen, auch *andere* Überzeugungen, Präferenzordnungen u. ä. eine Rolle spielen, sowie die Veränderung von Handlungszielen. Dies alles ist jedoch für die Frage, auf die es mir hier ankommt, nämlich ob Erklärungen individuellen Verhaltens im Sinne einer Erklärung durch strukturierende Ursachen dem entspricht, was wir üblicherweise unter einer Erklärung durch Gründe verstehen, entbehrlich. Und in der Tat will mir diese Behauptung nicht einleuchten.

Überzeugungen sind, wie Dretske richtig sieht, tatsächlich an die *Möglichkeit*, Gründe für Handlungen abzugeben, gekoppelt. Aber eine Überzeugung muss ebenso gut ein Grund dafür sein können, warum man eine andere Überzeugung hat oder eine bisher bestehende Überzeugung verwirft. Überzeugungen, im primären Sinne, sind tatsächlich nur dann solche, wenn wir sie selbst *mit Gründen* ändern können. Gründe in diesem Sinne entsprechen jedoch keineswegs strukturierenden Ursachen im Sinne Dretskes, also Ursachen von Verursachungen von Ereignissen (wie etwa Körperbewegungen), die auf der Grundlage einer behavioristischen Lerntheorie analysierbar sind. Wie Dretske selber sagt, sind Gründe für Handlungen etwas, von dem der Akteur weiß (oder zumindest wissen kann). Ich würde dies eher so ausdrücken, dass ein Akteur die Gründe für seine Überzeugungen oder für seine Handlungen normalerweise selber vorbringen können muss. Wenn, um eines von Dretskes Beispielen zu nehmen, der Grund dafür, dass ich zum Kühlschrank gehe, der ist, dass ich ein Bier wollte und der Überzeugung bin, dort könnte ich eines finden, so wäre die Tatsache, dass mein Nervensystem durch einen Konditionierungsprozess so ‚strukturiert‘ wurde, bzw. die Funktion erhielt, bei einem bestimmten Stimulus (etwa Bierwerbung im Fernsehen) bestimmte motorische Bewegungen zu verursachen, sicherlich nicht *im gleichen Sinne* ein Grund. Solche Erklärungen sind doch gewiss nicht das, was ein Akteur als *Handlungsgrund* angeben würde und was andere als solchen verstehen oder akzeptieren könnten. Dies liegt insbesondere daran, dass ein Handlungsgrund, anders als ein kausaler Mechanismus, der einen motorischen Output bewirkt (bzw. die kausale Ursache der Verursachung durch diesen kausalen

Mechanismus) nicht zwangsläufig zu einer Handlung führen muss. Und dies gilt auch dann, wenn man – retrospektiv – einen Grund als für die Ausführung einer Handlung tatsächlich maßgeblich nennt. Denn auch dabei haben wir es nur dann mit einem Grund für eine Handlung zu tun, wenn dieser die Ausführung der Handlung *nicht* im Sinne einer kausalen Zwangsläufigkeit bewirken *musste*. Wäre es nicht so, hätten wir keinen Grund, überhaupt von einer *Handlung* zu sprechen.

Eine behavioristische Terminologie, wie Dretske sie in Anschlag bringt, liefert keinen plausiblen Anhaltspunkt für eine Unterscheidung zwischen Verhalten und Handeln. Es ist doch umgekehrt gerade die Pointe solcher Beschreibungen, diese Unterscheidung letztlich einzuebnen. Die Verwendung des Wortes „Grund“ im Sinne strukturierender Ursachen ist im Hinblick auf Erklärungen natürlicher Vorgänge (etwa warum ein Baum im Herbst seine Blätter verliert oder warum jemand zurückzuckte, als er/sie mit der Hand versehentlich die heiße Herdplatte berührte) sicherlich üblich und unproblematisch. Und auch für Tierverhalten oder menschliche *Verhaltensmuster* mag es Gründe geben, die nicht dem entsprechen, was wir als Gründe für *Handlungen* verstehen. In diesem Sinne von „ein-Grund-sein-für“ können wir jedoch nicht ein Verhalten erklären, wenn wir „Verhalten“ *im Sinne einer Handlung* verstehen.

Gründe lassen uns einerseits *verstehen*, was überhaupt passiert. So ist zum Kühlschrank zu gehen etwas, was wir eine *unvollständige Handlung* nennen können (wie etwa auch, irgendwo einzusteigen). Solche Handlungen machen erst dann wirklich Sinn, wenn sie mit einem weitergehenden Zweck verbunden sind, also beispielsweise als Handlung des Bierholens begründet, bzw. erklärt werden. Man hätte ja auch aus einem *anderen Grund* zum Kühlschrank gehen können, etwa um nachzuschauen, ob es genügend Eiswürfel für die Cocktailparty gibt. Andererseits *rechtfertigen oder legitimieren* Gründe (als *Begründungen*) auch bestimmte Handlungen oder Einstellungen. Im Hinblick auf Verantwortbarkeit im sozialen Sinne etwa hängt die Richtigkeit, Erlaubtheit oder Akzeptabilität von Handlungen natürlich auch von deren Begründung vor einem Hintergrund sozialer Normen ab²⁵⁰.

Weder der Verstehens-, noch der Rechtfertigungsaspekt von Gründen haben etwas mit der Rolle zu tun, die Dretske ihnen zuschreiben möchte, also einer Erklärung von Verhalten durch die Ursache(n) der Verursachung einer Körperbewegung, bzw. *qua* strukturierender Ursache. Etwas vereinfacht gesagt, soll ja, was wir üblicherweise unter dem Verstehen einer Handlung mittels der Kenntnis ihrer ‚Gründe‘ (etwa: „X war der Überzeugung, dass noch ein Bier im Kühlschrank ist“) verstehen, der Kenntnis einer Erklärung eines Verhaltens durch (historische) Tatsachen von der Art entsprechen, dass

²⁵⁰ Beispielsweise ‚darf‘ auch der abstinenten Alkoholiker zum Kühlschrank gehen und Bier holen, *weil* es nicht für ihn, sondern seinen Gast bestimmt ist.

die Verursachung eines bestimmten motorischen Output K durch ein Token eines ‚natürlichen Anzeigers‘ I für eine Eigenschaft E in der Vergangenheit in hinreichend vielen Fällen zur Befriedigung eines natürlichen Bedürfnisses B führte, wodurch das Verhalten $I \rightarrow K$ in Gegenwart von E ‚verstärkt‘ wurde. (bzw. I eine Funktion F erhielt). – Man stelle sich einmal vor, jemand würde auf die Frage, *warum* er/sie eine bestimmte Handlung ausgeführt hat, eine entsprechende Antwort geben! Jedenfalls kann uns Dretske nicht weismachen, er würde damit der Konzeption unser selbst als „human agents“, die die *Autorität* hinsichtlich der Gründe ihres Tuns haben, gerecht. Dieses Verständnis unserer selbst aber ist, wie Dretske selbst feststellt, „something that we human agents will not soon give up“²⁵¹.

Es bleibt, wie ich meine, auch ein Stück weit unklar, wie Dretske eigentlich vom Prozess der innerlichen Verursachung von Körperbewegungen (also Verhalten im Sinne Dretskes) zu Handlungen kommt. Er selbst hatte doch behauptet, dass (typ-) gleiche Körperbewegungen ganz verschiedenen Handlungen entsprechen können (je nach dem, mit welchen Gründen sie ausgeführt werden). Woher aber wissen wir, *dass* wir es überhaupt mit einer *Handlung*, bzw. einem Verhalten, dass durch Gründe zu erklären ist, zu tun haben und mit *welcher* (welchem)? Es scheint hier nicht sehr hilfreich, darauf zu antworten, Handlungen seien eben das, was wir mit Gründen erklären und welche Handlung es sei, hänge von den Gründen ab. Denn was mögliche erklärende Gründe sind, hängt ja wiederum vom Handlungstyp ab. Bereits die symbolische Geste des Abschied-Winkens etwa ist ja nichts, was sich, wie die Flucht vor Fressfeinden, einfach als ein natürliches Verhalten, also unterhalb der Stufe *kultureller Konventionen*, verstehen lässt. Welche Gründe für eine Erklärung oder eine Begründung überhaupt in Frage kommen können, hängt deshalb, zumindest in vielen Fällen, von der *Interpretation* eines Tuns als eine *bestimmte* Handlung *vor dem Hintergrund des jeweiligen sozialen Kontextes* ab. Wie ein solches Interpretieren und Verstehen wiederum auf der Grundlage natürlicher Eigenschaften erklärbar wäre, dazu hat uns Dretske nichts gesagt. Insbesondere verliert er – im Gegensatz zu Ruth Millikan, als eine der anderen wichtigen Vertreter des Teleofunktionalismus – kaum ein Wort über Sprache, sprachliche Bedeutung oder Verständigung und mögliche Zusammenhänge mit dem Erklären oder Verstehen von Handlungen durch Gründe.

Grundsätzliche Bedenken möchte ich zudem bezüglich der von Dretske vertretenen explanativen, bzw. rekonstruktiven Hierarchie anmelden. Gemäß den in dieser Arbeit vertretenen philosophischen Grundannahmen, denke ich nicht, dass propositionaler Gehalt (‚Repräsentation‘) oder Bedeutung unter der Voraussetzung einer Ontologie von

²⁵¹ Dretske (1988); im Vorwort, S. x. Uns als letzte Autorität hinsichtlich unserer Gründe und damit auch des Gehalts unserer Überzeugungen zu verstehen, unterscheidet Dretske übrigens, wie wir noch sehen werden, auch von Millikan.

Dingen oder Eigenschaften erklärt werden kann, sondern dass, umgekehrt, unsere möglichen ‚ontologischen‘ Verständnisse von Kategorien von Dingen oder Eigenschaften auf der Grundlage propositional strukturierter Verständigungsverhältnisse rekonstruiert werden müssen. Da ich darauf noch mehrfach zurückkommen werde – als nächstes im Zusammenhang mit Millikans korrespondenztheoretisch motivierter explanativer Voraussetzung von Tatsachen – werde ich dies allerdings hier nicht diskutieren.

Dretske hat es bewusst vermieden, eine Definition oder wenigstens eine genauere Bestimmung des Funktionsbegriffs vorzuschlagen. Auch dies, sowie insbesondere eine unterschiedliche Einschätzung der explanativen Rolle selektionistischer Erklärungen, unterscheidet seine Position von der Millikans.

Millikan

Für Ruth Millikan sind psychologische Erklärungen, anders als für Dretske, eine Form *biologischer* Erklärungen. Aber konsequenter noch als dieser, betont sie von vorneherein den *normativen Charakter* eines für eine teleologische Erklärung von Verhalten in Anschlag zu bringenden *Funktionsbegriffs*. Ihre grundlegende These ist, dass Bedeutung und Repräsentationalität (bzw. Überzeugen, Wünsche, Gedächtnis, Wahrnehmung, Sprache, zweckgerichtetes Verhalten u. ä.) biologisch-funktionale Kategorien sind und dass Erklärungen in der Biologie üblicherweise einen normativen Begriff *Eigentlicher Funktion* („*proper function*“) verwenden. Eine entsprechende Funktion zu haben, soll heißen, dass ein System oder ein Träger diese *normalerweise* ausführen *sollte* (wobei sie ‚normal‘ in einer spezifischen Weise bestimmt). Die mit dem Begriff der Eigentlichen Funktion verknüpften Redeweisen von ‚gestaltet-sein-für‘ oder ‚eigentlichen Zwecken‘ sind, nach Millikan, im Rahmen einer neo-darwinistischen Theorie *naturalistisch integrierbar*.

Wie wir schon im Abschnitt über die Diskussion um ein angemessenes Verständnis des Funktionsbegriffs im Zusammenhang mit dem Computermodell des Geistes gesehen haben, behaupten die Vertreter eines normativen oder teleologischen Verständnisses des Funktionsbegriffs, dass dieser, zumindest in einer wichtigen Klasse von Fällen (wie etwa der Biologie), nicht im Sinne faktischer Eigenschaften, also durch die aktuellen Fähigkeiten oder Dispositionen von Funktionsträgern oder Systemen, analysierbar ist;– auch nicht im Sinne von statistischen Qualifizierungen, wie einer Mehrzahl oder eines ‚Durchschnitts‘. Worauf es nach Millikan ankommt, ist der

historische Ursprung. Ihre rekursive Definition ‚Eigentlicher Funktionen‘ in *Language, Thought and Other Biological Categories (LTBOC)* ist langwierig und recht komplex. Sie selbst formuliert später eine ‚Grobfassung‘ des zentralen Teils der Definition wie folgt:

„ (...) for an item to have a function *F* as a ‚proper function‘, it is necessary (close to sufficient) that one of these two conditions should hold. (1) *A* originated as a ‚reproduction‘ (to give an example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to the possession of the properties reproduced, have actually performed *F* in the past, and *A* exists because (causally historically because) of this or these performances. (2) *A* originated as a product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under those circumstances, normally causes *F* to be performed by means of producing an item like *A*. Items that fall under condition (2) have ‚derived proper functions‘, functions derived from the functions of the devices that produce them.“²⁵²

Millikan präzisiert diese Definition in *LTOBC* unter anderem durch eine Definition von ‚Reproduktion‘, einer Erläuterung dessen, was genauer unter dem kausal-historisch gemeinten ‚weil‘ („because“) zu verstehen ist und einer Definition von ‚normalen Bedingungen‘. Ich werde versuchen, mich auf das im Hinblick auf unsere Fragen wesentliche von Millikans Überlegungen zu konzentrieren ohne dabei systematische Verkürzungen verschmerzen zu müssen.

Gemäß der zitierten Definition soll, intuitiv gesprochen, die *Eigentliche Funktion* eines ‚Dinges‘ eine derjenigen seiner Eigenschaft sein, von deren *tatsächlichem Wirken* bei den kausal-historischen Vorgängern des ‚Dinges‘ dessen *gegenwärtige Existenz* abhängt. Einfache Fälle von ‚Dingen‘ (wir können auch sagen: Funktionsträger oder Vorrichtungen), die Eigentliche Funktionen besitzen, sind Körperorgane oder instinktives Verhalten. So ist die Blutzirkulation die Eigentliche Funktion des Herzens, weil es die Tatsache, dass Herzen das Blut zum Zirkulieren brachten ist, die für die Replikation von Herzen, also für den Fortbestand dieses Typs von Organ, verantwortlich ist. Es kommt dabei nicht darauf an, dass ein *heutiger* Funktionsträger wirklich in der Lage ist, seine Eigentliche Funktion zu erfüllen; – ja, nicht mal darauf, ob die *Mehrheit* der Funktionsträger tatsächlich in der Lage ist oder war, die Funktion zu erfüllen, bzw. etwas gemäß seiner Funktion auszuführen. Man denke nur an die Milliarden von Spermien, die ihrer eigentlichen Bestimmung, eine Eizelle zu befruchten, nicht nachgekommen sind.

²⁵² R. Millikan: *White Queen Psychology and other Essays for Alice* (1993), S. 13 f.

Ein wesentliches Element von Millikans Definition Eigentlicher Funktionen ist der Begriff der *Normalen Erklärung*²⁵³. Normale Erklärungen erklären die Existenz einer bestimmten Funktion, vereinfacht gesagt, durch eine Antwort auf die Frage, *wie* bestimmte reproduzierte Typen von Funktionsträgern *historisch* eine bestimmte Funktion erfüllt haben. Dabei nimmt Millikan an, dass es eine jeweils *naheliegendste*, bzw. am wenigsten detaillierte Erklärung gibt. Diese beschreibt bestimmte Eigenschaften eines Funktionsträgertyps *FT*, fügt einige Bedingungen, unter denen *FT* *historisch* die entsprechende Funktion *F* tatsächlich ausführte (,Normale Bedingungen‘) und Naturgesetze hinzu und schließt daraus, wie dies alles zur Ausführung von *F* führt.

Mit einer solchen Bestimmung des Funktionsbegriffs scheint das Phänomen der *Fehlfunktion* unproblematisch zu werden: Auch die Augen eines Blinden haben die Eigenschaft, Sehen zu ermöglichen als ihre Eigentliche Funktion; erst dies macht sie überhaupt zu Augen. Sehen zu ermöglichen ist in diesem Sinne der ,von der Natur‘ für *jedes* Auge ,vorgesehene‘ Zweck. Dass dies im Fall von Blindheit gerade nicht gegeben ist, ändert daran nichts. Augen hat man, gemäß dem selektionistischen Erklärungsmuster, überhaupt nur deshalb, weil unsere Vorgänger mit ihnen unter bestimmten Bedingungen tatsächlich sehen konnten, dies einen Vorteil im ,Kampf ums Überleben‘ darstellte und sich dadurch die *Wahrscheinlichkeit der biologischen Reproduktion* erhöhte. In Fällen von Funktionsträgern, die sich phylogenetisch entwickelt haben stellen, nach Millikan, die vorherrschenden *normalen Bedingungen*, also die, unter denen die Funktion *historisch* ausgeführt wurde diejenigen Bedingungen dar, an die der Träger biologisch *angepasst* ist. Die jeweilige neuronale Realisierung der Eigentlichen Funktion von Augen gehöre dabei nicht zur naheliegendsten Erklärung dafür, warum es sie gibt.

Mit dieser Beurteilung des explanativen Stellenwerts selektionistischer Erklärungen – um einen solchen Typ von Erklärung handelt es sich ja bei Millikans ,normalen Erklärungen‘ – und ihrer ,adaptionistischen‘ Interpretation des Funktionsbegriffs steht Millikan in offenem Gegensatz zu Dretske, Sober und Cummins. Denn diese würden, wie wir gesehen haben, die Existenz eines biologischen Funktionsträgers, bzw. die Ausführung einer Funktion durch die Kausalität entsprechender genetischer Voraussetzungen erklären. Nach dieser Sichtweise könnten selektionistische Erklärungen zum Beispiel nur die relative Häufigkeit des Auftretens von Organismen mit Augen erklären.

²⁵³ Den großen Anfangsbuchstaben werde ich auch in der Folge im Zusammenhang mit Millikans Begriff von ,Normal‘ verwenden, um, ebenso wie im Falle von „Eigentlich“ (-e Funktion), auf den spezifischen Charakter von Millikans Verständnis dieser Ausdrücke hinzuweisen. Übrigens redet sie später (1993), auch von ,Normalisierenden Erklärungen‘.

Ich möchte zu dieser (fortdauernden) Auseinandersetzung nicht abschließend Stellung beziehen. Allerdings erscheint mir Millikans Ansatz im Hinblick auf eine naturalistische Analyse von ‚intentionalen Phänomenen‘ *prima facie* erfolversprechender, als die ihrer kausal-funktionalen Konkurrenz.

Das wirklich überraschende an Millikans Vorschlag ist, dass sie den ursprünglich im Hinblick auf biologische Erklärungen eingeführten Begriff der Eigentlichen Funktion auf üblicherweise als nicht-biologisch verstandene Phänomene, wie erlerntes Verhalten, Gewohnheiten, Handeln, sprachliche ‚Mittel‘²⁵⁴ sowie deren organisch realisierte kognitive Grundlagen übertragen will. Auch diese würden nämlich, so Millikan, reproduziert, sie ‚überlebten‘, weil/wenn ihr Wirken in der Vergangenheit für die sich ihrer bedienenden Subjekte vorteilhaft war. Biologisch-funktionale Erklärungen, die in dem Sinne kognitiv sind, dass die jeweiligen Funktionen auf Verhaltensprozesse Bezug nehmen, sind, nach Millikan, Gegenstand der ‚Biopsychologie‘. Mit ihrem Verständnis dessen, was ‚Biopsychologie‘ erklärt, setzt sie sich dabei von stärker reduktiv ambitionierten Varianten naturalistischer Philosophien der Bedeutung und des Geistes ab. So bestimmt Millikan – hierin Dretske nicht unähnlich – den Gegenstand der in biologischen Kategorien gedachten Psychologie nicht im Sinne isolierbarer Dinge, ahistorischer Eigenschaften natürlicher Arten oder statistisch durchschnittlichen Verhaltens, sondern als Stufen fortlaufender Ereignisse, bzw. als das Normale Verhalten von Systemen (Organismen) im Zusammenhang mit ihrer Geschichte und ihrer Umgebung. Zentrales Anliegen ‚biopsychologischer‘ Erklärungen sei es *nicht, Gesetze zu entdecken*, weder universale, noch statistische. Vielmehr gehe es darum, diejenigen Mechanismen zu verstehen, „that contribute to the cyclical processes that constitute development, maintenance, and reproduction for the various species“ (1993; S. 215). Dabei sei der Grad des Scheiterns vieler solcher Mechanismen außerordentlich hoch, weshalb ‚Biopsychologie‘ auch keine Wissenschaft sei, deren explanativer Wert sich an ihrer Vorhersagegenauigkeit bemessen ließe. Zwar glaubt auch Millikan, dass wir das Verhalten von (auch menschlichen) Organismen häufig korrekt vorhersagen können. In Übereinstimmung mit der hier vertretenden Position bestreitet sie jedoch, dass die Grundlage dieser Möglichkeit ein Verhalten verursachender *innerer Mechanismus* sei, der von einer reduktiven Theorie auf der Grundlage einer ‚intentionalen Überzeugung-Wunsch-Theorie‘ zu beschreiben sei. Ihre Gründe hierfür entsprechen im Prinzip jenen, die oben bereits gegen die Möglichkeit einer definatorischen Individuierung propositionaler Einstellung durch (kausal-) funktionale Rollen angeführt wurden: Erstens sei die

²⁵⁴ Millikan benutzt auf Englisch den Ausdruck „*devices*“. Gemeint sind etwa Wörter, Betonungen und syntaktische Formen.

jeweilige physische Konstitution derjenigen Subjekte, die möglicher Gegenstand ‚intentionaler‘ Zuschreibungen sein können zu verschieden, und zweitens würden Normen, die durch Teleo-Funktionen beschrieben werden schon deshalb häufig nicht erfüllt, weil die Bedingungen ihrer Erfüllung untrennbar *mit einer kontingenten Umwelt verkoppelt* sind.

Die Möglichkeit einer mehr oder minder genauer Verhaltensvorhersage – auch beim Menschen – beruht nach Millikan auf Erfahrungen bezüglich bisheriger *Regularitäten im Verhalten* von Individuen und der Gruppe, der das Individuum angehört. Leute einer Kultur bewegten sich, ob nun beim Einkaufen oder etwa im Straßenverkehr, faktisch im Rahmen bestimmter Verhaltensmuster:

„Most people are more likely than not to meet what others consider to be their business and social obligations, to conform to general expectations concerning what is appropriate or seemly and, very important, to do the things they have said they will do. Beyond this, we project ahead patterns observed in the past for particular individuals. Known personality traits, character traits, and habits serve as our guides. Of course, such knowledge merely limits the boundaries of people’s likely behaviour. It does little to nothing toward actually *determining* behaviour in its variety. But seldom do we make an attempt to predict others’ behaviours in much more detail than this.“ (Millikan 1993, S. 225)

Diese Beschreibung der Grundlagen unserer generellen Verhaltenserwartungen wirkt bereits um einiges plausibler, als die von Naturalisten üblicherweise unterstellte Vorhersagbarkeit von Verhalten durch Überzeugungs-Wunsch-Theorien. Allerdings will Millikan die für das Gelingen sozialer Praxis – etwa *qua* Handlungskoordination – erforderlichen Kompetenzen auf Fähigkeiten im Sinne einer „method of brute correlation“ beschränken. Damit unterläuft sie jedoch bereits wieder das konzeptuelle Niveau der von ihr selbst in Anschlag gebrachten normativen Grundlagen wechselseitiger Verhaltens- oder Handlungserwartung, wie etwa das Wissen um geltende soziale Normen (etwa Verpflichtungen) oder die allgemeine Unterstellung zu erwartenden sinnvollen, bzw. vernünftigen Handelns. Denn selbst wenn man die Normativität wechselseitiger Verhaltenserwartung in naturalistischen Begriffen, also etwa behavioristisch, im Sinne sanktionsgestützter positiver oder negativer Verstärkungen von Verhaltenserwartungen, formuliert, hat man es ja nicht mit wechselseitigen Beschreibungen auf der Grundlage ‚bloßer Korrelationen‘ oder Regularitäten zu tun.

Ähnlich wie Dretske, setzt sich Millikan aber von dem unter Naturalisten üblichen *Credo* ab, Gründe als (‚auslösende‘) Ursachen von Verhalten oder Handlungen zu verstehen. Sie stellt fest, dass „(...) reasons cannot be, as such, causes“ (a.a.O., S. 226), da eine Funktion zu haben grundsätzlich ein ‚kausal impotentes Faktum‘ bezüglich

eines Dinges sei. Dies erscheint einleuchtend. Denn dass etwas eine bestimmte Funktion hat, kann nicht die *Ursache* dafür sein, dass es bestimmte Dinge tut. Auch nach Millikan hat etwas eine Funktion, weil es eine bestimmte Geschichte hat, und dies ist offensichtlich keine Eigenschaft, der eine bestimmte ‚kausale Kraft‘ in direkter Weise zukommt. Diese Geschichte soll dabei aber – anders als bei Dretske – *auch* im Fall von Normalen Erklärungen im Zusammenhang mit ‚propositionalen Repräsentationen‘ (Überzeugungen) eine *phylogenetische* sein können. Jedenfalls müssten Erklärungen von Verhalten im Sinne der Zuschreibung ‚intentionaler Zustände‘ eher als eine Subsumtion von Verhalten unter *biologische Normen*, als unter kausale Gesetze verstanden werden²⁵⁵.

Dies bedeutet natürlich nicht, dass Millikan teleofunktionale Verhaltens- bzw. Handlungserklärungen als Erklärungen außerhalb eines Rahmens kausaler Ordnung verstanden wissen will. Ihrer Ansicht nach ‚umschreiben‘ („*circumscribe*“) die mit teleofunktionalen Begriffen arbeitenden Normalisierenden Erklärungen bestimmte, auf Ursachen ausgerichtete physische Erklärungen. Eine erschöpfende ‚Analyse‘ hinsichtlich der Art und Weise, in der ein Funktionsträger, gegeben seine spezifische Geschichte, Normalerweise arbeiten würde schließlich die Ebene erreichen, auf der dessen normale physische Struktur und die normalen physischen Bedingungen so beschrieben werden können, dass eine kausale Erklärung des Wie der Ausführung der Funktion nach Naturgesetzen möglich würde.

Ich hatte bereits auf die hingewiesen, dass Millikan durch die Art und Weise, in der sie Sprache und ihre Funktionen thematisiert, eine gewisse Sonderstellung innerhalb des naturalistischen Lagers einnimmt. Weder sei, so Millikan, Satz- oder Wortbedeutung auf der Grundlage von Sprecherbedeutung zu analysieren, noch sei das Umgekehrte möglich. Dementsprechend wäre die Bedeutungshaftigkeit von Sätzen ohne Bezugnahme auf den Umstand zu beschreiben, dass Sätze typischerweise gebraucht würden, um Gedanken auszudrücken und zu übermitteln. Andererseits könnte aber auch ohne eine Bezugnahme auf Sprache erklärt werden, was Überzeugungen, Wünsche oder ‚Intentionen‘ sind. Gegen einen Griceschen Bedeutungsintentionalismus – und damit im Widerspruch zum innerhalb des naturalistischen Lagers dominierenden

²⁵⁵ Im einfachsten Fall könnte ein Phänomen funktional schlicht dadurch erklärt werden, dass es als das Resultat von etwas bestimmt wird, dessen Eigentliche Funktion darin besteht, dieses Phänomen hervorzurufen. Entsprechende Erklärungen bestehen aber etwa auch im Aufzeigen des Ortes, den ein Ereignis in einer Serie von in einer funktionalen Beziehung stehenden Elementen einnimmt.

IBS-Programm – betont Millikan in *LTOBC* die *Unabhängigkeit der Funktionalität* (bzw. der ‚Intentionalität‘) *sprachlicher Mittel*:

„What makes conventional signs special is that they have their very own proper functions, quite apart from how they happen to be handled by particular speakers and hearers on particular occasions“ (S. 70)

Zur Erläuterung des Stellenwertes einer von den jeweiligen Sprecherintentionen unabhängigen konventionellen Ausdrucksbedeutung verwendet sie den Begriff der ‚*stabilisierenden*‘ oder – spezifiziert auf Sprache – ‚*standardisierenden Eigentlichen Funktion*‘:

„The stabilizing and standardizing proper function (or functions) of a language device is that hypothesized function (or functions) that tends at the same time to keep speakers using the device in standard ways and to keep hearers responding to it in standard ways, thus stabilizing its function or functions“ (S. 31 f.)

„Just as speakers will continue to speak only if there is some correlation between what they say and how hearers respond, so hearers will continue to listen and to follow certain patterns of translation into understanding and sometimes into belief or intention and action only if these responses are of some use in relation to the hearer’s own projects in the world – of use at least in some critical proportion of cases. And if hearers stop listening or stop responding in ways that show some uniformities, speakers in turn will stop speaking. Hence one of the proper functions of the various language devices must be to do things that keep hearers responding in the old ways with some degree of uniformity.“ (S. 31)

Die Grundidee ist also, dass wenn ein sprachliches Mittel nicht in hinreichend vielen Fällen seine Funktion derart erfüllte, dass einerseits Sprecher durch dessen Aktualisierung gegenüber (willkürlichen) Hörern, und andererseits Hörer durch ihre Reaktion auf dessen Aktualisierung durch (willkürliche) Sprecher, jeweils ihre Zwecke erfolgreich realisierten, keine Motivation bestünde, sich dieses Mittels erneut in gleicher Weise zu bedienen, bzw. entsprechend zu reagieren. Ohne stabilisierende Eigentliche Funktion würde ein sprachliches Mittel „simply die out“ (ebd.). Die Idee der stabilisierenden oder standardisierenden Eigentlichen Funktion erscheint damit gleichsam als eine naturalisierte Version des Konzepts eines *Sprachgebrauchs* im Sinne einer Praxis gemeinsamer, institutionell standardisierter Regolorientierung.

Direkte Eigentliche Funktionen sind für Millikan solche, die reproduziert werden. Und *sprachliche* Mittel könnten nur reproduziert werden, *weil* sie stabil sind. Mechanismen mit sich reproduzierenden Funktionen können jedoch, nach Millikan, allgemein so beschaffen sein, dass sie sich nur in einem, an bestimmte

Umweltbedingungen angepassten, Verhalten mit *abgeleiteter* Eigentlichen Funktion als funktional *aktualisieren*. Dementsprechend könnten, im Fall der menschlichen Sprache, Äußerungen eine von der direkten Eigentlichen Funktion *abweichende* abgeleitete Eigentliche Funktion (bzw. ‚Bedeutung‘) in Form von spezifischen Intentionen haben, mit denen Sprecher sich sprachlich äußern. Dies sei jedoch nur aufgrund der Tatsache möglich, dass *standardisierte* Bedeutungen als *direkte* Eigentliche Funktionen gegeben sind, also dafür gesorgt ist, dass die sprachlichen Mittel überhaupt reproduziert werden. Alle neuen, das heißt nicht-reproduzierten, Sprachverwendungen hätten dementsprechend *abgeleitete* Eigentliche Funktionen.

Die Art der Funktionalität sprachlicher Standardisierung unterscheidet sich dabei, so Millikan, im Hinblick auf die *Modi* sprachlicher Äußerungsmöglichkeiten: So sei die stabilisierende Funktion des *imperativischen* Modus, *Befolgung hervorzurufen*, die des *indikativischen* Modus, Informationen zu übermitteln, also *wahre Überzeugungen* hervorzurufen. Darauf möchte ich hier nicht weiter eingehen. Wichtiger für uns sind andere Fragen. Denn die stabilisierende Eigentliche Funktion macht für Millikan nur *einen*, nämlich den *pragmatischen Aspekt* der Bedeutung öffentlicher Sprachen aus. Der *zweite Aspekt* soll die Funktion betreffen, die wir normalerweise unter Rubriken, wie ‚*semantischer Wert*‘, ‚*intentionaler*‘ oder ‚*propositionaler Gehalt*‘ beschreiben. Es ist dieser zweite Aspekt, auf den ich mich in der Folge konzentrieren möchte²⁵⁶.

Lässt sich propositionaler Gehalt und sprachliche Bedeutung auf der Grundlage einer evolutionistisch begründeten Korrespondenzbeziehung analysieren?

Millikan wählt den Begriff des ‚*intentionalen Ikons*‘ zur Bezeichnung all dessen, was den allgemeinen Bezugspunkt ihres teleofunktionalistischen Explikationsversuchs von ‚Bedeutung‘ und ‚Intentionalität‘ bildet. Darunter soll all dasjenige fallen, was im weitesten Sinne als *Zeichen* gedeutet werden kann, wie etwa (nicht-natürliche) Signale, Elemente von Sprachsystemen (z.B. Sätze), aber auch ‚innere‘ Repräsentationen. Und obgleich Millikan ‚Intentionalität‘ nicht für ein eindeutig definierbares Phänomen hält, bestimmt sie als den wesentlichen Zug aller Formen intentionaler Ikonen die Funktion, *der Welt* in einer bestimmten Weise *entsprechen zu sollen*. Dies können wir für den uns

²⁵⁶ Millikan entwickelt in *LTOBC* ein recht komplexes Netzwerk explizierender Erläuterungen und Definitionen, dem ich hier – im Rahmen meiner allgemeinen Naturalismuskritik – sicher nicht voll gerecht werden kann. Andererseits gilt hier wir überall, dass man eine gewisse Distanz zum technischen Begriffsapparat halten muss, wenn man sich auch mit den grundlegenden Annahmen der jeweiligen Positionen auseinandersetzen will.

interessierenden Fall ‚propositionaler Intentionalität‘ vorläufig (und vereinfachend) auch so ausdrücken: Nur weil indikativische Sätze oder Überzeugungen *Normalerweise* der Welt in einem bestimmten Sinne *korrespondieren* – zu korrespondieren also ihre Eigentlichen Funktion ist – sind sie überhaupt für die *weitergehenden Zwecke* von Sprachverwendern, bzw. Überzeugungsträgern funktional. Wäre eine solche Korrespondenz nicht der Normalfall, würden Sprachverwender, bzw. Überzeugungsträger nicht überleben und könnten sich dementsprechend auch nicht fortpflanzen. Damit aber würden natürlich auch die entsprechenden ‚intentionalen Mittel‘ nicht reproduziert. Allerdings geht Millikan in *LTOBC* einen langen theoretischen Weg von der Definition Eigentlicher Funktionen bis zur ihrer naturalistischen Bestimmung propositionaler ‚Intentionalität‘. Ich werde zunächst in aller Kürze versuchen, die wesentlichen Stationen dieses recht komplizierten Weges möglichst ‚untechnisch‘ nachzuzeichnen.

Ein erster wichtiger Schritt ist die Einführung des Begriffs der *relationalen* Eigentlichen Funktion. Er betrifft jene Fälle, in denen eine Funktion Normalerweise im Hervorbringen von etwas besteht, das in einer bestimmten Relation zu etwas in der Umgebung des Systems steht. Die je situative Realisierung einer relationalen Eigentlichen Funktion, also etwa der konkrete Fall einer Anpassung der Pigmentfarben eines Chamäleons an die Farben seiner Umgebung, nennt Millikan eine *adaptive* Funktion. Adaptive Eigentliche Funktionen sind, im Gegensatz zu relationalen, *abgeleitete* Eigentliche Funktionen, da sie sich nicht reproduzieren (was ja ein Kriterium für *direkte* Eigentliche Funktionen sein soll). Sie können gegebenenfalls sogar ein einmaliges historisches Ereignisse sein. Alles, was individuell erlernt wird, fällt, nach Millikan, unter adaptive Funktionen.

Unter ‚intentionalen Ikonen‘ versteht Millikan dann direkte relationale oder abgeleitete Eigentliche Funktionen, die (Normalerweise) zwischen zwei *kooperierenden* Vorrichtungen (*devices*), nämlich einer *produzierenden* und einer ‚interpretierenden‘, stehen. Diese seien durch genetische Determination oder Lernprozesse so gestaltet oder standardisiert, dass sie zueinander passen, also die Anwesenheit und Kooperation des einen eine Normale Bedingung für das Eigentliche Funktionieren des anderen ist. Im Gegensatz etwa zu den Vorrichtungen zur Produktion und zur ‚Interpretation‘ von Bientänzen wären die entsprechenden Bientänze selbst adaptiv. Und im Fall propositional-sprachlicher Sätze müsste dann gelten, dass schon die entsprechenden Mechanismen selbst an ein jeweiliges sprachliches Umfeld *angepasst* sind, da sie erst durch standardisierende *Lernprozesse* ihre (relationale) Eigentliche Funktion erhalten. – Doch so weit sind wir noch nicht. (Bei späterer

Gelegenheit²⁵⁷ vermeidet Millikan übrigens die fragwürdige Redeweise von ‚Interpretation‘ und redet stattdessen von ‚Konsumption‘).

Zunächst bestimmt Millikan in einem weiteren Schritt *Repräsentationen* als all jene Formen ‚intentionaler Ikonen‘, die in mittelbaren *Inferenzen* fungieren können. Aber auch Konsumenten von Repräsentationen versteht Millikan noch als Vorrichtungen, die durch einen *Prozess der natürlichen Auslese* so gestaltet wurden, dass sie mit gewissen Produzenten von Repräsentationen kooperieren – wobei diese Rollen im Fall der Verständigung durch Sprache auf verschiedene Individuen verteilt sind. Die Produzenten von Repräsentationen seien ihrerseits durch natürliche Auslese daraufhin gestaltet, Repräsentationen für einen Konsumenten zu erzeugen, welche den Bedingungen in der Welt nach einer *Korrespondenzregel* entsprechen (was aber nicht ausschließe, dass Repräsentationen auch durch Kombinationen anderer Repräsentationen, gemäß bestimmter Inferenzregeln, erzeugt werden können). „Korrespondenz“ bedeutet für Millikan eine umkehrbar eindeutige Funktion im mathematischen Sinne, bei der die Relata beider Seiten einer entsprechenden Beziehung in systematischer Weise kovariieren. Eine solche Regel sei es, die in der naheliegendsten Normalen Erklärung des praktischen Erfolgs eines Konsumenten zum Zuge komme. Es ist demnach eine *Normalitätsbedingung* für die Eigentliche Funktion eines Konsumenten, dass die konsumierten Repräsentationen dem, was sie repräsentieren, *gemäß einer Korrespondenzregel entsprechen*. Und dies bedeutet für Millikan, dass eine Repräsentation nur dann im genannten Sinne funktional ist, *wenn sie wahr ist*.

Insbesondere gegen Fodor und die Vertreter des oben diskutierten Funktionalismus kausaler Rollen besteht Millikan darauf, dass es der Aspekt des Konsumierens (‚Benutzens‘) der Repräsentation *als* Repräsentation ist – *nicht* die Art und Weise ihres Zustandekommens, der sie einerseits überhaupt erst zu Repräsentationen macht, andererseits aber auch ihren Inhalt bestimmt. Obgleich Repräsentationen immer durch Systeme *produziert* würden, deren Eigentliche Funktion es sei, solche Repräsentationen hervorzubringen, die der Welt nach gewissen Regeln entsprechen, so sei doch die Tatsache, *dass* es sich um eine nach Korrespondenzregeln gebildete Repräsentation handelt, aber auch, *was* diese Regel *ist*, vollständig durch den Konsumenten (bzw. Benutzer) der Repräsentation bestimmt.

Die Fähigkeit zu Produktion und Konsumption *propositional strukturierter* ‚Intentionalität‘ soll sich nun in mehrfacher Hinsicht von der Fähigkeit, lediglich einfache Repräsentationen zu produzieren oder konsumieren, unterscheiden. Dazu gehören, so Millikan, unter anderem die Differenzierbarkeit indikativischer und

²⁵⁷ R. Millikan: ‚Biosemantics‘ (1989).

imperativischer Modi, inferentielle Fähigkeiten, die Fähigkeit zur situationsübergreifender ‚Ding-Identifizierung‘, sowie Negierbarkeit bzw. Widersprüchlichkeit. Diesen Kriterien kann ich uneingeschränkt zustimmen. Millikan versteht die entsprechenden Fähigkeiten allerdings im Sinne natürlicher Eigenschaften von Organismen, bzw. als funktionale Eigenschaften natürlicher Organe, deren Struktur sie in beeindruckender Weise in den Teilen III und IV von *LTOBC* ausführlich beschreibt. Darauf kann ich hier nicht im Einzelnen eingehen. Stattdessen werde ich meine kritische Auseinandersetzung mit Millikans Position auf die grundsätzliche Frage nach dem explanativen Stellenwert ihres korrespondenztheoretischen Realismus konzentrieren. Dabei werde ich die These verteidigen, dass die genannten Kriterien überhaupt erst im Rahmen einer normativ-pragmatistischen Analyse der Bedingungen und Eigenschaften propositional-sprachlicher Verständigung erläutert werden können. *Zu verstehen, was es heißt, die natürliche Welt zu beschreiben, ist abhängig vom Verstehen der normativ-pragmatischen Struktureigenschaften unserer kognitiven Praktiken. Wir können also unsere kognitiven Praktiken (bzw. Fähigkeiten) nicht umgekehrt auf der Grundlage von (empirischen) Theorien über die natürliche Welt erklären.*

Um hier nochmals abzukürzen, bediene ich mich eines Zitats, in dem Millikan später, unter weitgehendem Verzicht auf den in *LTOBC* entwickelten technischen Begriffsapparat, wesentliche Züge ihrer Idee einer biologistischen Reformulierung der Korrespondenztheorie der Wahrheit wie folgt erläutert:

„Cognitive Systems are designed by evolution to make abstract pictures of the organism’s environment and to be guided by these pictures in the production of appropriate actions. This may involve deriving additional pictures by combining old ones in accordance with certain principles (inference, reasoning). It may also involve the formation of goal representations, pictures that guide the organism to produce or avoid what they picture (imagining, fearing that, desiring, intending). In the human case, biological design includes very general principles in accordance with which concepts are formed from appropriate sensory input, including input from perception through language. Forming adequate concepts is learning to represent or map in thought what *is* the same *as* the same. On this ability are built the abilities to make judgements, then appropriate inferences and also abilities to acquire appropriate desires and get them fulfilled. Like anything *else* the organism does, this all takes place in a highly principled way when it works, but, of course, it does not always work, or if it does, not always smoothly.“ (Millikan 1993, S. 11)

Der letzte Satz des Zitats gibt auch bereits einen Hinweis darauf, dass die korrespondenztheoretische Analyse der Funktionalität von Repräsentationen auch eine Lösung für den Problemzusammenhang von *möglichen Fehlrepräsentation* und

Unterbestimmtheit liefern soll. Der Inhalt oder ‚Sinn‘ einer Repräsentation soll ja indirekt durch eine Korrespondenzregel bestimmt sein, die den historischen, die Reproduktion ermöglichenden Erfolg des Konsumenten der Repräsentation Normal erklärt. Der *Bezug* einer Repräsentation (bzw. eines ‚intentionalen Ikons‘) soll dabei dem entsprechen was sie, gemäß einer solchen Erklärung, historisch *tatsächlich* abgebildet hat. Und es ist ein *bestimmter* Typ von Zuständen oder Ereignissen, unter Bezugnahme auf welchen die Reproduktion der Repräsentation Normal erklärt werden kann. Die dem Frege'schen *Sinn* entsprechende Eigentliche Funktion sei es dasjenige abbilden zu *sollen*, was den *historischen Bezug* der Repräsentation bildet. Eine Fehlrepräsentation (eine falsche Überzeugung, ein falscher Satz) tritt demnach dann auf, wenn diese Funktion insofern nicht erfüllt wird, als eine Repräsentation durch etwas anderes verursacht wird, als durch den Typ von Ereignissen, die den historisch begründeten Bezug bilden.

Dieser Gedanke lässt sich gut an unserem oben bereits diskutierten Beispiel der Bakterien mit ihren sich am Magnetfeld orientierenden Magnetosomen veranschaulichen: Die Indikatoren eines solchen Bakteriums würden nämlich nur unter einer Bedingung *dysfunktional* (im Sinne Millikans), nämlich dann, wenn sie nicht mehr die *Richtung sauerstoffarmen Wassers* anzeigen. *Wie* sie dies erreichen, spielt für den, im Rahmen der *nächstliegenden* Erklärung des biologischen Erfolges der Bakterien anzunehmenden, funktionalen Beitrag des Subsystems „Magnetosom“ keine Rolle. Gemäß einer solchen Erklärung repräsentierten die Magnetosome nicht etwa den jeweiligen magnetischen Pol, sondern den Ort sauerstoffarmen Wassers, denn dies sei es, was ihre Existenz erklärt.

Wenn man – *for the sake of the argument* – mal die naturalistische Unterstellung mitmacht, begrifflicher, bzw. propositionaler Gehalt könne überhaupt funktionalistisch ‚erklärt‘ oder ‚analysiert‘ werden, scheint Millikan damit eine durchaus plausible Lösung der genannten Probleme anbieten zu können. Allerdings hat ein solches ‚ätiologisches‘ Verständnis des Funktionsbegriffs, wie es Millikan (und in ähnlicher Weise auch Dretske) vorschlägt, starke Konsequenzen, deren vermeintliche Kontraintuitivität denn auch von ‚kausalistischen‘ Kritikern des Teleofunktionalismus ins Feld geführt wird²⁵⁸. So könnten, nach Millikans Verständnis, Repräsentationen selbst dann, wenn sie organismusintern jeweils auf *genau dieselbe* Weise physisch realisiert wären, in verschiedenen Spezies – mit abweichender Evolutionsgeschichte – einen *unterschiedlichen* Gehalt haben. Und umgekehrt könnten Repräsentationen mit demselben Gehalt (bzw. derselben Funktion) selbst innerhalb einer Spezies auf sehr unterschiedliche Art und Weise physisch realisiert sein. – Ist dies plausibel? Stellen wir

²⁵⁸ S. etwa Cummins (1989), S. 80 f.

uns etwa vor, durch einen gigantischen kosmischen Zufall hätte sich eine Anzahl zuvor willkürlich umherschwirrender Moleküle genau so zusammengesetzt, dass plötzlich ein exakter physikalischer Doppelgänger einer bestimmten Person entstünde²⁵⁹. Hätte dieser Doppelgänger die gleichen Repräsentationen (bzw. propositionalen Einstellungen) wie das Original? Hätte er überhaupt Repräsentationen? Nach der ätiologischen Grundidee, dass es bei Funktionen – und damit auch für den Gehalt von Repräsentationen – auf die phylogenetische und/oder ontogenetische *Geschichte* ihrer Träger ankomme, kann dies nicht der Fall sein. Nichts an einem solchen Doppelgänger kann eine Eigentliche Funktion im Sinne Millikans haben, da, wie sie bereits in *LTOBC* (S. 93 f.) betont, dessen Evolutions- oder Lerngeschichte einfach die *falsche* wäre. Dies ist tatsächlich eine starke Annahme, die offenbar nur vor dem Hintergrund eines robusten *Realismus* im Hinblick auf *Funktionen* im Allgemeinen und Repräsentationen (bzw. propositionale Einstellungen) im Besonderen Sinn macht. Jemand mit eher reduktiv-physikalistischen Intuitionen wird ihr sicherlich nicht zustimmen können. Jedenfalls steht Millikan damit in eindeutigem Widerspruch zu computationalen bzw. kausal-funktionalistischen Theorien propositionaler Einstellungen, für die sowohl die individuelle, als auch die Stammes-Geschichte kognitiver Eigenschaften keine (wesentliche) Rolle spielen *kann*.

Der Einwand hat aber auch noch eine erkenntnistheoretische Pointe. Denn unsere funktionalen Erklärungen setzen ja im Allgemeinen mit der Beobachtung von Kovarianzen an: Wir stellen fest, dass bestimmte Dinge oder Ereignisse in einem kausalen Zusammenhang stehen (ohne Herztätigkeit, keine Blutzirkulation). Dabei ist die Evolutionsgeschichte allerdings selten schon bekannt. Dies lässt sich, wie Cummins bemerkt²⁶⁰, auch auf die Funktion von Repräsentationen übertragen. Bei der Interpretation der ‚Bedeutung‘ eines Bientanzes etwa beginnen wir nicht mit der Selektionsgeschichte der sie hervorbringenden Mechanismen; eher erschließen wir sie durch die Beobachtung von Kovarianzen zwischen Verhaltensmustern und Umgebung und können *dann* ihre Selektionsgeschichte rekonstruieren. Wir wissen, dass Bienen normalerweise Blumen finden, wenn sie in Reaktion auf Bientänze losfliegen. Und wir wissen auch, dass Bienen Blumen zum Überleben brauchen. Letzteres bestärkt uns sicherlich noch in der Annahme, dass der Tanz den Bienen den Weg zu den Blumen weisen soll. Dennoch,

²⁵⁹ In der englischsprachigen Literatur wird ein solcher Doppelgänger – im Anschluss an ein Beispiel Davidsons – häufig als „*Swampman*“ bezeichnet (vergl. Davidson: ‚Knowing one’s own Mind‘ (1987)).

²⁶⁰ Cummins (1989), S. 85 f.

„(...) if it always happened that spectator bees flew off in the direction of the dance orientation, found a pile of rocks, milled around, then went home, we would, I think be justified in attributing the content *rocks in the direction of the dance orientation*, even though we would be mystified about the evolutionary significance of the whole business. It seems pretty clear that speculation about the evolutionary history (and even the learning history) of central cognitive mechanisms will be possible only after we have a pretty good idea what representations are actually required.“ (Cummins 1989, S. 85 f)

Cummins Einwand läuft darauf hinaus, dass *repräsentationale* Eigenschaften sowohl in *epistemologischer*, als auch in *explanatorischer* Hinsicht *grundlegender* sind, als *adaptionale*²⁶¹ (bzw. durch-Anpassung-einen-Vorteil-verschafft-habende) Eigenschaften. Und tatsächlich scheint es einleuchtend, dass wir die adaptionale Bedeutung einer kognitiven Funktion nicht verstehen können, bevor wir wissen, welche Bedeutung, welchen repräsentationalen Gehalt sie hat. Wie sollten wir die adaptionale Rolle einer Repräsentation überhaupt erklären können, ohne bereits zu *unterstellen*, dass es sich um eine solche, also eine semantisch zu charakterisierende Funktion handelt – und ohne zu *wissen*, um *welche*? Allerdings müssen wir wiederum Cummins entgegenhalten, dass auch seine Alternative wenig plausibel ist: Denn wären die von ihm vertretene *dispositional-funktionale* (also ahistorische) Erklärung von ‚Intentionalität‘ einer selektionistischen oder lerntheoretischen Erklärung vorzuordnen, sähe sich diese ja zumindest wieder mit den Problem der Unterbestimmtheit und dem Problem der Möglichkeit von Fehlrepräsentationen konfrontiert – die zu vermeiden Millikans Ansatz immerhin zu versprechen schien.

Ein häufig gegen evolutionistische Theorien des Geistes ganz allgemein formulierter Einwand lautet, dass selektionistische Erklärungen von ‚Intentionalität‘ spätestens dann jede Plausibilität verlören, wenn es sich nicht mehr um Fälle so einfacher, praktisch relevanter ‚Repräsentationen‘ handelte, wie jene, mit denen wir es in den üblicherweise angeführten Beispielen zu tun haben. Überzeugungen über den Ursprung des Universums etwa oder, dass Hermann Buhl der Erstbesteiger des *Nanga Parbat* war dürften, auch wenn sie wahr sind, kaum je für das historische Überleben und den reproduktiven Erfolg eines Organismus relevant gewesen und damit reproduziert worden sein. Dies an sich schon als einen starken Einwand gegen Millikan zu verstehen²⁶¹, wird allerdings dem theoretischen Niveau ihrer Position nicht gerecht, da sie solche Fälle ja ebenso als *abgeleitete* Eigentliche Funktionen deuten kann, wie etwa auch indexikalische oder andere ‚einmalige‘ Überzeugungen, Überzeugungen auf der Grundlage neu gebildeter Begriffe, idiosynkratisch verwendete Sätze ü. a.m. Und es

²⁶¹ So etwa in W. Lyons: *Approaches to Intentionality* (1995), S. 75 f; und bei A. Beckermann (1999), S. 337.

ist, wie wir gesehen haben, nur die *direkte* Eigentliche Funktion, von der Millikan behauptet, sie werde *reproduziert*. Die in Frage stehende *direkte* relationale Eigentliche Funktion der Produzenten von Repräsentationen soll ja lediglich darin bestehen, allgemein *solche hervorzubringen*, die gemäß einer Korrespondenzregel wahr sind. Wahre Überzeugungen mit *abgeleiteten* Eigentlichen Funktionen könnten dann nicht nur Folgerungen aus Überzeugungen mit *direkten* Eigentlichen Funktionen sein, sondern auch das Resultat sowohl individueller, als auch gesellschaftlicher *Lernprozesse* auf der *Grundlage* evolutionär gestalteter Mechanismen ihrer Produktion und Konsumption. Demnach müssten zwar die *Mechanismen* (bzw. die kognitiven *Voraussetzungen*), die es *ermöglichen*, dass wir nicht nur neue Überzeugungen oder Sätze, sondern darüber hinaus auch *neue Begriffe* erlernen oder bilden können, durch ihre Selektionsgeschichte erklärt werden können, *nicht* aber die *einzelnen* neuen Überzeugungen, Sätze oder Begriffe *selbst*:

„Unlike evolutionary adaption, learning is not accomplished by *random* generate-and-test procedures. Even when learning involves trial and error (probably the exception rather than the rule), there are principles in accordance with which responses are selected by the system to try, and there are specific principles of generalisation and discrimination, etc., which have been built into the system by natural selection. How these principles normally work, that is, how they work given normal (i.e. historically optimal) environments, to produce changes in the learner’s nervous system which will effect the furthering of ends of the system has, of course, an explanation – the normal explanation for proper performance of the learning mechanism and of the states of the nervous system it produces.“
(Millikan 1989, S. 292)

In dieser Allgemeinheit formuliert scheinen mir dies plausible Annahmen zu sein. Es spricht manches für die Richtigkeit einer solchen Form selektional-evolutionärer Erklärung der *Entwicklung* unserer kognitiven *Vermögen*, wie etwa der Lernfähigkeit. Und wir können Millikan auch darin zustimmen, dass auch solche kognitiven Fähigkeiten in einem adaptionistischen Sinne teleofunktional beschrieben werden können, die selbst nicht direkt aus einer (darwinistischen) Selektionsgeschichte ‚herleitbar‘ sind, aber auf der Grundlage von *organischen Mechanismen* hervorgebracht werden, bei denen dies der Fall ist²⁶². Doch es ist eine ganz andere Frage, ob man damit auch schon eine naturalistische Analyse, bzw. Integration, von Geist und Bedeutung geleistet hat, also insbesondere die *Frage*, worin Propositionalität besteht – *was es heißen kann, dass etwas einen propositionalen Gehalt hat oder zum*

²⁶² Dies hat etwa Fodor bei seiner Kritik an Millikan nicht berücksichtigt; s. ders. (1987), S. 116-17.

Ausdruck bringt – naturwissenschaftlich beantwortet hat. Davon kann wie ich meine keine Rede sein.

Um dies zu sehen, müssen wir uns zunächst vergegenwärtigen, welchen Sinn die Rede von Korrespondenz im Zusammenhang mit propositionalen Phänomenen überhaupt haben kann. Wenn wir etwa sagen, dass eine Beschreibung den Tatsachen entspricht oder korrespondiert, so ist dies doch eigentlich nur eine Weise, zum Ausdruck zu bringen, dass man sie für wahr hält. Was aber sind Tatsachen, und was heißt es, dass etwas wahr ist? – Ich denke die einzig plausible Antwort hierauf kann nur lauten, dass Tatsachen nichts anderes sind, als das, was wahren Überzeugungen entspricht oder durch wahre Sätze (bzw. Behauptungen) zum Ausdruck gebracht wird. Und umgekehrt ist eine Überzeugung dann wahr oder bringt ein Satz dann etwas Wahres zum Ausdruck, wenn sie, bzw. er den Tatsachen korrespondiert. Welche Überzeugungen oder Sätze wahr sind, was die Tatsachen sind, das ist natürlich eine interessante Frage. Die Antwort darauf gibt jedoch wie ich meine keine Korrespondenzbeziehung.

Wenn es richtig ist, dass wahre Überzeugungen oder Sätzen und Tatsachen in diesem Sinne intern aneinander gekoppelt sind, kann diese Beziehung keiner *explanativ* hilfreichen Korrespondenztheorie auf die Beine helfen. Es ist schon häufiger darauf hingewiesen worden, dass es in metaphysische Verwirrungen führt, Tatsachen als ‚Wahrmacher‘, bzw. als irgendeine Art von an sich bestehenden Entitäten (in der Welt) zu verstehen, denen Aussagen korrespondieren, wenn sie wahr sind²⁶³. Doch obwohl Millikan die Korrespondenzbeziehung als eine zwischen Repräsentationen und Verhältnissen in der Welt versteht – nicht etwa als eine zwischen Elementen einer Repräsentation und Dingen oder Eigenschaften –, kann man ihr, wie wir gleich noch sehen werden, nicht den Vorwurf machen, Tatsachen in diesem Sinne zu ‚verdinglichen‘. Zudem vertritt auch sie die Ansicht, ‚pure‘ Korrespondenztheorien der Wahrheit seien ‚leer‘: „If any correspondence theory is to avoid vacuousness, it must be a theory that tells what is *different* or *special* about mapping relations that map representations onto representeds“ (*LTOBC*, S. 86 f). Dieser Unterschied bzw. diese Besonderheit könne aber nicht rein formal, sondern nur durch empirische Theorien über kausale Zusammenhänge erläutert werden, also durch Physik, Physiologie, Biologie, bzw. ‚Biopsychologie‘ und Evolutionstheorie:

„The Position is that intentionality is grounded in external natural relations, Normal and/or proper relations, between representations and representeds, the

²⁶³ *Locus classicus* der neueren Debatte ist hier sicherlich P. F. Strawsons ‚Truth‘ (1950); ähnlich Einflussreich später aber auch Davidsons ‚A Coherence Theory of Truth and Knowledge‘ (1986).

notions „Normal“ and „proper“ being defined in terms of evolutionary *history* – either of the species or the evolving individual or both“ (LTBOC S. 93)

Es stellt sich aber die Frage, ob damit schon viel gewonnen ist. Denn auch wenn man, wie Millikan, anstatt eine Ontologie von Tatsachen als ‚Wahrmacher‘ gutzuhießen, im Grundsatz die Sache so versteht, dass *was* eine Repräsentation wahr macht, (im einfachsten Fall) eine, in Begriffen evolutionärer Historie zu spezifizierende, externe natürliche Relation zu einer entsprechenden Tatsache, *dass* einem Ding oder einer Substanz eine bestimmte Eigenschaft zukommt sein soll, hätte man der Rede von Korrespondenz doch nur dann irgend einen substantiellen Sinn gegeben, wenn dabei Dinge, Substanzen, Relationen und Eigenschaften irgendwie als etwas von den Bedingungen des Habens von Überzeugungen oder des Machens und Verstehens von Aussagen *Unabhängiges verständlich* gemacht werden kann. Man müsste also eine Welt intrinsisch individuierter Dinge oder Substanzen mit intrinsischen Eigenschaften oder intrinsisch relationiert unterstellen, und zwar in einem Sinne von Dingen, Substanzen, Eigenschaften und Relationen, der ungefähr üblichen Formen von Gegenstands- oder Substanzidentifizierung und von Klassifizierungen entspricht, damit ‚Repräsentationen‘ durch die Verhältnisse in der Welt überhaupt ‚wahr gemacht‘ werden könnten. Um es in mit einem einschlägigen Begriff Putnams auszudrücken: Millikan müsste, so scheint es, eine Art von ‚*Metaphysischem Realismus*‘ vertreten²⁶⁴.

Bekanntermaßen wirft eine solche metaphysische Konzeption von ‚Realität‘ viele – etwa ‚sinnkritische‘ – Fragen auf: So scheinen wir doch, um diese Konzeption ihrerseits als wahr oder falsch beurteilen bzw. verstehen zu können, aufgefordert, gleichsam von ‚*innerhalb*‘ unseres begrifflich-propositional strukturierten Weltverständnisses über beide Seiten des ‚Zaunes‘ zu blicken, um dann festzustellen, dass es auf der ‚außerhalb‘ liegenden Seite – erstaunlicherweise(?) – ungefähr so aussieht, wie wir uns das vorgestellt haben. Wir müssten demnach etwa sagen, dass die weltliche Seite der Entsprechungsrelation weitgehend so beschaffen ist wie das, was wir *im Rahmen unserer Praxis des Identifizierens und Klassifizierens* behaupten oder

²⁶⁴ Zu Putnams Verständnis von ‚Metaphysischem Realismus‘, s. etwa ders.: *Meaning and the Moral Sciences* (1978) und *Vernunft, Wahrheit und Geschichte* (1981/1982). Zusammenfassend lässt sich Putnams Begriff des ‚Metaphysischen Realismus‘ als eine Position verstehen, nach der selbst eine nach allen denkbaren Vernunftkriterien ‚ideale Theorie‘ von der Welt *falsch* sein könnte. Eine solche Konzeption hält er vor allem deshalb für ‚metaphysisch‘, weil nicht (empirisch) überprüfbar und deshalb sinnlos. Allerdings wird eine solche Argumentationsweise dann problematisch, wenn man – wie Putnam – meint, sie durch die Gegenkonzeption eines ‚internen Realismus‘ ergänzen zu müssen. Auf welcher ‚nicht-metaphysischen‘ Grundlage aber sollte sich wiederum die Richtigkeit einer solchen Konzeption beurteilen lassen? Und was soll überhaupt ‚intern‘ bedeuten, wenn es kein ‚extern‘ geben kann?

glauben, allerdings mit dem Unterschied, auch *unabhängig* von dieser Praxis, gleichsam ‚von sich her‘ individuiert und klassifiziert zu sein. Dabei müssten wir die entsprechende ‚natürliche‘ (kausale) Korrespondenzbeziehung anhand einer *Beschreibung* konstatieren können, die *ihrerseits* als den natürlichen Verhältnissen korrespondierend verstanden werden muss (usw.).

Nun ist eine solche, etwas polemisierende Darstellung selbst problematisch – was heißt hier ‚innerhalb‘ und ‚außerhalb‘? – und vielleicht auch nicht ganz fair (Millikan würde sie, wie wir noch sehen werden, zurückweisen). Immerhin scheint aber klar zu sein, dass zunächst der Versuch anzutreten wäre, Eigenschaften (und Relationen) irgendwie als etwas von einer sprachlichen Praxis begrifflichen Klassifizierens und/oder Dinge irgendwie als etwas von einer sprachlichen Praxis des Identifizierens und Reidentifizierens Unabhängiges (Vorgängiges) zu analysieren, um mit einer Korrespondenztheorie überhaupt etwas anfangen zu können. Und zumindest in Bezug auf Eigenschaften und Relationen ist es genau dies, was Millikan als möglich unterstellt:

„Properties and relations (...) must be in the natural world in such a way that their selfsameness or *identity* is there too. The identity of properties and relations must be an *objective* identity, not identity or sameness that depends upon language. It must be an identity upon which the univocity or equivocity, meaningfulness or emptiness, of our actual language categories depends, not one that is created by these categories“ (LTOBC S. 111)

Damit wendet sich Millikan explizit gegen eine philosophische Tradition, deren moderne linguistische Variante sie als „*Meaning Rationalism*“ bezeichnet. Nach Millikan entspricht Bedeutung ja *objektiven* Eigentlichen Funktionen von Repräsentation (etwa eines Satzes). Diese Funktionen könnten uns durch eine *empirische Theorie* erklärt werden, welche die kausalen Beziehungen zwischen Repräsentationen und der Welt durch historische Fakten und unter Voraussetzung selektiver Mechanismen identifiziert. Was nun Positionen des ‚Bedeutungsrationalismus‘, wie etwa platonistische, verifikationistische, inferentialistische, kohärentistische, formalpragmatische bzw. *geltungsorientierte* Theorien ganz allgemein tatsächlich verbindet, ist, dass sie gerade bestreiten, Wahrheit und Bedeutung seien aus einer *Beobachterperspektive* beschreibbare *Eigenschaften der natürlichen Welt*. Jene Beziehungen, die Bedeutungen und Gehalte bestimmen, werden dort ja als nur aus einer subjektbezogenen- bzw. *Teilnehmerperspektive* zugänglich gedeutet und in diesem Sinne gleichsam ‚*diesseits des Geistigen*‘ angesiedelt. Wichtig

ist dabei, dass es im Rahmen der Tradition des ‚Bedeutungsrationismus‘ (der sich ja auch die hier vertretene Position im weiteren Sinne zuordnet) bekanntermaßen durchaus naheliegt zu sagen, dass die ‚Welt‘ – oder vorsichtiger gesagt, unser *Weltzugang* –, *jedenfalls im Hinblick auf die durch ihre Eigenschaften und Relationen individuierten Dinge, in gewissem Sinne* durch (,unsere‘) Bedeutungsverständnisse erst ‚gemacht‘ wird.

Was damit meiner Ansicht nach sinnvollerweise gemeint sein kann lässt sich – auch unabhängig von spezifischen Rekonstruktionsbemühungen hinsichtlich der Bedingungen von Begrifflichkeit und propositionalem Gehalt (etwa im Stile Brandoms) oder gar einer hermeneutisch-idealistischen Verabsolutierung der *welterschließenden* Rolle der Sprache – anhand einfacher Beispiele verdeutlichen. So liegt es doch in vielen Fällen auf der Hand, dass was die Tatsachen sind *auch* von unseren *faktischen Klassifizierungen* bzw. *Individuierungen* abhängt. Wenn wir etwa zwischen Bergen und Hügeln, Kindern und Jugendlichen, zwischen Bäumen und Sträuchern oder zwischen verschiedenen Farben unterscheiden, sollte klar sein, dass die Extensionen der jeweiligen Begriffe nicht sozusagen ‚zwingend‘ sind. Wo genau die Grenze zwischen Bäumen und Sträuchern oder der Übergang von Grün zu Blau verläuft ist offenbar eine relativ willkürliche Angelegenheit. Es macht hier keinen Sinn, sich vorstellen zu wollen, es gäbe intrinsische, von unseren (mehr oder minder genauen) semantischen *Standards* bzw. von unseren normativ orientierten *faktischen Klassifikationsurteilen unabhängige*, Tatsachen – etwa hinsichtlich der Anzahl der Bäume, der Berge oder der blauen Dinge in der Welt. *Was* hier die Tatsachen sind bzw. überhaupt *mögliche* Tatsachen sein können, hängt doch offensichtlich auch davon ab, welche Verwendungsweisen entsprechender Ausdrücke als *zulässig* angesehen werden.

Ähnliches gilt offenbar auch für unsere Praxis der Bestimmung der *Identität* von Gegenständen durch die Zeit. Macht es Sinn sich vorstellen zu wollen, es wäre irgendwie intrinsisch festgelegt, unter welchem Maß an Veränderung in der Zeit ein Gegenstand noch *derselbe* ist und wann nicht mehr? Wie viele Teile an einem Auto können ausgewechselt werden bevor es nicht mehr dasselbe Auto ist? Es dürfte jedenfalls klar sein, dass es bei der Frage, was die Tatsachen sind in vielen Fällen sehr darauf ankommt *was* man als die *relevanten Kriterien* der Gegenstandsklassifizierung und *was* als die *relevanten Kriterien* der Gegenstandsidentität versteht. Dem stimmt auch Millikan zu.

Etwas anders als in den bisher genannten Beispielen scheint sich die Sache allerdings bei sogenannten ‚Termen für natürliche Arten‘ darzustellen, also jener sprachlichen Klasse, die schon in unserer Diskussion von Theorien ‚direkter Referenz‘ eine wichtige Rolle spielte und deren weltliche Relata Millikan als „Substanzen“

bezeichnet. In solchen Fällen sehen wir eine den möglichen unterschiedlichen Erscheinungsformen einer Substanz (eines Stoffes) *zugrundeliegende Mikrostruktur* (wie etwa bei Wasser H₂O oder bei Tigern eine bestimmte genetische Ausstattung) durchaus als letztgültiges Kriterium der Klassifizierung an. Und diese erscheint ja, anders als in den oben genannten Fällen, als etwas von der *Natur selbst* (in Form von Naturgesetzen) Vorgegebenes. Könnte man nicht zumindest in Bezug auf natürliche Arten doch von Substanzen mit intrinsischen Eigenschaften sprechen, also auch von sprach- oder Tatsachen der Welt, die als von jeglicher Form der ‚Repräsentation‘ völlig unabhängig gedacht werden können?

Ein moderner ‚Bedeutungsrationalist‘ wird dies mit guten Gründen bestreiten. Von einem ‚internen‘ epistemischen Sandpunkt aus betrachtet ist das Auffinden der kriterialen mikrostrukturellen Eigenschaften natürlicher Arten ja Gegenstand fortlaufender Forschungsbemühungen – und wird es wohl auch bleiben. Die Idee dessen, was Bernard Williams eine „*Absolute Conception of the World*“ genannt hat²⁶⁵, also einer Beschreibung der Welt, die von unserer Perspektive, von unseren kontingenten organischen und kulturellen Eigenheiten unabhängig ist, muss demnach eher als ein *forschungsleitendes Ideal* verstanden werden, also als eine Beschreibungsform, der wir uns annähern sollten, wenn wir Objektivität im Sinne eines höchstmöglichen Grades an Allgemeinheit anstreben. Auf jeden Fall aber kann auch eine solche Konzeption nach Auffassung des ‚Bedeutungsrationalisten‘ keiner *explanativen* Korrespondenztheorie der Wahrheit auf die Beine helfen, denn dies würde darauf hinauslaufen zu behaupten, dass unsere empirischen Theorien *deswegen wahr sind, weil* bestimmte Dinge sich *tatsächlich* (in Wahrheit) so verhalten, wie einige unserer empirischen Theorien behaupten – nämlich diejenigen, *auf deren Grundlage eine korrespondenztheoretische Erklärung der Wahrheit unserer Theorien gegeben wird*. Wir hätten es also mit einer Art von Zirkularität zu tun, wie sie sich immer dann einstellt, wenn die Unterscheidung zwischen Genese und Geltung unterschlagen wird. Das heißt, selbst wenn wir annehmen, wir könnten an einem Punkt ankommen, an dem unsere Theorien über die natürliche Welt allen empirischen Falsifikationsversuchen oder anderen Arten von Einwänden dauerhaft standhielten, müsste jeder Versuch einer *empirisch gehaltvollen korrespondenztheoretischen Erklärung* dieses Faktums (im Sinne Millikans oder auf andere Weise) seinerseits aufgrund von Kriterien beurteilt werden, die nicht bereits die Wahrheit der empirischen Annahmen voraussetzten. Doch auch in explanativer Hinsicht scheinen mir korrespondenztheoretische Erklärungsversuche der Möglichkeit der Wahrheit unserer Überzeugungen oder Sätze als nicht hilfreich: Stellen wir uns vor, wir hätten eine nach allen als maßgeblich

²⁶⁵ B. Williams: *Ethics and the Limits of Philosophy* (1985).

erachteten Kriterien *als wahr zu beurteilende* Menge an naturwissenschaftlichen Theorien, weshalb wir schließlich zu der Ansicht gelangten, über eine gleichsam ‚endgültige‘ Beschreibungen der natürlichen Welt zu verfügen, und nun würde jemand die Behauptung aufstellen, dass unsere Theorien *deswegen* wahr sind, *weil sie gemäß einer bestimmten Theorie* die intrinsischen Tatsachen der Welt abbilden. Es scheint mir nicht so, als könnten wir einer solchen Behauptung irgendeinen explanativ erhellenden Sinn abgewinnen, geschweige denn sie als ihrerseits *empirisch überprüfbar* verstehen. Denn natürlich müsste eine Korrespondenztheorie, die sich als Teil einer naturalistischen Theorie der Bedeutung und des Geistes versteht *ihrerseits empirisch überprüfbar* und dementsprechend als wahr beurteilbar sein²⁶⁶.

Ich halte die vorstehenden Argumente auch in Bezug auf Millikans teleofunktionale Version einer Korrespondenztheorie für stichhaltig, wenngleich ihr ‚metaphysischer Realismus‘ in einem eher abstrakten, strukturellen Verständnis der ‚objektiven Ontologie‘ der Welt besteht:

„The assumption that there must be one ideal or final articulation of a world affair, and ultimately of the world as a whole, that gets things ontologically *right*, that shows what the affair or the world is really composed of, has driven great philosophical engines in its time. But it may be that ontology can ultimately be made simpler by dropping this assumption. Interesting transformations and corresponding invariances, structure rather than kinds of *things*, would be left as the basic subject matter of ontology, things and their properties being derived from structure rather than vice versa“ (LTOBC, S. 109)

Millikan möchte die intrinsischen (‚selbstgleichen‘) Relata von ‚intentionalen Ikonen‘ oder von Repräsentationen also nicht im Sinne üblicher Gegenstandindividuiierung und -Klassifizierung verstanden wissen. Die systematische 1–1-Entsprechung soll an der Basis eine zwischen den möglichen ‚Transformationen von Repräsentationen‘ (bei Sätzen: möglichen Substitutionen ihrer Elemente) und den möglichen

²⁶⁶ Selbst wenn wir annehmen, wir könnten eine Korrespondenzbeziehung auf der Grundlage einer ‚absoluten Konzeption der Welt‘ im Sinne Millikans verständlich machen, bleibt zunächst auch noch in einer anderen Hinsicht unklar, ob dies für Millikans explanative Zwecke überhaupt tauglich wäre. Denn die Eigenschaften, Relationen, Substanzen oder Dinge, die in einer solchen Konzeption eine Rolle spielten, erscheinen doch als im Hinblick auf Millikans evolutionistischen Erklärungsansatz eher kontraproduktiv. Von solchen Arten Tatsachen hat ja wohl kaum ein selektiven Mechanismen ausgesetzter Organismus bisher gewusst. Zudem dürfte *diese Art* von Tatsachen beim ‚sich Herumschlagen mit der Welt‘ bei unseren stammesgeschichtlichen Vorfahren kaum eine unmittelbare Rolle gespielt haben. *Relevante* Tatsachen müssten hier doch eher den Charakter von solchen haben, wie dass sich links ein Fressfeind nähert oder dass das Alpha-Weibchen paarungsbereit ist; Tatsachen die in einer *perspektivenunabhängigen* Beschreibung der Welt kaum vorkommen dürften. So ist ja etwa ‚Feind‘ ein durchaus *wertender* Begriff, der nur in Bezug auf die *Perspektive* desjenigen Organismus, für den er eine Gefahr darstellt, Sinn macht.

‚Transformationen ihrer Relata‘ (möglichen Veränderungen der Verhältnisse in der Welt) sein. Nun scheint dies zwar für Millikans Beispielfall des Bienenanzes eine durchaus plausible Annahme, da hier eine bestimmte, quantitativ bestimmbare Variation des Bienenanzes einer bestimmten, quantitativ bestimmbaren Veränderung der (räumlichen) Beziehung zwischen der Sonne, dem Bienenstock und dem Nektar entspricht. Lässt sich Vergleichbares jedoch auch für möglich Substitutionen von Satzelementen behaupten? Wie sollen wir hier eine strukturelle Entsprechung zu Veränderungen der Verhältnisse in der Welt verstehen? Ich glaube nicht, dass Millikans Analogisierung dieser Fälle besonders einleuchtend ist, denn anders als bei Bienenanzes haben wir in propositionalen Sprachen ja keine von der jeweiligen Bedeutung *unabhängig* beschreibbaren, *quantitativen* Veränderungen, an denen wir eine Entsprechungsbeziehung im genannten Sinne festmachen könnten. Wir rufen nicht etwa *lauter* „Rot!“, wenn wir es mit einem besonders intensiven Farbton zu tun haben. Leider bleibt uns Millikan eine plausibilisierende Erläuterung für das, was im Hinblick auf propositionale Repräsentationen mit einer strukturellen Entsprechungsbeziehung zwischen ‚Transformationsmöglichkeiten‘ gemeint sein kann, schuldig.

Dennoch sollten wir uns noch etwas genauer ansehen, wie Millikan ihre als strukturell verstandene ‚objektive Ontologie‘ entwickelt. Die ‚Basis‘ sollen Substanzen und Eigenschaften bilden, die in Beziehung zueinander bestimmt werden. Dabei soll eine Substanz eine ‚selbstgleiche‘ relativ zu einer Menge von Eigenschaftsbereichen sein, bezüglich deren sie jeweils, *gemäß natürlicher Notwendigkeit*, eine Eigenschaft unter Ausschluss aller anderen Eigenschaften aus diesem Bereich hat. Eine solche Eigenschaft wäre für die Substanz *Gold* etwa die atomare Struktur, nicht aber ein Volumen. Korrelativ dazu wäre eine Eigenschaft eine ‚selbstgleiche‘ als die, die sie relativ zu einem bestimmten Bereich von gegensätzlichen Eigenschaften und zu einer Menge von Substanzkategorien ist, deren Mitglieder immer, *gemäß natürlicher Notwendigkeit*, eine Eigenschaft unter Ausschluss aller anderen Eigenschaften dieses Bereichs haben. Die atomare Struktur von Gold ist eine ‚selbstgleiche‘ Eigenschaft, die nur dieser Substanz und keiner anderen Substanz, die naturnotwendig (auch) über ihre atomare Struktur individuiert ist, zukommt.

Auf der *Basis* dieser strukturellen Bestimmung einer ‚objektiven Ontologie‘ der Welt will Millikan dann die epistemisch bedingten, von unseren repräsentationalen Vorrichtungen ‚gemachte‘ Ontologie von dauerhaften Objekten als eine der Vereinheitlichung von räumlichen und zeitlichen Teilen zu einem Ganzen erklären, die für bestimmte Zwecke nützlich ist. Es würde, wie gesagt, eine eigene Abhandlung erfordern, jeweils auf Einzelheiten ihrer differenzierten Überlegungen einzugehen. Auch so dürfte jedoch klar geworden sein, dass die ‚Korrespondenz‘ zwischen

Repräsentationen und der Welt im Sinne Millikans bestenfalls als eine *Strukturisomorphie* verstanden werden könnte. Die ‚objektive Welt‘ soll ja primär nicht aus Dingen (mit bestimmten Eigenschaften), sondern aus Eigenschaften (und ihren Relationen) bestehen, die über Substanzen und wechselseitige Exklusionsprinzipien individuiert sind. Dem sollen die über ihre jeweiligen Unvereinbarkeiten des relevanten Eigenschaftsbereichs spezifizierten Begriffe korrespondieren. Doch für so etwas, wie Korrespondenz im Sinne einer Strukturisomorphie bräuchte man auf der ‚repräsentationalen‘ Seite etwas, das *in einem quantifizierbaren Sinne strukturiert* ist. Doch dies sind propositionale bzw. begriffliche Formen von ‚Repräsentationen‘ ja gerade nicht.

Nun soll darüber hinaus auch das Unvereinbarkeitsprinzip selbst, also der Satz vom auszuschließenden Widerspruch, natürliche Prinzipien der Unvereinbarkeit ‚parallelisieren‘:

„In the outer world substances and properties display an abstract structure showing systematic exclusion of property and contrary property upon common ground. This parallels inner consistency of belief. Coherence can be a test of correspondence because coherence in thought corresponds to an abstract feature of the world. The law of noncontradiction reflects the (or a) structure of being. It is a template of the general structure of world affairs as it should reflect in thought – a template at least of that kind of world affair that language having subject-predicate structure is designed to map.“ (*LTOBC*, S. 300 f.)

Der hier naheliegende Vorwurf des Psychologismus scheint Millikan nicht zu beunruhigen. Im Gegenteil vertritt sie hinsichtlich der Bedingungen der Möglichkeit von wahren (empirischer) Theorien offensiv eine Position radikaler naturalistischer Selbstbezüglichkeit: Die epistemisch kriterialen Regeln selbst, sogar die apriorischen Regeln der Logik, wie insbesondere das Prinzip vom auszuschließenden Widerspruch, seien ihrerseits nichts anderes als durch eine evolutionsgeschichtliche Lernphase in uns installierte kognitive Mechanismen, die strukturelle ‚Prinzipien‘ der natürlichen Welt widerspiegeln:

„(...) the law of noncontradiction as applied to representations reflects a thoroughly natural structural principle rather than an a priori metaphysical or rational principle, or the workings of our language (or say the workings of a Kantian understanding). (...) *If* the law of noncontradiction is grasped ‚a priori‘, this must be so only in the sense that nature, via evolution, has built this grasp into us as a mirror of reflection (possibly only a sufficing reflection) of a structural principle in the natural world with which we must deal in order to survive.“ (*LTOBC*, S. 257 f.)

Dies ist einerseits schon fast beunruhigend konsequent, andererseits aber auch starker Tobak. Denn was sollen wir darunter verstehen, dass der Satz vom auszuschließenden (,zu vermeidenden‘) Widerspruch ein ‚natürliches strukturelles Prinzip‘ reflektiert? Ist dieses Prinzip ein Naturgesetz? – Offenbar nicht, doch was ist es dann? Es geht hier nicht nur um eine in Widersprüche führende ‚psychologistische‘ Unterschlagung der Unterscheidung zwischen Genese und Geltung. Der Punkt bezüglich des fundamentalen Prinzips vom auszuschließenden Widerspruch ist, dass es gar nicht möglich ist, dass wir in irgendeiner Weise feststellen könnten, er gälte gar nicht. Dieses Prinzip bildet nämlich die konstitutive Grundlage der Möglichkeit von Sinn überhaupt²⁶⁷. Für die Natur oder Naturvorgänge ist es dagegen kennzeichnend, dass es sich auch anders verhalten *könnte*, das heißt, es wäre prinzipiell möglich, dass nicht nur die Anfangsbedingungen, sondern auch die Naturgesetze andere hätten sein können. Es ist zwar auch eine gute Frage, in welchem Ausmaß wir uns tatsächlich die natürlichen Gegebenheiten als variierbar vorstellen können, ohne gleichsam den Boden der Vergleichbarkeit der Verhältnisse unter den Füßen zu verlieren. In gar keinem Sinne aber kann die natürliche Welt als nicht-strukturiert bzw. nicht-unterschieden gedacht werden. Und da eine Eigenschaft etwas ist, dessen Gegenteil (im Sinne der Negation) möglich sein muss, kann die Strukturiertheit oder Unterschiedenheit keine noch so grundlegende *natürliche Eigenschaft* der Welt sein, die wir – ob angeboren oder nicht – gleichsam durch die Logik ‚parallelisieren‘. Wie immer der Satz vom zu vermeidenden Widerspruch ‚in unseren Kopf (ge-)kommen‘ (sein) mag, wir können ihn jedenfalls nicht als eine quasi evolutionär erworbene *empirische Wahrheit über die Welt* verstehen.

Ich hatte oben bereits deutlich gemacht, dass durch die hier vorgebrachten Einwände evolutionistische Theorien der Entwicklung geistiger und sprachlicher *Fähigkeiten* nicht grundsätzlich zurückgewiesen werden sollen. Dies gilt auch für einen wichtigen Gedanken, der Millikans korrespondenztheoretischer These *zugrundeliegt* und der sich etwa dahingehend zusammenfassen lässt, dass sich *nur* unter der Voraussetzung irgendeiner Form von Rückkopplung zwischen geistigen oder sprachlichen Akten oder Fähigkeiten (bzw. ihrer Vorstufen) und der diesen Akten oder Fähigkeiten *vorgegebenen* natürlichen Welt eine neo-darwinistische Erklärung ihrer Existenz und Entwicklung überhaupt *möglich* ist, und dass diese Form der Erklärung dafür einfach die plausibelste und am besten belegte ist, die auf dem ‚Markt‘ ist²⁶⁸. Diese Idee lässt sich in unbedenklicher Weise auch so formulieren: Wir denken und sprechen überhaupt nur deswegen, *weil* unsere – kultur- und naturgeschichtlichen –

²⁶⁷ S. hierzu E. Tugendhat/ U. Wolf: *Logisch-semantische Propädeutik* (1983), Kap. 4.

²⁶⁸ Für eine frühe Ausarbeitung dieses Gedankens, s. Konrad Lorenz: *Die Rückseite des Spiegels – Versuch einer Naturgeschichte menschlichen Erkennens* (1973).

Vorgänger von den jeweiligen kognitiven Vorstufen dieser Fähigkeiten irgendwelche überlebens- und reproduktionsrelevanten Vorteile hatten. Bestimmte kognitive Fähigkeiten oder Leistungen mögen dabei ursprünglich gleichsam ‚zufällig‘ entstanden sein. Jedenfalls können wir und konnten unsere Vorgänger damit nur deshalb erfolgreich sein – und so bestimmte Weisen des Denkens (oder Sprechens) biologisch oder über Lernprozesse reproduzieren – weil wir und sie zumindest die für eine kritische Masse an praktischen Erfolgen relevanten Aspekte der Welt erfassen oder erfassten. Der selektive Anpassungsdruck im Hinblick auf ein *angemessenes* Erfassen jeweils relevanter Bereiche unserer Umwelt muss also auch auf die Art und Weise des Aufbaus unseres Gehirns (insbesondere des ‚Erkenntnis-‘, und ‚Sprachermöglichungsapparates‘) einen prägenden Einfluss gehabt haben²⁶⁹. Das heißt, eine jede Form von Theorie über eine Anpassungsgeschichte unserer kognitiven Vermögen muss unterstellen, dass es – in irgendeiner Form – so etwas, wie eine ‚*Ansich-Welt*‘ im Sinne *intrinsischer natürlicher Strukturen* gibt, an die angepasst wurde. Denn nicht nur die ganze Rede von einer phylogenetischen Anpassung der Erkenntnisfähigkeit, sondern auch die Vorraussetzung der *Möglichkeit* individuellen oder gesellschaftlichen *Lernens* macht offenbar nur dann Sinn, wenn *etwas* Gleichförmiges als vorgegeben angenommen wird, *das* durch Lernen besser oder *angemessener* erkannt, erfasst oder beschrieben werden kann²⁷⁰. Was durch die oben vorgebrachte Kritik bestritten werden sollte war lediglich, dass dies explanativ, im

²⁶⁹ Ein gängiger genereller Einwand gegen evolutionistische Erklärungen unseres Erkenntnisvermögens lautet, dass solche Erklärungen schon deshalb nicht stimmen könnten, weil auch Spezies mit sehr geringen kognitiven Vermögen evolutionär sehr erfolgreich waren. Ich halte diesen Einwand allerdings für nicht stichhaltig. Denn zu behaupten, dass die gesteigerte Intelligenz des *homo sapiens* ein Überlebensvorteil war und ist, schließt ja nicht aus, dass andere Fähigkeiten (oder ihre Kombination) nicht *auch* gute ‚Überlebensstrategien‘ bilden können: Ein besseres Erkennen der Welt mag eine mögliche Überlebensstrategie darstellen, weit springen zu können oder sich massenhaft fortzupflanzen andere. Dass dies keine Beliebigkeit erzeugt, lässt sich an vielen Beispielen von Arten zeigen, die unter geänderten Umweltbedingungen ihren Vorteil, bzw. ihre ‚biologische Nische‘ verloren und ausstarben. Zudem gilt es zu berücksichtigen, dass das Argument von der Anpassung an intrinsische Strukturen der Welt ja in ähnlicher Form für nicht-kognitive Vermögen gilt (beispielsweise ermöglichen Flügel ja nur in Anpassung an bestimmte Bedingungen das Fliegen).

²⁷⁰ Bezüglich der vieldiskutierten Frage nach der Möglichkeit wissenschaftlichen Fortschritts hat John Worrall (in ders.: ‚Structural Realism: The Best of both Worlds?‘ (1989)), im Anschluss an Poincaré, ebenfalls ein strukturelles Verständnis von Realismus vorgeschlagen. Was nachfolgende Theorien von ihren prognostisch durchaus ebenfalls erfolgreichen Vorgängern erbten (und ausbauten), seien Wahrheiten im Hinblick auf die mathematisch beschriebene *Struktur* der Phänomene, selbst wenn deren inhaltliche Aussagen über ihre *Natur* dabei jeweils vollkommen widerlegt wurden. Auch wenn dies manche Frage offenlässt – etwa nach dem Verhältnis ‚inhaltlicher‘ und ‚formaler‘ Wahrheiten, halte ich es für einen sehr interessanten Vorschlag, der zudem gut zu der hier vertretenen Sicht der Dinge zu passen scheint.

Sinne einer korrespondenztheoretischen Erklärung von Wahrheit und ‚Repräsentationalität‘, bzw. von propositionalem Gehalt und sprachlicher Bedeutung, nutzbar zu machen ist.

Natürlich ist eine jede Redeweise von einer ‚Welt an sich‘, wie schon gegen Kant immer wieder eingewendet wurde, aus verschiedenerlei, zum Teil ja bereits vorgetragenen, Gründen problematisch, weshalb ich mit Millikan auch lieber von intrinsischen ‚Strukturen‘ spreche. Aber was auch immer darunter zu verstehen sein kann, es gibt jedenfalls nichts her, was einer nicht-trivialen Korrespondenztheorie auf die Beine helfen könnte, also eine verständliche Konzeption einer vorgegebenen Welt in dem Sinne, dass sie *qua* einer Entsprechungsbeziehung propositionale Gehalte bzw. die Bedingungen der Wahr- oder Falschheit von ‚Repräsentationen‘ *erklären* könnte. Die Bedingungen, unter denen wir *über* die Welt reden oder denken können, lassen sich gewissermaßen nicht von der Welt her verstehen. Ihr Explizit-Machen lässt sich, wie ich behauptet habe, nur in einer Begrifflichkeit vollziehen, die am Ende auf verantwortbare Subjekte, also kompetente Teilnehmer einer Praxis Bezug nimmt, die zu Fragen im Sinne des Richtig oder Falsch (begründet) *Stellung beziehen können müssen*. Dabei müssen wir jeweils die Gründe oder Kriterien, die Bedingungen richtigen oder vernünftigen Urteilens kennen, um vernünftigerweise als wahr beurteilbare Überzeugungen haben oder zuschreiben, bzw. entsprechende Sätze äußern oder verstehen zu können. Und es ist die *kommunikative Struktur* unsere propositional-sprachlichen Praktiken und die jeweils als gültig (richtig, vernünftig) erachteten Klassifikations-, Individuations- und Identifikationskriterien, die konstitutiv für das sind, was überhaupt Dinge, Substanzen, Eigenschaften, Relationen – und damit die Tatsachen – sein können. Dazu werde ich unten noch einiges sagen. Jedenfalls gibt es unabhängig von solchen Praktiken gar keine Antwort auf die Frage, was Dinge, Eigenschaften oder Substanzen sind und welche es geben kann. Auch deshalb kann es keine Erklärung des Zusammenhangs von Wahrheit, Bedeutung und propositionalem Gehalt geben, der sich auf eine Form von Korrespondenzbeziehung zu einer vorgegebenen Welt von Dingen, Eigenschaften oder Substanzen beruft.

Ich denke, dass eines der grundsätzlichen Missverständnisse nicht nur korrespondenztheoretischer, sondern auch anderer klassischer Wahrheitskonzeptionen darin besteht, einen Satz, wie: „Das ist wahr“ als eine Art von Zuschreibung einer Eigenschaft zu deuten. Ich hatte ja oben bereits dafür plädiert, solche Sätze *primär* als Form der *Zustimmung*, der *expliziten Stellungnahme* zu einer durch einen indikativischen Satz (durch eine assertorische Äußerung) ausgedrückten Proposition (zu einem ‚Gedanken‘), bzw. zu einer Menge durch indikativische Sätze ausgedrückten Propositionen (etwa einer Theorie) zu verstehen. Im letzteren Fall hat der

Wahrheitsbegriff dabei auch die Funktion eine gewissermaßen ‚zusammenfassende‘ Behauptung eines ganzen Zusammenhangs von Behauptungen zu ermöglichen. Weitere sekundäre Funktionen ließen sich aufzählen. Keine dieser Funktionen entspricht jedoch einer *Beschreibung* eines überprüfungs- oder begründungstranszendenten *Zustandes* oder *Ereignisses*. Dass alle noch so gut begründeten Überzeugungen falsch sein können²⁷¹ heißt vor allem, dass es immer sein kann, dass wir bessere Gründe finden oder bisherige Begründungen in einem neuen Licht beurteilen (bzw. – im Extremfall – dass es solche besseren Gründe geben kann, selbst wenn sie uns nicht zugänglich sind²⁷²). Was vernünftigerweise geglaubt oder behauptet werden sollte, lässt sich natürlich nicht durch die ‚richtige Analyse‘ des Wahrheitsbegriffs herausbekommen, sondern wir halten umgekehrt vernünftigerweise das für wahr – und sagen dabei aus einer Teilnehmerperspektive: „Das *ist* wahr“ – für was wir gute Gründe zu haben glauben. Der *grundlegende* stellungnehmende Begriff scheint mir dabei der des Vernünftig-Seins. Wichtig ist jedenfalls, dass sich unser Verständnis des Sinns von Stellung nehmenden Begriffen nicht vollständig aus einer Beobachterperspektive einholen lässt, die diese im Sinne von Einstellungen relativiert, deren konstitutive Bedingungen wir *beschreiben* können. Und es entspricht einem *Kategorienverstoß*, den Begriff der Wahrheit, dessen primärer Sinn der einer (positiven) *Stellungnahme* zur *Behauptbarkeit propositionaler Gehalte* (etwa empirischer Sätze) ist, als eine natürliche Eigenschaft von bestimmten („repräsentationalen“) Arten von Zuständen oder Ereignissen – im Sinne von *Geschehnissen* – verstehen zu wollen, die ihrerseits *durch eine empirische Theorie erklärbar* wäre.

Über diese allgemeinen Einwände gegen nicht-triviale, also ‚metaphysisch-realistische‘ Korrespondenztheorien der Wahrheit hinaus lässt sich gegen Millikans spezifisch evolutionistische Variante zudem noch ein weiterer Einwand vorbringen, der es verdient, hier noch zur Sprache zu kommen. Er betrifft die tendenzielle Gleichsetzung des Faktums der Reproduktion einer ‚Repräsentation‘ (eines Überzeugungs- oder Satztyps) mit ihrer Wahrheit. Millikans Argument im Sinne der ‚besten Erklärung‘

²⁷¹ Was Rorty die „warnende“ Funktion des Wahrheitsbegriffs nennt (s. R. Rorty: ‚Pragmatismus, Davidson und der Wahrheitsbegriff‘ (1986/1990)).

²⁷² Wahrheit lässt sich deshalb in dem Sinne als rechtfertigungstranszendent verstehen, als es möglich ist sich vorzustellen, dass wir nie in der Lage wären, die Wahrheit einer bestimmten empirischen Behauptung zu überprüfen. Dabei könnten wir den Sinn einer solchen Aussage aber nur dadurch verstehen, dass wir Bestätigungsbedingungen für ihre *Komponenten* aus Zusammenhängen mit anderen möglichen Behauptungen kennen, die nicht ihrerseits als rechtfertigungstranszendierend verstanden werden können. Diese Möglichkeit stellt demnach lediglich einen, zu den primären Fällen der Zustimmung zum Erfülltsein der Behauptbarkeitsbedingungen und zur Unterscheidung des Die-Wahrheit-Sagens vom Lügen derivativen, Grenzfall dar.

unterstellt ja (durchaus im Einklang mit dem klassischen Pragmatismus jamescher Prägung), dass *wahre* Repräsentationen eher oder häufiger zu erfolgreichem Verhalten führten als unwahre und deshalb reproduziert würden. Dementsprechend wäre auch ein Sprachgebrauch, durch den sich häufig etwas Wahres zum Ausdruck bringen lässt einer Sprachgemeinschaft dienlicher, als ein solcher, mit dem dies weniger häufig möglich ist (was den ‚reproduktiven Erfolg‘, also die Tatsache fortgesetzter Verwendung, erklärte).

Dies klingt zunächst plausibel. Problematisch an einer solchen Angleichung von Wahrheit und durch Verhaltens- oder Handlungserfolg ermöglichte Reproduktion (von Repräsentationen) ist aber bereits, dass es *unterhalb einer bestimmten Ebene kognitiver Fähigkeiten und kommunikativer Praktiken* als zumindest fragwürdig erscheint, überhaupt *Wahrheitsfähigkeit* zu unterstellen, um dann die Reproduktion der ‚repräsentationalen Vorrichtungen‘ (und entsprechender Verhaltensmuster) durch die Wahrheit der verhaltensleitenden Repräsentationen erklären zu wollen. Ein Frosch, dessen Verhalten etwa wir *im Sinne* einer Maxime „Schnapp nach allen Dingen einer bestimmten Größe, die sich in deinem Gesichtsfeld bewegen“ interpretieren könnten, kann durchaus eine hohe Anzahl an Fehlschüssen oder Treffern auf Ungenießbares zustande bringen, was aber seinem *reproduktiven Erfolg* – und damit dem der verhaltensauslösenden ‚Repräsentationen‘ – keinen Abbruch tut. – Ganz im Gegenteil! Sollen wir hier wirklich sagen, der Frosch habe, zumindest im Erfolgsfalle, *wahre* Repräsentation, *weil* diese sich, *qua* der erfolgreichen ‚Gesamtstrategie‘ der Frösche, reproduzieren?²⁷³ Millikan selbst hatte doch eine ganze Menge an Bedingungen für Propositionalität aufgezählt, die aber von den kognitiven Fähigkeiten der Frösche sicher nicht erfüllt werden. Sie nennt aber keinen plausiblen Grund dafür, schon unterhalb von Kompetenzen, die propositionale ‚Repräsentationalität‘ ermöglichen, von Wahrheitsfähigkeit zu reden.

Außerdem scheinen sich auch *Überzeugungen* und *Sätze* ziemlich erfolgreich zu reproduzieren, die nach meinem Dafürhalten unwahr sind, wie etwa solche des Aberglaubens oder mit religiösem Inhalt. – Oder irre ich hier *schon deshalb*, weil ich die andauernde Reproduktion von entsprechenden Überzeugungen oder Sätzen nicht zum Anlass nehme sie für wahr zu halten? Mitglieder moderner Gesellschaften scheinen dagegen dazu zu neigen, zu Überzeugungen zu kommen, von denen man annehmen kann, dass sie ihre biologische Reproduktion eher hemmen, als fördern. Man denke nur an das ‚Wissen‘ um Möglichkeiten der Schwangerschaftsverhütung oder um mögliche Probleme, mit denen sich zukünftige Generationen konfrontiert sehen

²⁷³ Ein solcher Einwand wird bereits in Dennett (1978) vorgebracht. Ähnliches findet sich aber auch bei Fodor, Putnam und anderen.

dürften. Sind solche Überzeugungen deshalb eher falsch?²⁷⁴ – Sind die ‚kognitiven Vorrichtungen‘, die solche ‚Repräsentationen‘ hervorbringen, deshalb dysfunktional?

Hier haben wir es offenbar Schwierigkeiten von der Art zu tun, mit denen sich auch der klassische Pragmatismus mit seinem Motto „It’s truth that works“ konfrontiert sieht. Dies müssen wir hier nicht weiter ausführen. Wir werden uns statt dessen nachfolgend ausführlich mit einer Variante naturalistischer Philosophien der Bedeutung und des Geistes auseinandersetzen, die von den bisher gegen reduktiv-materialistische, bzw. gegen – sowohl die ontologischen Unterstellungen mentaler Zuschreibungen betreffend als auch im ‚metaphysischen‘ Sinne – ‚realistische‘ Positionen vorgebrachten Argumenten weitgehend unberührt geblieben ist, nämlich dem Interpretativen Materialismus.

²⁷⁴ Beim Vergleich zwischen modernen und eher traditionellen Gesellschaften kann darüber hinaus sogar der Verdacht aufkommen, dass sich die Gesamtmenge der von den Individuen einer Gesellschaft üblicherweise erkannten Wahrheiten umgekehrt-proportional zu deren biologischer Reproduktionsrate verhält.

III INTERPRETATIVER MATERIALISMUS

1 Quine, Davidson und Dennett

Positionen, die ich als „Interpretativer Materialismus“ bezeichne, lassen sich zunächst vereinfachend als eine holistische, ‚weiche‘ Form von Behaviorismus verstehen und bilden, nach dem Scheitern des klassischen Behaviorismus von Watson, Skinner, Carnap u.a., die wichtigste naturalistische Alternative zum ‚Realismus‘ reduktiver Kognitionstheorien. Während aber Quine wohl eher dem Eliminativismus zuzurechnen ist – aus systematischen Gründen aber hier behandelt wird – scheint mir die gemeinsame Einordnung der je für sich einflussreichen Positionen Dennetts und Davidsons unter diesem gemeinsamen Label, trotz einiger noch zu diskutierender Unterschiede, aufgrund wichtiger geteilter Annahmen als durchaus gerechtfertigt. So ist es insbesondere eine zentrale These beider Theoretiker, dass ‚intentionale Zustände‘ oder ‚Ereignisse‘ nicht durch ihre tokenidentischen *physischen* Eigenschaften *individuiert* sind, sondern solche Individuierungen nur im Rahmen eines mit Optimierungsannahmen (Rationalität und Wahrheit ‚maximierenden‘ bzw. optimierenden Prinzipien) operierenden *Zuordnungsschemas*, bzw. einer ‚interpretativen‘ *Theorie* auf der Grundlage von *Verhaltensbeobachtung* möglich sind. Die einzigen ‚interpretationsunabhängigen‘ Fakten einer ‚Intentionalen Psychologie‘ oder ‚Bedeutungstheorie‘ sollen in ihrer behavioralen Datenbasis bestehen. Die *innere physische* Natur psychischer Zustände oder Ereignisse soll dagegen für Theorien der Bedeutung und Geistes – zumindest auf der Ebene ‚intentionaler‘ Beschreibung – keine Rolle spielen und zwar nicht einfach deshalb, weil wir zu diesen aus kontingenten epistemischen Gründen keinen Zugang hätten, sondern weil es einfach keine eindeutigen, gesetzesartigen Zusammenhänge zwischen den physischen und den ‚intentionalen‘ Beschreibungen von Zuständen oder Ereignissen gäbe. Beide Theoretiker halten allerdings daran fest, dass propositionale- und Bedeutungszuschreibungen *Formen der Beschreibung physischer Zustände* oder *Ereignisse* sind.

Die vom Interpretativen Materialismus vertretene Relativierung des ‚Intentionalen‘ auf die epistemischen Bedingungen seiner Zuschreibung hat Auswirkungen auf die Form der Semantik sprachlicher Bedeutungen und propositionaler Einstellungen. Sowohl Dennett, als auch Davidson vertreten Varianten des *Holismus*: Überzeugungen oder Sätze haben danach nicht als einzelne, sondern nur im Rahmen eines

‚Interpretationsschemas‘ einen Gehalt oder eine Bedeutung, welches viele Überzeugungen und andere Einstellungen bzw. – bei Davidson – viele Bedeutungen sozusagen ‚auf einen Schlag‘ zuordnet und dabei die Beziehungen der Einstellungen oder Sätze untereinander berücksichtigt. Dabei sollen die behavioral verfügbaren Belege in Verbindung mit den Optimierungsannahmen den Interpreten *nicht* auf ein bestimmtes Interpretationsschema als einzig möglichem festlegen. Die Tragweite dieser sogenannten ‚Unbestimmtheitsthese‘, von der oben schon mehrfach die Rede war, ist allerdings auch zwischen Dennett und Davidson umstritten. Jedenfalls bildet sie einen wichtigen Pfeiler der anti-reduktionistischen Argumentation vor allem bei Dennett. Sie scheint aber auch mit einigen der hier vorgebrachten Einwände gegen die zuvor diskutierten ‚realistischen‘ Vorschläge zu naturalistischen Identitäts- oder Individuierungsbedingungen von Einstellungen und Gehalten in Einklang zu stehen.

Darüber hinaus hat die holistische Relativierung der genannten Individuationsbedingungen durch Inanspruchnahme von Optimierungsannahmen offenbar auch den Vorzug – ähnlich wie Millikans teleologischer Funktionsbegriff – einen *konzeptuellen Spielraum für die Möglichkeit von Irrtümern* (Fehlrepräsentationen), bzw. eine entsprechende Unterscheidbarkeit von ‚Soll-‘ und ‚Ist-Zuständen‘ zu ermöglichen. Dies scheint nämlich bereits dadurch gewährleistet, dass die Individuierung ‚intentionaler Zustände‘ oder ‚Ereignisse‘ grundsätzlich von einer eindeutigen Festlegung auf unmittelbare Ursachen entkoppelt wird. Insgesamt scheinen also manche der bisher gegen naturalistische Positionen vorgebrachten Einwände nicht für den Interpretativen Materialismus zu gelten.

Von besonderem Interesse im Zusammenhang mit dem Thema dieser Arbeit ist vor allem Davidsons Sprachphilosophie. Davidsons Position ist nämlich nicht nur äußerst einflussreich, sie stellt tatsächlich auch einen der wenigen ernstmachenden Vorschläge zur *Möglichkeit einer empirischen Operationalisierung wahrheitssemantischer Bedeutungstheorien* dar, die das *für sprachliches Verstehen hinreichende Wissen von Kommunikationsteilnehmern* angeben sollen. Zudem geht Davidson, im Gegensatz zu den meisten der bisher diskutierten Autoren, mit der hier vertretenen Ansicht konform, dass Sprachlichkeit eine Voraussetzung für die Zuschreibbarkeit von propositionalen Einstellungen (im vollwertigen Sinne) ist. Berücksichtigt man außerdem noch seine Inanspruchnahme von Rationalitätsprinzipien als apriorische Postulate einer adäquaten ‚Interpretationstheorie‘, wird vielleicht verständlich, warum so manch einer zu glauben scheint, Davidson als so etwas wie die ‚analytische Ausgabe‘ eines sprachphilosophisch aufgeklärten Hermeneutikers deuten zu können. Dies ist jedoch ein Irrtum. Schon Davidsons wahrheitssemantisch begründeter Interpretationsbegriff ist eigentlich ein ganz anderer als der üblicher hermeneutischer Verständnisse. Und auch

Davidson muss, wie wir noch genauer sehen werden, nach den hier vertretenen Kriterien insofern als Naturalist gelten, als er die Kompetenz zum Verstehen sprachlicher Äußerungen nicht als eine *Fähigkeit zur – ihrerseits der Kritik durch Sprecher ausgesetzten – geltungsbezogenen Stellungnahme* zu (normalerweise adressierten) sprachlichen Handlungen, sondern letztlich als eine Fähigkeit zur Erstellung einer spezifischen Form (wahrer) *empirischer Theorien*, nämlich solcher, die Aussagen über quasi *vorgefundene* ‚rationale Äußerungsereignisse‘ machen, analysieren will.

Dies hat selbstverständlich Konsequenzen für seine Auffassung davon, was es in philosophischer Hinsicht über Bedeutung zu sagen gibt. Unter weitgehender Vernachlässigung von eher technischen Fragen nach der grundsätzlichen Möglichkeit einer umfassenden extensional-wahrheitssemantischen Analyse aller sprachlichen Äußerungsformen, denen wir in irgendeinem Sinne eine Bedeutung zuschreiben können, werde ich zu zeigen versuchen, dass Davidsons Theorie der Bedeutung gerade dadurch, dass sie einem als vorverstanden unterstellten Wahrheitsbegriff die zentrale Rolle in der Theoriearchitektur zuweist, die eigentlich interessanten Phänomene und Fragen außen vor lässt. Meine Kritik wird darauf hinauslaufen, dass Davidsons ‚Theorieprogramm‘ an einer szientistisch motivierten Beschneidung der explanativen Ansprüche an sprachphilosophische Fragestellungen krankt, die aus unvoreingenommener Perspektive nicht überzeugen kann.

Da nun aber die maßgebliche Grundlage zum Verständnis nicht nur von Davidsons, sondern auch von Dennetts Position der holistische Behaviorismus Quines bildet, ist es sinnvoll, sich zunächst kritisch mit den für unsere Diskussion relevanten, äußerst einflussreichen Thesen Quines auseinanderzusetzen. Wie sich später zeigen wird, ist Dennett mit seiner Relativierung des ontologischen Status von ‚intentionalen Zuständen‘ in mancher Hinsicht näher an Quines eliminativen Ambitionen, als Davidson. Insbesondere deutet Dennett, im Unterschied zu Davidson, Quines berühmte, in *Word and Objekt* entwickelte These von der Unbestimmtheit der Referenz als eine Art ‚Interpretationsrelativismus‘ bezüglich des ‚Intentionalen‘ ganz allgemein.

Quine im Hintergrund

Quine ist ausgewiesenermaßen kein Freund von solchen ‚Dingen‘, wie Bedeutungen, Propositionen, inferentiellen Beziehungen, Eigennahmen und ähnlichem. Es der sogenannte ‚Bedeutungskeptizismus‘ Quines, vor dessen Hintergrund die ganze Art

und Weise, in der Dennett und Davidson das Thema ‚Intentionalität‘ angehen überhaupt erst verständlich wird. Dieser Skeptizismus beruht zunächst auf der allgemeinen naturalistischen Voraussetzung, dass es *keine klare methodologische Trennung* zwischen der Behandlung philosophischer und solcher Fragen gibt, die Gegenstand empirischer Forschung sein können. Er beruht des weiteren insbesondere auf der szientistisch-ontologischen Voraussetzung, dass Bedeutungen, Propositionen u. ä. im Sinne *ontologischer* Setzungen beurteilt werden müssten – die sich aber als vorwissenschaftliche Fiktionen erwiesen, da es für sie *keine präzisen*, empirisch operationalisierbaren *Individuationskriterien* gibt, weshalb sie auch nicht Teil einer *wissenschaftlich begründeten Ontologie* sein sollten. Und er beruht schließlich auf der behavioristischen Voraussetzung, dass die *einzigsten objektiven Fakten* im Hinblick auf unsere Redeweise über Bedeutung in den jeweiligen *Dispositionen* bestehe, aufgrund von bestimmten äußeren Stimuli Geräusche zu äußern. Der residuale, wissenschaftlich brauchbare ‚cash value‘ unseres ‚konfusen‘, vorwissenschaftlichen Bedeutungsbegriffs besteht für Quine also in demjenigen, was sich an Generalisierungen über verbales Verhalten aus der Perspektive eines behavioristisch²⁷⁵ orientierten Beobachters erfassen lässt.

Der Quine üblicherweise zugeschriebene semantische Holismus, nach dem die Bedeutung der Ausdrücke einer Sprache mehr oder minder stark von der Bedeutung anderer – im Extremfall *aller* anderen – Ausdrücke dieser Sprache abhängt, ist zunächst ein *epistemisch* begründeter Holismus. In ‚*Two Dogmas of Empiricism*‘ (*TDE*), als dem *locus classicus* des quineschen Holismus, entwickelt Quine seine Überlegungen tatsächlich in Bezug auf Aussagen (*statements*), und es spricht manches in seinen Formulierungen dafür, dass er dort auch einen holistischen *Verifikationismus* vertritt, nach dem die ‚Bedeutung‘ von (assertorischen) Sätzen von ihren Bestätigungsbeziehungen abhängen und diese in einem komplexen Netzwerk mehr oder weniger eng miteinander verwoben sind. Darauf weist schon der Diskussionskontext von *TDE* hin. Zielscheibe seiner berühmten Kritik an der Analytisch-synthetisch-Unterscheidung ist dort ja vor allem die im *Logischen Empirismus* (insbesondere von Carnap) vertretene Ansicht, die Bedeutung von Aussagen sei entweder durch semantische Regeln oder durch Bestätigungsbeziehungen zu einem Bereich möglicher Sinneserfahrung, also in *verifikationistischer* Weise *apriori definierbar*. Dabei verfolgt Quine eine doppelte Argumentationsstrategie: Zum

²⁷⁵ Diese Bezeichnung bezieht sich ganz allgemein auf materialistische Positionen, die die Existenz mentaler Zustände oder Ereignisse leugnen (in welcher Form auch immer). Es besteht jedoch große Unterschiede zwischen dem Behaviorismus Quines und etwa dem Skinners oder auch Carnaps (für den ja *individuelle* mentale Zustände behavioral-dispositional *definierbar* sein sollen).

einen versucht er zu zeigen, dass es *keine zirkelfreie Definition* des Apriorischen geben kann und glaubt daraus folgern zu können, dass die ganze Unterscheidung zwischen apriorischen und synthetischen Sätzen in sich zusammenbricht. Und andererseits tritt er für ein *holistisches* Bild empirischer Signifikanz ein, indem er argumentiert, dass es *keine apriorischen Bestätigungsbeziehungen zwischen Aussagen und Sinneserfahrungen* gibt. Die Instanz der Überprüfung bildeten nicht einzelne Aussagen, sondern immer die ‚Theorie‘ als Ganze: „(...) our statements about the external world face the tribunal of sense experience not individually but only as a corporate body“ (Quine 1953, S. 41). Eine Sprache wird dabei gleichsam als eine Theorie vorgestellt, die wir von der Welt haben. Im Falle widerspenstiger Erfahrungen gäbe es dann jeweils verschiedene Möglichkeiten, die Theorie entsprechend zu modifizieren. Und dies gilt auch im Hinblick auf die Entscheidung, ob eher empirienaher oder eher empirie-ferne Aussagen als wahr aufrechterhalten oder als falsch verworfen werden müssen²⁷⁶. Es gibt nach Quine keine sozusagen ‚auf eigenen Füßen stehenden‘ Einheiten empirischer Signifikanz unterhalb einer Sprache/Theorie als Ganzer. Die unterschiedlichen Möglichkeiten Anpassung an die Erfahrung innerhalb einer Theorie entsprechen dabei – zumindest nach der verifikationistischen Lesart – unterschiedlichen Auswirkungen auf die ‚Bedeutung‘ aller anderen Terme der Theorie.

Nun kann es durchaus unklar erscheinen, wie dieser augenscheinliche semantische Verifikationismus Quines in *TDE* mit dessen später mehr in den Vordergrund tretenden radikalen Naturalismus zusammengehen soll – was insbesondere im Hinblick auf den in *Word and Objekt* durchexerzierten SprachBehaviorismus und die mit diesem einhergehenden These von der *Unbestimmtheit der Referenz* gilt. Ich denke, dass häufig, auch unabhängig von einer verifikationistischen Interpretation von *TDE*, tatsächlich eine grundsätzliche Spannung zwischen den Begriffen und Verständnissen, die Quine zur Erläuterung seiner Position in Anspruch nimmt und seiner naturalistischen Zurückweisung eben solcher Begriffe und Verständnisse besteht²⁷⁷: So soll eine naturalistische Perspektive auf ‚Intentionales‘ einerseits ja nur beobachtbares Verhalten oder, auf einer anderen Ebene der Beschreibung, neuronale Ereignisse zum Gegenstand haben können. Andererseits redet Quine aber auch über Referenz, Bestätigungsbeziehungen oder empirische Äquivalenz, in einer Begrifflichkeit also, deren Sinn doch zunächst nur aus einer ‚vernunft-internen‘ oder normativen Perspektive, bzw. nicht allein als beobachtbare Kausalrelationen, verständlich sein kann. So verwendet er ja etwa auch den Begriff „Aussage“, insbesondere im

²⁷⁶ Quine war allerdings nicht der erste, der einen epistemischen Holismus vertreten hat. Er selbst weist auf Pierre Duhem als einen Vorgänger hin.

²⁷⁷ Auf eine Variante dieser Spannung werden wir auch im Zusammenhang mit Davidsons Quine-Kritik noch eingehen.

epistemischen Argumentationskontext von *TDE*, eigentlich im Sinne von Sätzen, mit denen wir bestimmte Bestätigungsbeziehungen zu anderen ‚Aussagen‘ oder zu Sinneserfahrungen assoziieren, müsste dieses Verständnis aber, gemäß seinem radikal naturalistischen Anspruch, eigentlich in einer Begrifflichkeit rein kausaler Geschehnisse *reformulieren* können. Wie dies in plausibler Weise – also zumindest durch so etwas, wie eine ‚reduktive Entspannung‘ der Unterscheidung zwischen Genese und Geltung – vor sich gehen könnte, hat Quine allerdings nirgends wirklich ausgeführt²⁷⁸. Es wäre, vorsichtig formuliert, jedenfalls durchaus möglich, dass durch eine solche Reformulierung nicht nur die Plausibilität, sondern sogar die grundsätzliche Verständlichkeit seiner epistemisch begründeten Argumente Schaden nehmen würde.

In *Word and Objekt* entfaltet Quine die Konsequenzen seiner These, nach der das, was sich über Bedeutung überhaupt theoretisch sagen lässt dem entspricht, was ein an der Erklärung oder Vorhersage von Sprachverhalten interessierter Beobachter in einer Situation ‚Radikaler Übersetzung‘, also dem Versuch einer sich lediglich an behavioralen Daten und beobachtbaren Umständen orientierenden Übersetzung einer völlig fremden Sprache in die eigene herausbekommen kann. Bekanntermaßen²⁷⁹ behauptet Quine – ganz im Sinne des Holismus von *TDE* – auch hinsichtlich der epistemischen Bedingungen empirischer Theoriebildung über sprachliches Verhalten, dass sich *wechselseitig widersprechende Übersetzungen gleich gut* durch die behavioralen Daten *bestätigt* bzw. zu vergleichbar leistungsfähigen Vorhersagen von sprachlichem Verhalten führen können. Auch hier hätte man, angesichts von Beobachtungen, die mit der bisher in Anspruch genommenen Theorie (dem angefertigten Übersetzungshandbuch) kollidieren, mehrere in explanativer und prognostischer Hinsicht gleich gute Möglichkeiten der Anpassung, zum Beispiel über unterschiedliche Änderungen im ‚Individuationsapparat‘. Eine Entscheidung für eine bestimmte Übersetzung als einzig richtiger lässt sich nach Quine also nicht durch die Fakten erzwingen; unterschiedliche Übersetzungen können empirisch äquivalent sein. (Ich möchte hier nur nochmal darauf hinweisen, dass Quine bei seine Rede von der Beurteilung der Leistungsfähigkeit einer Theorie anhand ihrer empirischen Konsequenzen – gleichsam metasprachlich – eine normative Begrifflichkeit in Bezug auf inferentielle Beziehungen und die Vergleichbarkeit von Theorien *in Anspruch nimmt*, die er eigentlich als ‚unwissenschaftlich‘ desavouieren müsste.)

²⁷⁸ Quine verweist hier lediglich auf den Mechanismus des Spracherwerbs, der – irgendwie – gewissermaßen den Weg von der ‚Kausalität ohne‘ zur ‚Kausalität mit Vernunft‘ eröffne. Vergl. W. v. Quine (1974/1976).

²⁷⁹ Ich verzichte hier auf die Details der einschlägigen Überlegungen Quines im 2. Kapitel von *Word and Objekt*.

Übrigens soll die Unbestimmtheit auch für die unterste Ebene der ‚sensorischen Bestätigungsbasis‘ gelten. Zwar will Quine bestimmten Sätzen eine feststehende ‚Stimulus-Bedeutung‘ im Sinne derjenigen Menge an Reizungen (von sensorischen Nerven) zuordnen, die jeweils unmittelbar ‚Zustimmung‘ oder ‚Ablehnung‘ hervorrufen würde. Das heißt, Sätze sollen ‚kognitiv äquivalent‘ bzw. ‚stimulus-identisch‘ – und in diesem Sinne bedeutungsgleich – sein, wenn ein auf sie bezogenes Zustimmungs- bzw. Ablehnungsverhalten durch die gleichen sensorischen Stimuli hervorgerufen wird²⁸⁰. Offensichtlich entspricht ‚Stimulusidentität‘ aber nicht dem, was wir üblicherweise unter der propositionalen Gleichheit einfacher empirischer Sätze verstehen, da die ‚Identität‘ (bzw. Gleichheit) von Quine ja lediglich ‚physikalistisch‘ bestimmt wird und nach seiner Auffassung der *referentielle* Aspekt selbst auf der Ebene von ‚Stimulus-Bedeutung‘ in den Strudel der Unbestimmtheit hineingezogen wird²⁸¹. Allerdings räumt Quine ein, dass es für unser *Alltagsverständnis* unverzichtbar sein könnte, einen Begriff von Bedeutungsgleichheit zu unterstellen. Lege man an solche Bestimmungen jedoch empirisch exakt zu operationalisierende Kriterien, wie Zustimmungs- oder Ablehnungsverhalten zu Geräuschfolgen unter beobachtbaren Bedingungen an, erwiesen sie sich als für Zwecke wissenschaftlicher Theoriebildung über sprachliches Verhalten weitgehend unbrauchbar.

Quines Dogmen

Für eine kritische Auseinandersetzung mit Quines sprachphilosophischer Position müssen wir uns hier nicht auf eine detaillierte Diskussion darüber einlassen, ob die von ihm in *TDE* vorgebrachten Argumente gegen den Analytizitätsbegriff im Einzelnen stichhaltig sind und worin ihre Konsequenzen genau bestünden²⁸². Offenbar lassen sich

²⁸⁰ Hatte Quine ‚Stimulus-Bedeutung‘ in *Word and Objekt* noch als etwas Intersubjektives verstanden, so neigt er mittlerweile zu einer eher individualistischen Konzeption (vergl. Quine: *Theorien und Dinge* (1981/1985)), da, wie er nun denkt, eine *Gleichheit* sensorischer Stimulation bei unterschiedlichen Sprechern kaum je realisiert sein dürfte.

²⁸¹ Bekanntes Beispiel Quines ist die Übersetzbarkeit von „Gavagai“ als Hase oder unabgetrennter Hasenteil. Diese Relativierung der Ontologie auf ‚Übersetzungshandbücher‘ ist für Quines szientistische Ambitionen insofern unproblematisch, als die Konzeption der *Dinge* ohnehin „nichts weiter (sei) als ein Begriffsapparat, der uns hilft, die Aktivierung unserer Sinnesrezeptoren unter Berücksichtigung früherer Aktivierungen unserer Sinnesrezeptoren vorherzusehen und im Griff zu halten“ (Quine 1981/1985, S. 11).

²⁸² Dies gilt etwa auch für die strittige Frage, ob nicht bereits aus der Kritik an einer klaren Trennbarkeit von Analytischem und Synthetischem, also zwischen solchen Sätzen, die nur aufgrund ihrer Bedeutung wahr sein und solchen, deren Wahrheit auch vom ‚Zustand der Welt‘ abhängen soll, schon die These von der Unbestimmtheit der Bedeutung notwendig folgt. Gegen

im Hinblick auf Quines ‚Bedeutungskeptizismus‘ aber zwei grundlegende Argumentationsmuster hervorheben, die sich deutlich unterscheiden. Das erste Argument können wir das ‚Unbrauchbarkeitsargument‘ nennen und in folgender Weise zusammenfassen:

(I.) A. Alles, was in wissenschaftlich bzw. philosophisch relevanter Weise über Bedeutung gesagt werden kann, muss auf beobachtbaren Fakten im Sinne von Regelmäßigkeiten im Zustimmungs- oder Ablehnungsverhalten zu Lautfolgen (Sätzen) unter bestimmten Bedingungen basieren.

B. Auf dieser Grundlage lässt sich, außer ‚Reizbedeutung‘, kein klarer Begriff von Bedeutungsgleichheit und damit kein explanativ hilfreicher Bedeutungsbegriff gewinnen.

Deshalb:

C. Es gibt keinen philosophisch/wissenschaftlich relevanten Begriff von Bedeutung.

Die andere, epistemische Argumentationsfigur, welche wir das ‚Holismusargument‘ nennen können, lässt sich – eher informell – wie folgt wiedergeben:

(II.) D .Der Gehalt einer Aussage (bzw., die Bedeutung eines Satzes) entspricht den Bedingungen ihrer (seiner) empirischen Bestätigung/Widerlegung im Hinblick auf einen sensorischen Input.

E. Die Bedingungen der empirischen Bestätigung/Widerlegung einer Aussage (eines Satzes) durch einen sensorischen Input hängen von den inferentiellen Beziehungen zwischen den für wahr gehaltenen Aussagen (Sätzen) der ganzen Theorie (Sprache) ab, zu der die Aussage (der Satz) gehört.

Es gibt unterschiedliche, aber kognitiv bzw. explanativ gleich gute Möglichkeiten, die Theorie/Sprache an den sensorischen Input anzupassen.

Deshalb:

G. Der (empirische) Gehalt einer Aussage (die Bedeutung eines Satzes) ist nicht ‚autonom‘ und daher auch nicht apriori bestimmbar.

Diese Vorschläge zu einer zusammenfassenden Formulierung der maßgeblichen Argumentationsmuster Quines lassen im Hinblick auf ihre genauere Interpretation sicher manche Fragen offen. Im Grundsatz scheinen sie mir aber zumindest gängige Verständnisse wesentlicher Züge von Quines bedeutungskeptischen Überlegungen wiederzugeben.

einen solchen Zusammenhang s. etwa: J. Fodor./ E. Lepore: *Holism: A Shopper's Guide* (1993), dafür: P. Boghossian: ‚Analyticity Reconsidered‘ (1996).

Für den unvoreingenommenen Betrachter spiegelt sich in der Prämisse (A) von (I.) offenbar eher ein szientistisches Vorurteil wieder, als dass Quine gute Gründe für sie vorbringt. Dieser Szientismus zeigt sich insbesondere in Quines eigener Beschränkung der möglichen Alternativen zum Behaviorismus auf einen gewissermaßen ‚zügellosen Platonismus‘, die mitzumachen ich aber keine Notwendigkeit sehe. Dagegen lässt sich (II.) auch unabhängig von Quines naturalistischen Voraussetzungen beurteilen. Eine durchaus ähnliche Form der Argumentation hatten ja auch wir gegen funktionalistische Definitionsversuche propositionaler Einstellungen in Anschlag gebracht. Es ist also ratsam, eine eher differenzierte Beurteilung von Quines Argumenten vornehmen.

Zunächst möchte ich Quine im Ergebnis insoweit grundsätzlich Recht geben, als wir häufig keine klare Grenze zwischen material-analytischen (bzw. apriorischen oder nicht revidierbaren) und synthetischen (bzw. empirisch revidierbaren) Inferenzbeziehungen ziehen können, ja es bezüglich vieler unserer begrifflicher Verständnisse womöglich gar keine nicht-revidierbaren Inferenzbeziehungen gibt. Dies meine ich aber nicht, wie Quine, im Sinne von (I.), also als Konsequenz des Anspruchs, dem Begriff der Bedeutung selbst eine empirische Bedeutung zu geben, sondern im Hinblick auf Bedeutungsverstehen *aus der normativen Perspektive von urteilenden Sprechern*. So hat ja insbesondere auch Putnam eindringliche Belege dafür geliefert, dass es in vielen Fällen nicht ganz klar ist, ob es so etwas, wie einen apriorischen deskriptiven ‚Kern-Sinn‘ (was Putnam „Stereotypen“ nennt) von Termini gibt, den wir im Lichte neuer Erkenntnisse nicht zu revidieren bereit wären²⁸³. Zumindest ist es häufig unklar, worin dieser Kern bestehen sollte. Was sollten beispielsweise die analytischen, nicht-revidierbaren Folgerungen sein, die die Bedeutung des Ausdrucks „Tiger“ bestimmen? Ist eine notwendige Bedingung dafür, dass etwas ein Tiger ist, dass es Streifen hat? Wohl kaum, denn es ist durchaus vorstellbar, dass wir Exemplare entdecken, die wir mit guten Gründen als Tiger klassifizieren, die aber keine Streifen haben. Auch ist es sicherlich nicht *unmöglich*, dass die Tiere, die wir „Tiger“ nennen und zoologisch bisher der Gruppe der (Groß-) Katzen zuordnen, eigentlich in eine andere Gruppe gehören. Ja nicht mal, dass Tiger tatsächlich Tiere sind, hält Putnam für ein apriori gültiges Urteil im strengen Sinne, denn wir könnten uns durchaus eine – wenngleich zugegebenermaßen ziemlich abwegige – Situation vorstellen, in der wir dahinterkämen, dass die Dinge, die wir als „Tiger“ bezeichnen, tatsächlich vom Mars ferngesteuerte Roboter sind. Ich würde allerdings meinen, dass es in einem solch extremen Fall nicht mehr viel Sinn machte, zu sagen, wir hätten etwas über *Tiger* gelernt. Eher müssten wir dann wohl sagen, wir hätten festgestellt, dass es eigentlich gar keine Tiger gibt, und dass das, was wir dafür gehalten haben, ferngesteuerte Roboter sind.

²⁸³ Putnam (1975). Vergl. auch: ders.: (1986) und (1988/1991).

Extensionsgleichheit allein liefert nämlich, wie ich denke, kein hinreichendes Kriterium für einen sinnvollen Begriff von intersprachlicher und innersprachlicher (zeitübergreifender und/oder intersubjektiver) Bedeutungsgleichheit, selbst wenn wir mitunter der Bedeutung von Termen ein hohes Maß an ‚Resistenz‘ gegenüber Überzeugungsunterschieden oder -Änderungen zugestehen können. Aber es ist richtig, dass wir bei solchen Fragen wohl keine scharfe apriorische Trennungslinie zwischen bedeutungskonstitutiven und bloß ‚kontingenten‘ Eigenschaften ziehen können, bzw. keine unbezweifelbaren Kriterien für die Unterscheidung zwischen einem Bedeutungswandel einerseits und Überzeugungsunterschieden im Rahmen gleicher Bedeutungsverständnisse andererseits zur Verfügung haben. Entsprechende *Urteile* können dabei unter Gesichtspunkten theoretischer oder praktischer Relevanz mehr oder weniger gut begründet sein.

Wie die meisten der in der Literatur diskutierten Beispiele, fällt auch das Tiger-Beispiel in den Bereich der Begriffe (bzw. Terme) natürlicher Arten. Sicherlich lassen sich aber Unmengen von anders gearteten Beispielen anführen, für die Ähnliches gilt, auch wenn es dabei nicht um ein Entdecken neuer Eigenschaften in Bezug auf eine Extension, sondern um die Variabilität der Extension von Begriffen geht. So will ja auch Wittgenstein am Beispiel des Begriffs des Spiels verdeutlichen, dass das, was unterschiedliche Praktiken zu Spielen macht nicht als ein gemeinsamer definitiver ‚Kern‘ verstanden werden sollte, sondern eher im Sinne eines lockeren Verbundes von Gemeinsamkeiten, von ‚Familienähnlichkeiten‘, wobei keine klare Grenze der Erweiterung vorgegeben ist.

Allerdings sollte man hier nicht überziehen, indem man sich, um nochmals mit Wittgenstein zu reden, von einer einseitigen Diät von Beispielen ernährt. Es gibt doch offenbar eine Vielzahl von Fällen, in denen es einen guten Sinn hat, von ‚analytischen‘, nicht-empirischen Wahrheiten zu sprechen wie etwa, dass Rot eine Farbe ist oder dass Demokratie eine Form sozialer Verhältnisse charakterisiert. Diese Beispiele stellen natürlich nicht Fälle einer *definitiven* Synonymie dar, wie sie Quine besonders interessieren, und die ohnehin eher die Ausnahme darstellen (wie „Junggeselle“ bedeutet „unverheirateter Mann“). Ich meine begriffliche Zusammenhänge, die für unsere jeweiligen Verständnisse so zentral sind, dass wir der Behauptung, es könnte anders sein (jedenfalls außerhalb sehr besonderer Kontexte), keinen klaren Sinn geben können. Ein solches Verständnis legt uns weder auf die Annahme von eigenständigen ‚Bedeutungsentitäten‘, noch auf die platonistische Vorstellung fest, Begriffe entsprächen der Bedeutung von sprachlichen Ausdrücken im Sinne von Regeln, durch die alle bisherigen und zukünftigen Möglichkeiten ihrer richtigen Verwendungen eindeutig bestimmt sind. Wie wir uns in unklaren, neuartigen oder strittigen

Anwendungsfällen über die Bedeutung sprachlicher Ausdrücke (bzw. über unser begrifflichen Verständnisse) einigen bzw. verständigen können, ist eine mehr oder minder offene und mitunter recht komplexe, auch von Relevanz- und kontextbedingten Faktoren abhängige Angelegenheit. Von einer entsprechenden Unklarheit oder Offenheit zu reden macht jedoch nur Sinn auf der Grundlage der Möglichkeit vernünftiger Urteile hinsichtlich einer begrifflichen Kontinuität in der Ausdrucksverwendung.

Aus einer begründeten Skepsis gegenüber einer *klaren Trennbarkeit* analytischer und synthetischer Fragen folgt jedenfalls *nicht*, dass es *keinen Sinn* hat, überhaupt zwischen Fragen der Bedeutung und empirischen Fragen zu *unterscheiden*. Hier wie anderswo können Unterscheidungen auch dann wichtigen Zwecken dienen, wenn es im Einzelfall keine am Maßstab physikalischer Genauigkeit bemessenen präzisen Grenzlinien gibt. Selbst im Hinblick auf Terme für natürliche Arten muss eine solche Unterscheidung aufrechterhalten werden, denn um überhaupt von Entdeckungen von Eigenschaften hinsichtlich der Extension eines Terms reden zu können, brauchen wir *irgendwelche* (mehr oder minder variablen) *deskriptiven* Reidentifikationsmerkmale, die konstant bleiben. Die für uns primär relevanten, gewissermaßen an der ‚Erscheinung‘ orientierten, (Re-) Identifikationsmerkmale etwa von Tigern würden sich ja auch durch die Entdeckung, dass sie vom Mars ferngesteuerte Roboter sind kaum ändern. Und selbst wenn die Erscheinungsweise sich radikal ändern sollte, müssen wir auf einen gleichsam ‚verborgenen‘ kausalen Zusammenhang zurückgreifen, der die veränderten Exemplare genealogisch *an diejenigen koppelt*, deren Erscheinungsweise den Standardgebrauch *ursprünglich* anleitete. Zur Bedeutung des Terms „Tiger“ muss es also zumindest gehören, dass etwas, was wir noch zu Recht so bezeichnen können, *in irgendeiner Hinsicht* – ob hinsichtlich zu entdeckender, natürlicher Eigenschaften, wie etwa eine gemeinsame Naturgeschichte *oder* hinsichtlich der ‚bloßen‘ Erscheinung (den ‚Stereotypen‘) – unseren primären, nämlich *am Erscheinungsbild orientierten Paradigmen* für die standardmäßige Verwendung des Ausdrucks „Tiger“ gleichen muss.

Da Quine aber der Ansicht ist, dass die fragliche Unterscheidung nach empirisch-wissenschaftlichen Kriterien nicht haltbar ist, spricht dies, wie ich meine, eher dafür, dass dessen Anspruch an einen brauchbaren Bedeutungsbegriff fragwürdig ist, als dieser Begriff selbst. Dementsprechend kann man Quines Kritik am Bedeutungsbegriff ihrerseits (bzw. seinen behavioristischen Residualbegriff von Bedeutung seinerseits) in mehrfacher Hinsicht kritisieren. So würde ich der frühen Kritik von Grice und Strawson²⁸⁴ insofern zustimmen, als Quine in seiner Argumentation in szientistischer

²⁸⁴ P. Grice/ P. F. Strawson: ‚In Defence of a Dogma‘ (1956).

Manier Exaktheitsmaßstäbe an den Begriff der Bedeutung heranträgt, die dem Sinn unserer gängigen Praktiken der Klärung von Fragen der Bedeutungsgleichheit oder -Invarianz (bzw. der Bedeutung ganz allgemein) einfach nicht angemessen ist. Diese Kritik gegenüber Quines Ansprüchen an einen passablen Bedeutungsbegriff findet ihre Entsprechung in Wittgensteins Bemerkung, Exaktheit sei relativ auf Zwecke (Phil. Unt. § 88). Exaktheit ist ja kein Wert an sich, der wie selbstverständlich an sprachliche Bedeutung herangetragen werden kann oder sollte. Unser Verständnis von Bedeutung oder Bedeutungsinvarianz bemisst sich an den jeweiligen praktischen Möglichkeiten und Notwendigkeiten, entsprechende Fälle von in gewisser Hinsicht gleichen Verwendungen eines sprachlichen Ausdrucks zu beurteilen. Die Erwiderung Quines hierauf lautet notorisch, dies sei umso schlimmer für unseren alltagsüblichen Bedeutungsbegriff. Diese Riposte ist jedoch unangebracht, da sie ihrerseits auf der allgemeinen szientistischen Voraussetzung beruht, dass Fragen der Bedeutung, genau wie jedes andere Phänomen, welches überhaupt zum Thema weitergehender intellektueller Bemühungen werden kann, einer empirisch-wissenschaftlichen Behandlung zugänglich sein müssen. Doch diese Voraussetzung ist solange lediglich ein Dogma, wie für sie keine unabhängigen Gründe vorgebracht werden. Im Gegenteil scheint ja eher einiges dafür zu sprechen, dass ein Begriff, der sich so weit von unseren üblichen Verständnissen entfernt hat, wie der Bedeutungsbegriff Quines, nicht mal mehr beanspruchen kann, ein ‚entmythologisiertes‘, auf die ‚harten Fakten‘ heruntergekochtes Residual des entsprechenden, vorthoretisch bereits verfügbaren Begriffs darzustellen. Darüber hinaus scheint Quines Exaktheitsanspruch *per se* auch keinen wirklichen Nutzen im Sinne einer hilfreichen Kritik unseres ‚vorphilosophischen‘ Verständnisses der unterschiedlichen Facetten des Phänomens Bedeutung zu haben. Ich möchte dafür argumentieren, dass die ‚Unwissenschaftlichkeit‘ unserer Bedeutungsverständnisse, bzw. gerade deren ‚Ungenauigkeit‘ einen guten, nicht-eliminierbaren (Orientierungs-)Sinn hat und zudem ein wesentliches Element der Möglichkeit von Erkenntnisfortschritt bildet.

Die Quine besonders interessierenden Fälle, nämlich definatorische Synonomiebeziehungen zwischen unterschiedlichen sprachlichen Ausdrücken, stellen ja eher semantische Sonderfälle dar²⁸⁵. Synonyme im genannten Sinne stehen uns meist gar nicht zur Verfügung, wenn wir die Bedeutung eines Ausdrucks ‚innersprachlich‘

²⁸⁵ Ein auf solche Sonderfälle gestütztes Wissen scheint es allerdings zu sein, auf das sich, wie Searle zu bedenken gibt, auch Quine *intuitiv* berufen muss, wenn er bestimmte Vorschläge für eine kriteriale, bzw. definatorische Bestimmung von Analytizität als untauglich *beurteilt* (Searle (1969), S. 6 f.). Das heißt, die klaren Fälle bilden offenbar auch für Quine Paradigmen, welche die Verständnisgrundlage bilden, auf der Quine die einzelnen Versuche einer umfassenden extensionalen, bzw. behavioralen Bestimmung oder Definition von Synonomie verwerfen kann.

thematisieren. Im Hinblick auf die meisten Fällen einer expliziten verbalen Bedeutungs(er)klärung ist es wohl angemessener, von *Erläuterungen*, die zugleich auch den präskriptiven Sinne von *Vorschlägen* haben können, als von exakten Definitionen reden. *Dass* unsere sprachlichen Ausdrücke meist keine scharfen Verwendungskriterien haben, ist, wie ich meine, ein wesentlicher Aspekt ihrer begrifflichen Funktion. Erst dadurch wird nämlich ermöglicht, was wir als *bedingte kreative Offenheit* unserer begrifflichen Verständnisse bezeichnen können, also die Möglichkeit klassifikatorischer *Gleichbehandlung auch von nicht-trivial neuen Fällen*, das heißt solchen, welche sich zwar in bestimmten Hinsichten von den bisherigen Fällen unterscheiden, in anderen, für die gemeinsame Klassifikation *als maßgeblich erachteten*, jedoch nicht (denken wir beispielsweise an mögliche Kontexte der Verwendung des Wortes „Musik“). Gemäß der in dieser Arbeit verteidigten Sicht der Dinge entspricht die bedingte kreative Offenheit sprachlicher Bedeutung *flexiblen* (wenngleich nicht beliebigen) *normativen Spielräumen* richtiger Unterscheidung und Klassifizierung, mit relativ klaren, über sich selbst hinausweisenden Paradigmen im ‚Zentrum‘ und einer zu den ‚Rändern‘ hin abnehmenden Selbstverständlichkeit im Hinblick auf die richtige Verwendung von sprachlichen Ausdruckstypen²⁸⁶. Selbst der Übergang zum metaphorischen Gebrauch kann dabei fließend sein.

Eine weitere wichtige Beschränkung, der jede philosophische Theorie gerecht werden muss, die überhaupt beanspruchen will, nicht gänzlich den Anschluss an zentrale Intuitionen unserer Bedeutungsverständnisse zu verlieren, besteht nach meinem Dafürhalten in der inhaltlichen Unterscheidbarkeit zwischen den Überzeugungen und den Mitteilungsabsichten eines Sprechers und der Bedeutung seiner Worte. Von besonderer Brisanz ist diese Beschränkung natürlich für alle im weiteren Sinne ‚intentionalistischen‘ Theorien (wie natürlich das IBS-Programm, aber wie wir sehen werden auch für Davidson). Einen spezieller, aber sehr wichtiger Aspekt dieser Beschränkung, den wir bereits im Zusammenhang mit Putnams referenztheoretischen Argumenten angesprochen hatten, betrifft insbesondere holistische Theorien, wie die Quines: Denn eine Theorie, die (in gewissermaßen fregeanischer Weise) Bedeutung eng an Überzeugungen koppelt und dabei den Bezug dem Sinn unterordnet, zugleich aber auch eine starke Form von Holismus hinsichtlich der Bestätigung oder Widerlegung von Überzeugungen bzw. Überzeugungssystemen vertritt, hat ja offenbar zur Konsequenz, dass jede Veränderung im Überzeugungssystem eine Veränderung der Bedeutungen der Sprache (des Sprechers oder der Sprecher) nach sich ziehen müsste. Und wie wir oben bereits einwendeten,

²⁸⁶ Und dies gilt am Ende auch, wenngleich auf erhöhtem Präzisionsniveau, für technische oder wissenschaftliche Termini.

hätte dies zur Folge, dass die Welt, die in der Sprache vor der Überzeugungsveränderung beschrieben wird in gewissem Sinne nicht mehr die gleiche wäre wie jene, welche die nachfolgende Sprache beschreibt. Überzeugungsänderungen könnten dann nicht mehr als Lernen oder Erkenntnisfortschritt begriffen werden, denn dazu ist doch offenbar ein *partiell überzeugungsinvarianter Aspekt sprachlicher Bedeutung* im Sinne einer (partiell) überzeugungsinvarianten Bezugnahme notwendig, der eine Unterscheidung ermöglicht zwischen einer Überzeugung, bzw. einer Überzeugungsänderung und demjenigen, *bezüglich dessen* wir unsere Überzeugung geändert haben. Diese Unterscheidung aufrechterhalten zu wollen, bedeutet natürlich nicht, zu behaupten, dass eine Beurteilung von Bedeutungsgleichheit als unabhängig von *jeglicher Überzeugung über* das als gleich Beurteilte möglich wäre. Aber es heißt zumindest, dass es Beurteilungsmaßstäbe für die Gleichheit oder Identität (von Eigenschaften bzw. Dingen) geben muss, die ein gewisses Maß an Überzeugungsveränderungen transzendieren.

Ein solches Verständnis von Bedeutung im Sinne von (partieller) Überzeugungsinvarianz scheint jedoch mit Quines Holismus unverträglich, insbesondere dann, wenn man wie Quine auch noch die Unbestimmtheit des Bezugs behauptet. Zwar postuliert Quine ‚Stimulus-Bedeutung‘ als dasjenige, was vom Holismus – und damit der Unbestimmtheit – nicht betroffen sei und für das es präzise Identitätskriterien gäbe. ‚Stimulus-Bedeutung‘, was immer man von ihr halten mag, kann jedoch nicht die benötigte Form von Bedeutungsinvarianz liefern, denn sie beinhaltet nicht jenen Bezug nehmenden Aspekt, den man für den Begriff einer Welt der Dinge braucht, über deren Eigenschaften man etwas lernen kann²⁸⁷. Eine ‚Stimulus-Bedeutung‘ sollen ja nur Beobachtungssätze (‚Gelegenheitssätze‘) haben, denen alle Mitglieder einer Sprachgemeinschaft in einer bestimmten Situation allein aufgrund ihrer ‚Sinneserfahrung‘ (bestimmter Nervenreizungen) zustimmen würden. Solche Äußerungen kann man aber kaum mit gutem Recht als propositionale Sätze begreifen, in denen strukturell die Unterscheidung von Prädikation und (situationstranszendierender) Identifikation angelegt ist. Dies ist auch Quines eigene Deutung: „(...) it seems clear, that (...) reification of bodies over time is beyond the reach of observation sentences and categorials“²⁸⁸. Wenn dem so ist, dann fragt es sich allerdings ganz grundsätzlich, wieso Quine glaubt, solche ‚Sätze‘ so verstehen zu können, dass sie bei einer Gelegenheit *wahr*, bei einer anderen *falsch* sein könnten. Ich werde noch ausführlich dafür argumentieren, dass wahr oder falsch nur propositional

²⁸⁷ Vergl. Putnam (1986). Für eine in diese Richtung gehende Kritik Quines vergl. auch D. Føllesdal: ‚Essentialism and Reference‘ (1986).

²⁸⁸ Quine: *Pursuit of Truth* (1992), S. 25. Mit „categorials“ meint Quine hier *rein prädikative* Verallgemeinerungen im Sinne von „wenn immer dies, dann jenes“.

strukturierte Sätze oder Äußerungen sein können, was aber die Möglichkeit von durch die Zeit reidentifizierbaren Körpern oder Gegenständen voraussetzt und damit von vornherein jenseits einer ‚rein prädikativen‘ Deutbarkeit liegt. An dieser Stelle begnüge ich mich zunächst mit der Feststellung, dass die einzige, gegenüber der jeweiligen Theorie (einem Überzeugungssystem) neutrale, sprachliche Grundlage, welche unter den Voraussetzungen Quines einen ‚zweisprachlichen‘ Vergleich ermöglichte, die ‚Stimulus-Bedeutung‘ von Beobachtungssätzen ist, die aber, wie Quine selbst einräumt, keinen Begriff einer objektiv erfahrbaren Welt von Dingen hergibt. Ohne einen solchen Begriff können wir aber, wie ich denke, nicht einmal davon reden, eine neue Theorie sei dann besser, wenn sie zukünftige ‚Beobachtungen‘ – sprich: Reizungen von Nervenenden – besser vorherzusagen ermöglichte, als die vorhergehende.

Quine bleibt in dieser Hinsicht selbst einem *empiristischen Dogma* treu, wenn er glaubt, dass *epistemisch Primäre* seien Nervenreizungen, aus denen wir dann irgendwie die materiellen Dinge (‚Körper‘) der Welt ‚aufbauen‘²⁸⁹. Mit seiner naturalistischen Transformation des Empirismus scheint Quine zwar den gegen den ‚Mythos des Gegebenen‘ (Sellars), also die ‚einfachen Ideen‘ oder ‚Sinnesdaten‘ der alten Empiristen vorgebrachten Einwände zu entgehen. Allerdings hat er dafür mit seiner Rede von Nervenreizungen zugleich in kategorialer Weise *das Thema* gewechselt. Ich habe bereits im Zusammenhang mit der Kritik am Kausal-Funktionalismus darauf aufmerksam gemacht, dass keine gangbare begriffliche Brücke von der physischen Redeweise über Nervenreizungen zur epistemischen (‚mentalistischen‘) Redeweise über Beobachtungen, wahre empirischen Überzeugungen, Erfahrungen, Wahrnehmungen usw. führt. Nur diese stellen aber jene Art von inferentiell relevanten Instanzen dar, an denen unsere ‚Theorien‘ sich bewähren oder scheitern können. Bloße Nervenreizungen, als die Quine unseren ‚epistemischen Input‘ letztlich bestimmen will, haben keine inferentielle Relevanz. Wir können aus ihnen nichts erschließen und sie folgen auch aus nichts. Sie sind als physische Ereignisse Gegenstand kategorial deskriptiver Erklärungen der kausalen Zusammenhänge der natürlichen Welt. Es gilt hier ein klares Entweder-Oder: Entweder begnügt sich Quine mit ‚ontologisch einwandfreien‘ physische Entitäten, allerdings auf Kosten eines Themawechsels; oder aber er beansprucht, einen – empiristisch gestimmten – Beitrag zu epistemologischen Fragestellungen zu liefern, der dann allerdings in hergebrachter Weise der *philosophischen* Kritik ausgesetzt ist. Wenn, wie Quine in *Die Wurzeln der Referenz*

²⁸⁹ Vergl. Quine (1974/1976), § 1 f. Quine lässt dabei allerdings immerhin angeborene Mechanismen der (gewissermaßen ‚ontogenetisch-apriorischen‘) Strukturierung der ‚Erfahrung‘ zu, Mechanismen, deren Angemessenheit durch evolutionäre Anpassungsprozesse erklärt werden könnte.

(§1) selbst einräumt, ‚Ideen‘ oder ‚Sinnesdaten‘ nicht dazu geeignet sind, aus ihnen die ‚Welt der Körper‘ zu erschließen, so erst recht nicht Nervenreizungen²⁹⁰.

Jedenfalls kann Quines ‚Stimulus-Bedeutung‘ von Beobachtungssätzen nicht die für ein Verständnis der kognitiven Vergleichbarkeit von unterschiedlichen Überzeugungssystemen oder Theorien notwendige überzeugungs- bzw. sinn-invariante Komponente von Bedeutung liefern. Dazu bedarf es der Möglichkeit *singulärer Bezugnahme*, die von vornherein auf der Ebene der Propositionalstruktur analysiert werden müssen. Diese semantische Funktion sollen für Quine aber nur *innertheoretische Variablen* übernehmen:

„I see reference, reification, and ontology no longer as a goal of science, but rather as a spin-off of quantification and the variables, these being in turn a mere technical aid in the forging logical links between observation sentences and theoretical sentences.“ (Quine: ‚Reply to Dagfinn Føllesdal‘ (1986), S. 115)

Mit dieser Depotenzierung der Bezug nehmenden Aspekte unserer Bedeutungsverständnisse beraubt sich Quine der Möglichkeit, jene unterschiedlichen identifizierenden und reidentifizierenden Potentiale von Sprachen heranzuziehen, durch die erst eine brauchbare Form von Theorien- oder Überzeugungswandel übergreifender, bzw. vergleichender Beurteilbarkeit verständlich wird. Lediglich über Variablen zu quantifizieren, deren Werte durch die gerade gültige wissenschaftliche Theorie bestimmt sind, leistet dies ja gerade nicht²⁹¹.

Quines Geringschätzung der semantischen Funktion von Eigennahmen ist zum einen darin begründet, dass er vermeiden will, sich durch unsere jeweils vorgegebene Sprache auf irgendeine Ontologie festlegen zu lassen²⁹²; meine Zweifel an der Möglichkeit eines solchen Verzichts habe ich bereits vorgetragen. Zum anderen gehört es ja zu Quines empiristischen Grundannahmen, dass die nicht nur dem Linguisten, sondern auch normalen Teilnehmer an Verständigungen und spracherlernenden Kindern allein zugängliche Grundlage des Bedeutungsverstehens Äußerungen von *ganzen Sätzen* unter beobachtbaren Bedingungen seien²⁹³. Diese Annahme ist aber *per*

²⁹⁰ Weshalb auch Davidson, wie wir unten noch sehen werden, mit Quines Konzept von ‚Stimulus Bedeutung‘ bricht.

²⁹¹ Darüber hinaus folgert Quine aus der These von der Unbestimmtheit bzw. Unerforschlichkeit der Bezugnahme auch eine *Relativität der Ontologie*, nach der diese, solange Struktur und empirischer Gehalt (im quineschen Sinne) nicht betroffen sind, auch bezüglich *einer* Sprache/Theorie „ohne wirklichen Verlust“ variiert werden kann (u. a. in Quine (1969)). Diese Radikalisierung muss uns hier aber nicht weiter interessieren (Davidson folgt Quine hierin auch nicht, da er das Postulat einer ontologischen Relativität, wie jede Art von Kultur- oder Sprachrelativismus, für selbstwidersprüchlich hält).

²⁹² Quine (1948).

²⁹³ ‚Sätze‘ im semantischen Sinne, also auch etwa Einwortsätze wie „Feuer!“.

se alles andere als selbstverständlich. Nicht nur Linguisten, sondern auch normale Sprecher wissen doch offenbar erheblich mehr, als in der Konzeption der Situation Radikaler Übersetzung zugelassen wird. Quines epistemische Beschränkungen können also nur in dem grundsätzlichen Sinne gemeint sein, dass jede Art von Theoriebildung am Ende auf Beobachtbares angewiesen ist. Damit vertritt Quine aber nicht nur eine Version der These vom öffentlichen Charakter all dessen, was für Bedeutung relevant ist (*nothing is hidden*) – eine These, der ich in ihrer allgemeinsten Form ohne Abstriche zustimme – sondern er gleicht Bedeutungsverstehen zugleich dem Verstehen kausaler Zusammenhänge auf der Grundlage wahrer empirischer Theorien über die Welt an – ein Vorgehen, das zu kritisieren eines der Hauptziele dieser Arbeit bildet.

Was den primären Spracherwerb betrifft, so genügen Entwicklungsstufen, auf denen Kinder noch keinen ‚Individuations-‘ bzw. ‚Identifikationsapparat‘ zur Verfügung haben, ohnehin nicht jenem Mindestmaß an sprachlicher Bedeutungskompetenz, um die es bei unseren Überlegungen geht. Und ich hatte bereits meine Überzeugung zum Ausdruck gebracht, dass jede Rede von Sätzen (bzw. Äußerungen), die wahr oder falsch sein können, bereits eine semantische bzw. praktische Kompetenz bezüglich der Subjekt-Prädikat-Struktur unterstellt. Wenn dies richtig ist, macht es grundsätzlich keinen Sinn, entsprechende Hinweise, wie etwa auf die wichtige ‚rigide‘ Bezugnahme-Funktionen von Eigennamen, durch die Behauptung einer semantischen Priorität des Sinns ganzer Sätze oder einer epistemischen Priorität (der Beobachtbarkeit) von vollständigen Äußerungen abwehren zu wollen. Ich werde im Zusammenhang mit der Erörterung der wahrheitssemantischen Position Davidsons dafür argumentieren, dass es plausibler ist, das Verstehen von Sätzen (bzw. Äußerungen), semantischen Strukturen und ihren subsententiellen Elementen als wechselseitig aufeinander verweisend anzusehen²⁹⁴.

Aus einer gewissen Distanz betrachtet, spricht manches für den gelegentlich²⁹⁵ gegen Quine erhobenen Vorwurf, die These von der Unbestimmtheit der Bedeutung laufe eher auf eine *reductio ad absurdum* hinaus, als auf eine plausible sprachphilosophische Position. Für mich stellt sich jedenfalls die Frage, wie wir (also auch Quine) überhaupt Begriffe sprachlich zur Verfügung haben – und ihre Verständlichkeit unterstellen – können, wenn diese keine sprachliche Grundlage haben sollen, die sie von anderen Begriffen unterschiedbar macht (wie ja etwa „Hase“,

²⁹⁴ Quines Vernachlässigung der propositionalen Struktur geht, wie ich denke, Hand in Hand mit seinem ‚deflationistischen‘ Verständnis des Wahrheitsbegriffs. Eine Kritik von Quines Theorie des Bedeutungsverstehens betrifft demnach auch dieses. Und in mancher Hinsicht gilt ähnliches auch für Davidson.

²⁹⁵ Beispielsweise auch in Searle: ‚Indeterminacy, Empiricism and the First Person‘ (1987) und Boghossian (1986).

„unabgetrenntes Hasenteil“ und „Hasenstadium“ auf der Ebene beobachtbarer Fakten im Hinblick auf die gesamte Evidenz äquivalent sein sollen). Wie kann uns Quine seine Beispiele überhaupt mitteilen und verständlich machen, wie hat er die entsprechenden Unterscheidungen unter beobachtbaren Umständen erlernen können, wenn die gesamte verfügbare Evidenz doch ihnen gegenüber indifferent sein soll?

Unstrittig dürfte sein, dass Quines spektakuläre Thesen zur Bedeutung weitgehend von den von ihm befürworteten Beschränkungen hinsichtlich der Möglichkeit abhängen, sprachliche Bedeutung und Verständigungskompetenz zu thematisieren. Ähnliches gilt auch für Davidson, insbesondere was die folgenreiche These anbelangt, dass propositionale Einstellungen und sprachliche Bedeutung nur in wechselseitiger Abhängigkeit voneinander zugeschrieben werden können. Und trotz mancher Unterschiede zu Davidson (vor allem in der Gewichtung des Faktors Sprache), teilt auch Dennett, zumindest in wesentlichen Punkten, Quines Beschreibung der epistemischen Ausgangsposition des Verstehens ‚intentionaler‘ Phänomene, eine damit einhergehende Unbestimmtheitsthese²⁹⁶ und eine Varianten des ‚Prinzips der Nachsicht‘ (‚*principle of charity*‘²⁹⁷). Von Quine in *Word and Objekt* zunächst nur als methodologische Maxime zur Übersetzung der logischen Konstanten in Anschlag gebracht, erfährt dieses Prinzip in den Positionen Davidsons und Dennetts allerdings eine erhebliche Aufwertung im Sinne von *gegenstandskonstitutiven Optimierungsannahmen*, die als *wesentliche Eigenheit* einer ‚Interpretationstheorie‘ bzw. einer ‚Intentionalen Psychologie‘ angesehen werden. Die Pointe dieser Annahmen besteht darin, dass wir Handlungen oder Verhalten gar nicht *verstehen* könnten, wenn wir nicht ein bestimmtes Quantum an *Rationalität bzw. Übereinstimmung und Wahrheit unterstellen*. Die wichtige Rolle, welche ‚Nachsichtigkeitsprinzipien‘ für den Interpretativen Materialismus spielen, muss allerdings im Kontext der jeweiligen theoretischen Annahmen Davidsons und Dennetts erläutert werden. Beginnen wir zunächst mit Davidson.

²⁹⁶ Dennett und Davidson vertreten allerdings abweichende Ansichten hinsichtlich der Tragweite der Unbestimmtheitsthese. Unter den genannten Voraussetzungen spricht nach meiner Auffassung einiges dafür, die These von der Unbestimmtheit der Bedeutung/Referenz, wie Dennett es tut, auch auf propositionale Einstellungen auszuweiten und von einer allgemeinen Unbestimmtheit des ‚Intentionalen‘ zu reden (ähnlich auch Føllesdal in ders.: ‚Indeterminacy and Mental States‘ (1990)).

²⁹⁷ In der neueren Debatte ist dieser Ausdruck von N. Wilson (ders.: ‚Substances without Substrata‘ (1959)) geprägt worden.

Davidsons ‚integrierte Theorie der Bedeutung und des Handelns‘

Obwohl sich Davidsons bedeutungstheoretische Überlegungen eng an Quine anlehnen, vermeidet er doch weitgehend dessen behavioristische Rhetorik. So scheut sich Davidson nicht, propositionale Einstellungen als empirische Bestätigungsbasis einer Bedeutungstheorie zuzulassen und ‚Für-wahr-Halten‘ an die Stelle von ‚Zustimmungsverhalten‘ treten zu lassen²⁹⁸. Außerdem schwebt ihm eine theoretische Beschreibung des *Bedeutungsverstehens* vor. Dies leiste jedoch kein Übersetzungshandbuch im Sinne Quines, sondern nur eine ‚Interpretationstheorie‘, da wir wissen könnten, welche Sätze einer Sprache Übersetzungen der einer anderen wären, ohne dabei eine der beiden Sprachen tatsächlich verstehen zu müssen. Dazu bedürfe es einer Theorie, welche ein Verständnis ermöglichende Interpretation einer natürlichen Sprache liefert, weshalb Davidson von „Radikaler Interpretation“ anstatt von „Radikaler Übersetzung“ redet. Allerdings scheint mir Davidson den Unterschied über zu betonen. Denn natürlich ist auch Quine an mehr interessiert, als der Übersetzung einer unbekanntes Sprache in eine andere unbekanntes, so dass die Differenz keine grundlegende, sondern eher terminologischer Art sein dürfte.

Davidsons Vorschlag ist jedenfalls, dass eine Bedeutungstheorie für eine natürliche Sprache die Form einer Wahrheitstheorie im Stile Tarskis annehmen sollte²⁹⁹. Er sucht nach einer Theorie, deren Kenntnis einen Sprecher in die Lage versetzte, die Äußerungen eines anderen Sprechers zu verstehen (bzw. zu ‚interpretieren‘), und er glaubt, dass Tarski uns die wesentlichen formalen Mittel hierfür geliefert hat. Während jedoch Tarski Bedeutung (bzw. Bedeutungsgleichheit) als gegeben voraussetzte, um den Wahrheitsbegriff zu analysieren, schwebt Davidson das Umgekehrte vor, also etwas über Bedeutung zu erfahren, indem er Wahrheit als zentralen, undefinierten Begriff verwendet. Damit ist natürlich bereits eine wichtige Vorentscheidung darüber gefallen, in welcher Form und in welchem Sinne diejenigen kognitiven Fähigkeiten beschrieben werden sollten (bzw. können), die kompetente Kommunikationsteilnehmer verfügen und die sie in die Lage versetzen, andere Sprecher zu verstehen und sich mit ihnen zu verständigen.

Eine weitere wichtige, an Quine anlehrende, Vorentscheidung Davidsons besteht darin, Bedeutungskompetenz an empirisches Wissen bzw. die Kenntnis einer empirischen Theorie anzugleichen. Davidson versteht eine verständnisermöglichende ‚Interpretationstheorie‘ als eine empirische Theorie, die bestimmten Adäquatheitsbedingungen genügen muss. So dürfe eine solche Theorie, neben der

²⁹⁸ Vergl. insbes. Davidson (1967) und (1973).

²⁹⁹ A. Tarski: ‚Der Wahrheitsbegriff in formalisierten Sprachen‘ (1935) und ders.: ‚Die semantische Konzeption der Wahrheit und die Grundlagen der Semantik‘ (1944).

Forderung, eine ‚Interpretationen‘ für jeden möglichen Satz zu liefern, den ein Sprecher (oder eine Gruppe von Sprechern) äußert, vor allem keine Begriffe wie „übersetzen“, „Bezug nehmen (auf)“ oder ähnliche verwenden, da diese als semantische (bzw. intensionale) Begriffe ihrerseits genauso erklärungsbedürftig seien, wie der Bedeutungsbegriff selbst. Wenn man sich dementsprechend zunächst auf eine Bedingung *extensionaler Adäquatheit* für eine Wahrheitstheorie im Stile Tarskis beschränkt und auf intensionale Begriffe verzichtete, müssten aus einer solchen Theorie, wenn sie richtig ist, alle wahren Theoreme (W-Sätze) der folgenden Form herausspringen:

S ist (in der natürlichen Sprache L) wahr genau dann, wenn p

Dabei steht „ S “ für einen ‚Namen‘ eines objektsprachlichen Satzes, während durch Einsetzungen für „ p “ dem so identifizierten Satz eine propositionale Beschreibung in der Metasprache zugeordnet wird. Sätze, die dieser sogenannten ‚Konvention W‘ genügen, wären natürlich auf Sprecher und Zeitpunkte zu relativieren um den indexikalischen Aspekten sprachlicher Bedeutung gerecht zu werden. Ist es aber überhaupt plausibel anzunehmen, dass bei allen möglichen Instantiierungen von W-Sätzen der durch die Variable „ p “ markierte Teil solcher Bikonditionale die *Bedeutung* eines objektsprachlichen Satzes einer Sprache L auf der linken Seite des Bikonditionals angeben würde? Nach Tarski sollten sich aus einer angemessenen Wahrheitstheorie alle wahren W-Sätze durch ein rekursives Verfahren aus einer endlichen Anzahl wahrer Axiome folgern lassen (allerdings nur für eine formalisierte Sprache). Davidson muss jedoch mit dem Problem fertig werden, dass eine ausschließliche Beschränkung auf die Bedingung *extensionaler Adäquatheit* auch W-Sätze, wie (W_1) zuließe:

(W_1) „Schnee ist weiß“ ist wahr (in Deutsch) genau dann, wenn das Gras grün ist.

Obwohl (W_1) der Konvention W genügt, sagt er uns natürlich nichts über die Bedeutung von „Schnee ist weiß“. Ohnehin zeigen einschlägige Beispiele, wie etwa „Wasser“ und „ H_2O “, dass Extensionsgleichheit allein keine hinreichende Bedingung für so etwas wie Bedeutungsgleichheit, richtige Übersetzung oder ‚Interpretation‘ sein kann (denn sicherlich sind „Wasser“ und „ H_2O “ nicht bedeutungsgleich). Es lassen sich mindestens zwei Strategien unterscheiden, mit denen Davidson dieses Problem zu lösen versucht.

Der erste, von Davidson in ‚Wahrheit und Bedeutung‘ vorgeschlagene, Lösungsversuch besteht darin, es für die Bedeutung (bzw. richtige ‚Interpretation‘) objektsprachlicher Sätze nicht auf einzelne W-Sätze ankommen zu lassen, sondern auf die Semantik ihrer lexikalischen Bestandteile plus ihrer syntaktischen Struktur bzw. darauf, wie die Theoreme aus der Theorie abgeleitet werden. Eine angemessene Wahrheitstheorie müsse ja wahre W-Sätze für alle möglichen Instantiierungen liefern, in denen die subsententiellen Bestandteile jeweils vorkommen können. Auch eine extensionale Theorie könne dann den Unterschied etwa der Bedeutung von „Schnee ist weiß“ und „Gras ist grün“ erfassen, weil aus ihr beispielsweise kein (falscher) W-Satz, wie etwa (W₂)

(W₂): „Schnee ist weiß“ ist dann und nur dann wahr, wenn das Gras weiß ist.

folgen darf. Dieser Vorschlag läuft darauf hinaus, den *holistischen* Charakter von Bedeutung und propositionalen Einstellungen ernst zu nehmen, das heißt, Bedeutung lässt sich nach Davidson einzelnen Äußerungen von Sprechern nur im Rahmen einer Theorie *aller* ihrer möglichen Äußerungen und Einstellungen zuschreiben.

Den zweiten Vorschlag diskutiert Davidson in seiner ‚Replik auf Foster‘ (1976). Danach seien W-Sätze als *Gesetze* zu verstehen, also als empirische Generalisierungen in Bezug auf Sprachen oder Sprecher, die kontrafaktische Aussagen ermöglichen. Die Äußerung „Schnee ist weiß“ (in Deutsch) könnte nach diesem Verständnis auch dann noch wahr sein, wenn das Gras nicht grün wäre, weshalb (W₂) wohl kein Gesetz darstellte. Ein erfolgreicher Interpret müsste dementsprechend mehr wissen, als die wahren W-Sätze zu kennen bzw. ableiten zu können. Er müsste auch wissen, dass es sich um Gesetze handelt. Dieser Vorschlag erscheint allerdings vor dem Hintergrund des *konventionellen*³⁰⁰ Charakters der Bedeutung von sprachlichen Ausdrücken nicht besonders einleuchtend. Ist es denn ein Naturgesetz, dass ein Satz (ein Satz im Munde eines Sprechers) bedeutet, was er bedeutet? Wir gehen doch wohl eher davon aus, dass derselbe *Ausdruckstyp*, der *faktisch* im Munde eines Sprechers (zu einem bestimmten Zeitpunkt an einem bestimmten Ort bzw. in einer bestimmten Sprachgemeinschaft) eine wahre Aussage zum Ausdruck bringt, auch eine falsche zum Ausdruck bringen *könnte* – nämlich dann, wenn derselbe Ausdruck eine andere Bedeutung hätte.

Davidson ist sich durchaus darüber im Klaren, dass der Versuch, alle sprachlichen Ausdrucksmöglichkeiten auch tatsächlich wahrheitsfunktional zu analysieren zumindest als ein unabgeschlossenes Projekt anzusehen ist. Bekannte Problemfälle

³⁰⁰ Mit ‚konventionell‘ ist hier gemeint, dass die Verbindung zwischen einem Ausdruck und seiner Bedeutung in keinem Sinne eine notwendige ist, auch nicht im Sinne einer natürlichen Notwendigkeit.

einer konsequent wahrheitsfunktionalen Analyse bilden etwa Zitate, Überzeugungs- und Handlungszuschreibungen, metaphorische Redeweisen oder sprachliche Modi (durch die die illokutive Kraft von Äußerungen bestimmt wird). Davidson will die wahrheitssemantisch analysierbaren Aspekte von Bedeutung als ‚wörtliche‘ von den pragmatischen Aspekten trennen und er hat versucht, für die genannten Schwierigkeiten Lösungsmöglichkeiten im Hinblick auf die ‚wörtliche Bedeutung‘ aufzuzeigen. Zu diesen – zum Teil eher technischen – Fragen möchte ich hier nicht Stellung beziehen. Ich werde allerdings später die grundsätzliche Frage aufwerfen, ob Wahrheit überhaupt der angemessene *Grundbegriff* für eine Theorie der Bedeutung und/oder propositionaler Einstellungen sein kann. Die unbestrittenen Schwierigkeiten wahrheitsfunktionaler Analysen könnten dann immerhin ein zusätzliches Indiz dafür liefern, dass dem nicht so ist.

Von entscheidender Wichtigkeit für die Beurteilung von Davidsons Ansatz ist auch die Frage, in welcher Weise sich eine ‚Interpretationstheorie‘ durch Belege überprüfen lassen könnte, die schon vor Beginn der ‚Interpretation‘ zur Verfügung stehen. Detaillierte Überzeugungen, Wünsche und Intentionen von Sprechern können ja in Situationen Radikaler Interpretation – wie Davidson selbst fordert – nicht herangezogen werden, bevor nicht einiges darüber bekannt ist, was deren Äußerungen bedeuten. Letzteres soll nach Davidson aber wiederum davon abhängen, was die Sprecher glauben und wünschen (bzw. von ihren doxastischen und volitiven Einstellungen). Seiner Auffassung nach ist der einzige Ausgangspunkt, der weder das eine, noch das andere voraussetzt, die *Einstellung des Für-wahr-Haltens*, welche sich in Zustimmungsverhalten ausdrückt. Das heißt Davidson geht davon aus, dass Zustimmung vor jeder ‚Interpretation‘ erkennbar ist³⁰¹. Ein in Situationen Radikaler Interpretation zulässiger Beleg für einen W-Satz einer Theorie *T* für einen Sprecher *X* (bzw. eine Gruppen von Sprechern *G*) hätte demnach etwa die Form:

(B) Sprecher *X* äußert zustimmend den Satz „*S*“ (zum Zeitpunkt *t*) unter den Bedingungen *C*.

Dem möglichen Einwand gegen Belege der Form (B), dass Sprecher sich doch irren könnten, begegnet Davidson, indem er zugibt, dass solche Belege für W-Sätze zwar für sich genommen nicht stichhaltig sind, in ihrer Summe der Theoriebildung aber insofern

³⁰¹ Dem Einwand, Für-wahr-Halten oder Zustimmung setze doch propositionale Gehalte in dem Sinne voraus, dass es sich auf etwas bezieht, das für wahr gehalten oder dem zugestimmt wird, begegnet Davidson in ‚The Structure and Content of Truth‘ (1990) mit dem Hinweis, dass dasjenige, dessen Voraussetzung es zu vermeiden galt, individuierte propositionale Einstellungen seien, nicht aber die Einstellung des Für-wahr-Haltens als solche.

Beschränkungen auferlegen, als wir als Interpreten nicht umhin könnten dem Sprecher möglichst viele wahre Überzeugungen zuzuschreiben. Wir könnten Sprecher schlicht nicht verstehen, wären wir nicht in der Lage eine ‚Interpretationstheorie‘ für sie aufzustellen, die eine weitgehende Übereinstimmung zwischen unseren – der ‚Interpreten‘ – und ihren Überzeugungen herstellte:

„Um aus den Äußerungen und dem Verhalten – selbst dem überspanntesten Verhalten – anderer schlau zu werden, müssen wir auf ihrer Seite eine Menge Vernünftiges und Wahres ausfindig machen. Wenn wir zu viel Unvernunft auf Seiten der anderen sehen, untergraben wir einfach unsere Fähigkeit zu verstehen, was es denn eigentlich ist, bezüglich dessen sie so unvernünftig sind. Wenn die rauhen Mengen an Übereinstimmung im Hinblick auf hausbackene Angelegenheiten der Aufmerksamkeit entgehen, liegt das daran, dass die gemeinsamen Wahrheiten zu zahlreich und zu fade sind, um sie zu erwähnen.“
(Davidson 1974, S. 221)

Bereits in dieser Formulierung Davidsons wird deutlich, dass das ‚Prinzip der Nachsichtigkeit‘ nicht lediglich als eine – gewissermaßen ‚nützliche‘ – *Maxime* im Sinne methodologischer Optimierungsannahmen, sondern eigentlich als eine notwendige Präsupposition im Sinne einer transzendentalen oder konstitutiven Unterstellung fungiert. Und als solche erscheint es tatsächlich zwingend: Es bedarf eines weiten Netzes an geteilten Überzeugungen, um überhaupt verstehen oder Irrtümer feststellen zu können. Wir können die Behauptung irgendeines Hinterweltlers, die Erde sei eine Scheibe, nur deshalb als belächelnswerten Standpunkt abtun, weil wir eine ganze Menge an sonstigen Überzeugungen – insbesondere über die Erde und über Scheiben – *teilen*. Allerdings ist diese wichtige Einsicht nicht an Davidsons Verständnis der Bedingungen von Verstehen und Verständigung bzw. daran gebunden, die Situationen Radikaler Interpretation als für normale Formen sprachlicher Kommunikation konzeptuell grundlegend anzusehen, und sie ist auch nicht erst in diesem Kontext formuliert worden³⁰².

Davidson geht schließlich so weit, zu behaupten, wir hätten gar keinen Grund, andere Wesen überhaupt als rational bzw. als Personen anzusehen, wenn eine ‚Interpretation‘ nicht möglich ist. Und ich denke, dass wir ihm hier grundsätzlich folgen sollten, wenngleich der grundlegende Gedanke nach meinem Dafürhalten in einer Weise formuliert werden muss (und kann), die von Davidsons speziellen

³⁰² Besonders auffällig ist die Ähnlichkeit von Davidsons *Maxime* der ‚Optimierung der Übereinstimmung‘ mit Gadamers ‚Vorgriff auf Vollkommenheit‘, als formaler Voraussetzung hermeneutischen Verstehens, gemäß der „nicht nur (...) ein Text seine Meinung vollkommen aussprechen soll, sondern auch (...) das, was er sagt, die vollkommene Wahrheit ist.“ (Gadamer (1960), S. 299).

Annahmen unabhängig ist: Würden wir einem möglichen Gegenstand des Verstehens nicht von vorneherein ein gewisses Mindestmaß an Rationalität oder Vernunft und Wahrheit (bzw. geteilte Überzeugungen) unterstellen, könnten wir erst gar nicht davon ausgehen, etwas vor uns zu haben, dem wir Geist oder geistigen Ursprung zuschreiben können.

Davidson hat erst nach und nach genauer erläutert, was er mit „rational“ oder „vernünftig“ meint. Zunächst muss dies natürlich heißen, dass wir die Zuschreibung logischer Widersprüche möglichst vermeiden. Aber auch zu viele empirische Irrtümer seitens des Sprechers sind auszuschließen. Darüber hinaus müssen Zuschreibung eines ‚intentionalen Profils‘ aber auch ein mehr oder minder kohärentes, stimmiges Gesamtbild hinsichtlich seiner Überzeugungen, seiner ‚Wünsche‘ und, wie erst in späteren Aufsätzen deutlich wird, seiner Absichten und Handlungen ergeben. Dazu müssen wir bei unserem zu ‚interpretierenden‘ Gegenüber weitgehend – am Anfang der ‚Interpretation‘ sogar praktisch vollständig – nur solche (‚evidentiellen‘) Begründungszusammenhänge zwischen Überzeugungen als für-richtig-gehalten unterstellen, die wir selbst (als Interpreten) für richtig bzw. gültig halten. In *Toward a Unified Theorie of Meaning and Action*‘ (1980) rückt Davidson dann aber auch die Handlungsrationalität ins Zentrum notwendiger Rationalitätsunterstellungen. Konsequenterweise sieht er eine ‚Interpretationstheorie‘ nun als integralen Bestandteil einer umfassenden ‚interpretativen Handlungstheorie‘, verstanden als Theorie der rationalen Erklärung sprachlichen und nichtsprachlichen Handelns auf der Grundlage individueller Präferenzen und Überzeugungen:

„Theory of meaning as I see it, and Bayesian decision theory, are evidently made for each other. Decision theory must be freed from the assumption of an independently determined knowledge of meaning; theory of meaning calls for a theory of degree of belief in order to make serious use of relations of evidential support. But stating these mutual dependencies is not enough, for neither theory can be developed first as a basis for the other. There is no way simply to add one to the other, since in order to get started each requires an element drawn from the other. What is wanted is a unified theory that yields degree of belief, utilities on an interval scale, and an interpretation of speech without assuming any of them.“
(Davidson 1981, S. 8)

Davidson glaubt, auf der schmalen, vorthoretisch feststellbaren Grundlage der Zustimmung zu Sätzen (bzw. der Präferenz bezüglich der Wahrheit eines Satzes gegenüber einem anderen oder dem Wunsch, ein Satz möge wahr sein) eine einheitliche Theorie des Handelns und der Bedeutung aufstellen zu können, die nicht nur Auskunft darüber liefert, was die Sätze (für einen Sprecher) bedeuten, sondern auch über die Grade an Wahrscheinlichkeit, die er der Wahrheit der Sätze zubilligt, sowie

den Graden von Präferenz, die er für bestimmte Zustände der Welt hat. Er beruft sich dabei unter anderem auf entscheidungstheoretische Überlegungen R. Jeffrey's³⁰³, nach denen sich die subjektive Wahrscheinlichkeit und die relative Wünschbarkeit von ‚Propositionen‘ (bzw. Sätzen) herausbekommen lässt, wenn wir die Präferenzen einer Person beobachten können und erst mal die logischen Operatoren identifiziert sind.

Nun betont Davidson allerdings auch immer wieder, dass die Zustimmung zu Sätzen nicht nur von den (Graden von) Überzeugungen (und subjektiven Präferenzen) der Sprecher abhängt, sondern auch davon, was die Sätze in deren Mund bedeuten. Dies erscheint allerdings insofern merkwürdig, als er andererseits ein bedeutungsintentionalistisches Modell von Äußerungsbedeutungen vertritt, das stark an Grice erinnert:

„What matters to successful linguistic communication is the intention of the speaker to be interpreted in a certain way, on the one hand, and the actual interpretation of the speaker's words along the intended lines through the interpreter's recognition of the speaker's intentions, on the other.“ (Davidson 1990, S. 311)

Wie, so möchte man fragen, kann hier noch Platz sein für einen Begriff wörtlicher Äußerungsbedeutung, die unabhängig von dem ist, was ein Sprecher glaubt oder beabsichtigt? Denn wir sollen Äußerungen ja, wie Davidson betont, so interpretieren, wie der Sprecher interpretiert zu werden beabsichtigt. Und die Absichten eines Sprechers verstehen wir ja nur, wenn wir seine Überzeugungen und Wünsche kennen.

Diese Frage erscheint umso dringlicher, als sich Davidson im Laufe der Zeit zunehmend einen ideolektischen Sprachbegriff zu Eigen gemacht hat. Wie er inzwischen glaubt, hat das Konzept einer *gemeinsamen* Sprache in einem bestimmten Sinne eigentlich keine empirische Grundlage³⁰⁴. Und dass Leute dazu neigen, so ähnlich zu sprechen, wie ihre Nachbarn, sei für die Möglichkeit von Verstehen *konzeptuell irrelevant*. Sprachliche Konventionen seien erst auf der Grundlage von erfolgreicher Verständigung möglich und nicht umgekehrt. Eine solche Sicht der Dinge ist im Rahmen der von Davidson unterstellten Vorgaben einer Verstehenstheorie aus der Perspektive eines Beobachters sprachlichen Geschehens tatsächlich konsequent. Denn aus der Perspektive einer Beschreibung von Ereignissen hat man es schließlich nur mit *Regularitäten* im Sprachverhalten zu tun. Und da dürfte es kaum zwei Sprecher geben, deren *faktisches* sprachliches Verhalten sich einem gemeinsamen Set an (grammatischen oder semantischen) Regeln unterordnen lassen.

³⁰³ R. Jeffrey: *The Logic of Decision* (1983); hierzu auch: Davidson (1990).

³⁰⁴ Siehe Davidson: ‚Kommunikation und Konvention‘ (1982) und insbes. ders.: ‚Eine hübsche Unordnung von Epitaphen‘ (1986a).

Dies wird noch genauer zu erörtern sein. Auf jeden Fall aber erscheint es unter diesen Annahmen umso berechtigter, zu fragen, wie überhaupt noch systematisch zwischen den Überzeugungen und der Bedeutung der Äußerungen eines Sprechers unterschieden werden kann. Wenn nämlich die Bedeutung sprachlicher Ausdrücke in systematischer Weise an Überzeugungen und Mitteilungsabsichten von Sprechern gekoppelt ist und zugleich ein die Möglichkeit *subjekttranszendenter* Bedeutung eröffnendes Konzept einer *geteilten* Sprache (wie auch immer dies genauer zu spezifizieren ist) nicht zur Verfügung steht, könnte ein überzeugungsunabhängiger Begriff von Bedeutung doch offenbar nur dann gewonnen werden, wenn Sprechern die Möglichkeit einer Täuschung im Sinne einer Abweichung ihrer Mitteilungsabsichten von ihren übrigen propositionalen Einstellungen (also wenn etwa jemand selbst nicht glaubt, was er sagt) nicht nur in Einzelfällen, sondern in systematischer Form eingeräumt würde. Doch dann hätte ein Radikaler Interpret wiederum keine Chance, an die propositionalen Einstellungen der Sprecher heranzukommen. Denn diesen muss ja nicht nur weitgehende *Rationalität*, sondern auch weitgehende *Aufrichtigkeit* unterstellt werden, um überhaupt eine ‚Interpretationstheorie‘ in Angriff nehmen zu können. Bevor wir uns eingehender mit den sprachphilosophischen Thesen Davidsons beschäftigen, will ich aber zunächst auf Davidsons und Dennetts Kritik am Reduktiven Materialismus bzw. am ‚harten Realismus‘ (Dennett) bezüglich des Mentalen eingehen.

Dennetts Theorie der Perspektivität des ‚Intentionalen‘

Ein wichtiger Unterschied zwischen Dennetts Theorie ‚Intentionaler Systeme‘ und Davidsons ‚Interpretationstheorie‘ besteht sicherlich in dem erheblich geringeren Stellenwert, den Dennett kommunikationstheoretischen Fragen im engeren, sprachphilosophischen Sinne beimisst. Dies zeigt ist schon darin, dass Davidson die Reichweite seiner Theorie ausdrücklich auf sprachliche Wesen einschränkt, während es Dennett um alle biologischen Organismen oder artifiziellen Systeme geht, bei denen eine Erklärung oder Vorhersage aus der ‚*Intentionalen Perspektive*‘ oder Einstellung (‚*Intentional Stance*‘) sinnvoll oder sogar notwendig wäre. Die Intentionale Perspektive unterscheidet Dennett, als eine von drei möglichen explanatorischen Perspektiven, von der physikalistischen und der funktionalen oder Gestaltungs- Perspektive (‚*Design Stance*‘)³⁰⁵.

³⁰⁵ D. Dennett: ‚Intentional Systems‘ (1971) und ‚True Believers‘ (1981a).

Aus der *physikalistischen* Perspektive werden Dinge oder Systeme im Sinne ihrer physischen Konstitution beschrieben oder erklärt. Sie entspricht den üblichen Erklärungsmethoden der Physik oder Chemie. Die *funktionale* Perspektive entspricht dem, was wir im Zusammenhang mit funktionalen Theorien und dem Computermodell des Geistes schon erörtert hatten. Sie beschreibt die interne Organisation eines Systems in einer Weise, bei der von der jeweiligen physischen Realisierung des Systems abstrahiert wird, um zu Vorhersagen des Systemverhaltens durch kausale Input-output-Generalisierungen zu gelangen, wenn die physischen Fakten (im engeren Sinne) unbekannt oder irrelevant sind. Sie kommt etwa bei der Erstellung von Computerprogrammen oder beim alltäglichen Verständnis von technischen Geräten, wie etwa einer Waschmaschine zum Zug. Die *Intentionale* Perspektive sei schließlich diejenige, aus der ein System durch die Zuschreibung von Überzeugungen und Wünschen, unter Berücksichtigung bestimmter Rationalitätsregeln, erklärt bzw. im Verhalten vorhergesagt wird. Sie abstrahiere nicht nur von der physischen Konstitution, sondern auch von der funktionalen Organisation eines Systems wenn diese entweder unbekannt oder irrelevant ist. Es ist die Intentionale Perspektive, die wir nach Dennett sowohl in der vielbemühten ‚Alltagspsychologie‘, als auch – in präziserer Form – in der wissenschaftlich-kognitiven Psychologie einnehmen.

Diese drei Perspektiven sind nach Dennett epistemisch und explanatorisch autonom, sie können jeweils ihren explanatorischen Zweck erfüllen, ohne dass auf ein nur aus den anderen Perspektiven zugängliches Wissen zurückgegriffen werden muss. Eine markante Besonderheit von Dennetts Position ist nun die These, dass die in der funktionalen bzw. Intentionalen Perspektive beschriebenen ‚Gegenstände‘ gegenüber denen der physikalistischen Perspektive schrittweise an ‚ontologischer Dignität‘ einbüßen. Vor allem sollten wir nach Dennett nicht glauben, dass mit den in Intentionaler Perspektive zugeschriebenen propositionalen Einstellungen *wirkliche* physische *Zustände* oder *Vorgänge*, etwa im Gehirn, *individuiert* würden. Vielmehr sollten Zuschreibungen propositionaler Einstellungen als ‚heuristischer Überzug‘, als idealisierende Annäherung an tatsächlich wirkende Mechanismen verstanden werden³⁰⁶. Dennoch könnten solche Zuschreibungen im Rahmen der ‚Alltagspsychologie‘ – hier ist sich Dennett mit dem ‚Realisten‘ Fodor einig – ein erstaunlich treffsicheres Instrument für die Vorhersage von Verhalten liefern, welches

³⁰⁶ Das kann man sich ungefähr so vorstellen wie die Diskrepanz zwischen einer arithmetischen ‚Überzeugung‘, wie: 20 geteilt durch 3, multipliziert mit 3, ist 20, und deren physischer Realisierung durch einen einfachen Taschenrechner: Dem Taschenrechner eine entsprechende ‚Überzeugung‘ zuzuschreiben, würde, trotz Unkenntnis über dessen Funktionsweise und inneren Aufbau, dennoch eine *ziemlich präzise* Vorhersage über die nach entsprechendem Tastendruck zu erwartende Reaktion auf dem Display ergeben, auch wenn die *faktisch wirkenden Mechanismen* das Erscheinen eines Tokens „19.999999“ verursachen werden.

es ermöglicht, aus Gründen ‚kognitiver Ökonomie‘ auf die Details einer Beschreibung der tatsächlich kausalwirksamen Mechanismen zu verzichten. Funktionale oder intentionale Zustände seien dabei aber insoweit für voll zu nehmen, als ihre Zuschreibung für die Erklärung und Vorhersage von Verhalten *nützlich* sein kann:

„*All there is to being a true believer is being a system whose behaviour is reliably predictable via the intentional strategy, and hence all there is to really and truly believing that p (for any proposition p) is being an intentional system for which p occurs as a belief in the best and (most predictive) interpretation.*“ (Dennett 1981a, S. 29)

Die Zurückweisung starker ‚realistischer‘ und reduktiver Ambitionen, insbesondere auch der Idee einer ‚Sprache des Geistes‘, ist aber nach Dennetts Einschätzung nicht gleichbedeutend mit der Leugnung der Möglichkeit von Objektivität in Bezug auf die Zuschreibbarkeit propositionaler Einstellungen oder einem ‚Anti-Realismus‘ gegenüber dem Phänomen ‚Intentionalität‘ ganz allgemein. Diese Konsequenz scheint sich nämlich aus Dennetts Behauptung zu ergeben, dass ein Intentionales System zu sein darauf hinauslaufe, Vorhersagen über dessen Verhalten mangels detaillierter Kenntnisse des inneren Aufbaus am besten durch die Zuschreibung propositionaler Einstellungen zu erzielen. Und dies scheint dazu zu führen, dass Geist oder Geistigkeit ein Phänomen ist, welches relativ zu den kognitiven Fähigkeiten des Interpretieren ist. Wenn wir uns etwa Marsianer vorstellen, welche, mit Fähigkeiten ‚Laplacescher Superphysiker‘ ausgestattet, die Erde besuchten, so wären diese nach Dennetts Vorgaben ja gegebenenfalls in der Lage unser Verhalten auf der Ebene der Inanspruchnahme mikrophysikalischer Gesetzmäßigkeiten, also ohne Inanspruchnahme propositionaler Einstellungen, extrem präzise vorherzusagen. Aus der Sicht der Marsianer wären wir dementsprechend keine Intentionalen Systeme. Dennett reagiert auf dieses Problem einerseits damit, dass sich die Marsianer dann immerhin noch *selbst* als Intentionale Systeme verstehen müssten, als Systeme, die beobachten, theoretisieren, vorhersagen und kommunizieren (Dennett 1981a). Allerdings wird durch dieses Argument – mal ganz abgesehen davon, dass die Vorhersagen der Marsianer sich auf physikalistische Vorhersagen von Körperbewegungen beschränken müssten – die Schwierigkeit nur verschoben, denn erstens könnten die Marsianer wiederum von kognitiv noch fortgeschritteneren Alpha-Centauren als physikalische oder funktionale Systeme verstanden und ihr ‚Verhalten‘ dementsprechend unter Verzicht auf ein ‚intentionales‘ Vokabular vorhergesagt werden. Außerdem könnten wir oder die Marsianer prinzipiell selbst auf eine kognitive Entwicklungsstufe gelangen, die dies ermöglicht. Überzeugender wirkt deshalb ein anderes Argument

Dennetts, nach dem den marsianischen ‚Super-Laplacianern‘ beim Verzicht auf die Zuschreibung propositionaler Einstellungen ein *objektiv existierendes Muster* entginge:

„Our imagined Martians might be able to predict the future of the human race by Laplacean methods, but if they did not also see us as intentional systems, they would be missing something perfectly objective: the *patterns* in human behaviour that are describable from the intentional stance, and only from that stance, and that support generalizations and predictions.“ (Dennett 1981a, S.25)

Dennett vergleicht Zuschreibungen propositionaler Einstellungen mit dem Postulieren abstrakter Objekte in der Physik, wie etwa Gravitationszentren. Hier wie dort seien die postulierten ‚Gegenstände‘ zwar nichts, was irgendwie unabhängig von unseren explanativen Interessen (bzw. den entsprechenden Perspektiven) materiell realisiert sei, deshalb aber nicht schon als bloße Fiktionen zu verstehen. Propositionale Einstellungen seien Postulate einer ‚Alltagspsychologie‘, die ihrerseits „might best be viewed as a rationalistic calculus of interpretation und prediction – an idealizing, abstract, instrumentalistic interpretation method, that has evolved, because it works and it works, because we have evolved“³⁰⁷. Zwischen Fällen, bei denen eine ‚intentionale‘ Beschreibung zwar durchaus üblich, aber durch eine physische ersetzbar sei (wie etwa bei Thermostaten) und solchen, bei denen dies im Rahmen unserer jetzigen Möglichkeiten offensichtlich nicht so ist (wie in Bezug auf uns selbst), lässt sich nach Dennett aber keine als intrinsisch zu deutende Grenze ziehen. Ontologische Fragen erlaubten häufig keine einfachen Ja-nein-Antworten, sondern müssten differenziert beurteilt werden³⁰⁸.

Aus dem letzten Zitat wird überdies deutlich, dass Dennetts Begründung für die Nicht-Relativierbarkeit bzw. Nichtreduzierbarkeit der Intentionalen Perspektive sowohl Elemente von Überlegungen Davidsons, als auch Millikans³⁰⁹ enthält. Nach Dennett orientierten wir uns in der Intentionalen Perspektive grundsätzlich an Prinzipien, wie den folgenden:

– Das Verhalten eines Systems in Intentionaler Perspektive ergibt sich aus dem, was für es *rational* ist, gegeben seine Überzeugungen und Wünsche, bzw. Bedürfnisse.

³⁰⁷ Dennett (1981), S. 48 f.

³⁰⁸ Dennett (1991).

³⁰⁹ In seinem standardmäßigen Verständnis des praktischen Syllogismus der ‚Alltagspsychologie‘ als einem Instrument der *Verhaltensvorhersage* fällt Dennett aber eindeutig hinter die oben angesprochene Position Millikans zurück.

– Überzeugungen sind als solche zuzuschreiben, die ein System haben *sollte*, gegeben seine perzeptuellen Kapazitäten, seine epistemischen Notwendigkeiten und seine Biographie.

– Wünsche bzw. Bedürfnisse eines Systems sind als solche zu verstehen, die es haben sollte, gegeben seine biologischen Bedürfnisse und die praktikabelste Weise, diese zu befriedigen.

Auch Dennett befürwortet also jene interpretativen Maximen, nach denen die Überzeugungen, die ein System (ein Organismus) hat *generell wahr* und *kohärent* sein müssen. Darüber hinaus formuliert er aber auch noch biologische und anthropologische – gewissermaßen ‚basishermeneutisch‘ – begründete Beschränkung, die bei Davidson bestenfalls implizit mitklingen: So müssten die zugeschriebenen Überzeugungen eines Organismus auch eine *elementare Relevanz für dessen Leben*, für seine Bedürfnisse und Überlebensfähigkeit haben. *Falsche* Überzeugungen zuzuschreiben sei dann nur auf der Grundlage einer speziellen Geschichte darüber möglich, wie der Irrtum auf Grund bestimmter Umweltbedingungen relativ zu den kognitiven Fähigkeiten des Organismus zu Stande kommen konnte. Analoges gilt natürlich erst recht für Wünsche: Auch diese seien vor dem Hintergrund biologischer Konstanten, wie Überlebens- und Fortpflanzungsdrang, also dem Bedürfnis nach Nahrung, Sicherheit, Gesundheit, Sex, Wohlergehen, Macht, Einfluss u. ä. m. zuzuordnen. Und auch im Falle ‚abnormaler Wünsche‘ müsste eine besondere Geschichte erzählt werden können.

Dennetts Begründung für die Behauptung, dass wir einem Organismus generell wahre und kohärente Überzeugungen zuschreiben müssen, ähnelt nun einerseits der Davidsons, besteht also in einem Hinweis auf die methodologische Notwendigkeit und Konstitutivität entsprechender Interpretationsprinzipien: „(...) the attribution of false belief, *any* false belief, requires a special genealogy, which will be seen to consist in the main in true beliefs.“ (Dennett 1981a, S. 18). Irrationalität sei nur vor dem Hintergrund einer die generelle Rationalität Intentionaler Systeme unterstellenden, holistischen Interpretationstheorie identifizierbar und erklärbar³¹⁰. Andererseits will er

³¹⁰ So wehrt sich Dennett in ‚Making Sense of Ourselves‘ (ders.:(1981b)) gegen den Vorwurf Stichts (in: ders.: ‚Dennett on Intentional Systems‘, (1981)), die idealisierenden Interpretationsmaximen der ‚Alltagspsychologie‘ würden den Tatsachen im Hinblick auf unsere kognitiven Beschränkungen nicht gerecht, mit dem Argument, dass zugeschriebene Irrtümer als *Fehler* im Rahmen einer ‚guten‘, also rationalen Prozedur zu verstehen sein müssten, nicht als *Befolgung* einer ‚schlechten‘, also ‚irrationalen‘ Prozedur. Generelle Rationalität ist auch für Dennett *konstitutiv* für die Möglichkeit des Verstehens – im weitesten Sinne – geistiger Phänomene: „This is not to say that we are always rational, but that when we are not, the cases defy descriptions in ordinary terms of belief and desire. There is no mystery about why this should be so. An intentional interpretation of an agent is an exercise that

dies aber zugleich auch mit *empirischen* Überlegungen untermauern, nämlich mit Annahmen im Rahmen der biologischen Evolutionstheorie:

„There is no point in ascribing beliefs to a system unless ascribed are in general appropriate to the environment, and the system responds appropriately to the beliefs. An eccentric expression of this would be: The capacity to believe would have no survival value unless it were the capacity to believe truth.“ (Dennett 1971, S. 17)

Später präsentiert Dennett diese Überlegung allerdings etwas differenzierter:

„treating each other as intentional systems works (to the extent that it does) because we really are well designed by evolution and hence we approximate to the ideal version of ourselves exploited to yield the predictions. But not only does evolution not guarantee that we will always do what is rational; it guarantees that we won't. If we are designed by evolution, then we are almost certainly nothing more than a bag of tricks, (...) and no better than our ancestors had to be to get by.“ (Dennett 1981, S. 51)

Dennetts evolutionstheoretische Argumentation entspricht grundsätzlich dem, was Millikan zugunsten einer Form von Korrespondenztheorie der Wahrheit anführt. Dessen kritische Bewertung soll hier nicht wiederholt werden. Das zweite Zitat zeigt aber auch, dass Dennett zwischen einer ‚Alltagspsychologie‘ und einer wissenschaftlichen Psychologie unterscheiden will. Die wissenschaftliche Psychologie in Form einer Theorie Intentionaler Systeme soll auch empirische Untersuchungen über Inkonsistenzen, Gedächtnisschwächen, ‚pragmatische Abkürzungen‘ und andere (vermeintliche oder tatsächliche) kognitive Defizite unserer ‚alltagspsychologischen‘ Denkweisen berücksichtigen. Schon ‚Alltagspsychologie‘ sei eine Mischung aus empirischem Verallgemeinerungswissen über die ‚intentionalen Gewohnheiten‘ unseres gesellschaftlichen Umfeldes und einer idealisierten Theorie Intentionaler Systeme. Auch eine wissenschaftliche Theorie Intentionaler Systeme müsse dementsprechend auf zwei unterschiedliche, sich jedoch wechselseitig beeinflussende Weisen betrieben werden: Einmal in Form einer *Kompetenztheorie*, das heißt, als abstrakte, idealisierende, holistische und instrumentalistische Theorie, die sich mit bereits existierenden ‚normativen‘ Disziplinen, wie Spieltheorie und Entscheidungstheorie überschneidet. Zum anderen als eine subpersonale kognitive Psychologie im Sinne einer *Performanztheorie*, die als konkrete mikrotheoretische Wissenschaft die tatsächlichen Realisierungen Intentionaler Systeme untersucht.

attempts to make sense of an agent's acts, and when acts occur that make no sense, they cannot be straightforwardly interpreted in sense making terms.“ (ebd. S. 87).

Letztere würde etwa im Sinne des Computermodell des Geistes zu zeigen versuchen, wie das Gehirn, verstanden als Syntaktische Maschine, in der Lage sein kann ‚so zu tun‘ als wäre es eine Semantische Maschine. Allerdings müsse die Performanztheorie den konzeptuellen Anschluss an die Kompetenztheorie waren, da sie sonst überhaupt keine Psychologie mehr sei, sondern bestenfalls abstrakte Neurophysiologie bzw. „pure internal syntax with no hope of semantic interpretation.“ (ebd., S. 64).

Was aber heißt für Dennett „semantische Interpretation“? Ein wesentlicher Unterschied zu Davidson besteht darin, dass Dennett – neben seiner stärkeren Ausrichtung an den technischen Möglichkeiten der Simulation von ‚Intentionalität‘ im Rahmen der KI-Forschung – ein Verfahren der semantischen Interpretation von Intentionalen Systemen vorschlägt, welches ohne die gleichzeitige Interpretation von sprachlichen Äußerungen auskommen soll. Insbesondere die Zuschreibung propositionaler Einstellungen durch eine wissenschaftliche Psychologie soll als Methode der Zuschreibung ‚begrifflicher Welten‘ („notional worlds“) verstanden werden³¹¹. Ihr Vorgehen soll dabei einem ‚methodischen Solipsismus‘ bzw. einer ‚Phänomenologie aus der Perspektive der dritten Person‘ insofern entsprechen, als propositionale Einstellungen so beschrieben werden müssen, dass ihr Gehalt dadurch bestimmt ist, wie den Subjekten die Welt *erscheint*, während die Frage nach der tatsächlich existierenden Welt im Stile von Husserls *Epoché* quasi ‚eingeklammert‘ wird. Dies bedeutet, dass eine ‚Psychologie begrifflicher Welten‘ die ‚intentionalen Zustände‘ und ‚Ereignisse‘ eines Systems zwar im Kontext seiner Umwelt beschreiben, sie aber nicht über ihre Wahrheitsbedingungen, sondern – als ‚Sinn‘ – relativ auf die kognitiven Fähigkeiten und die Perspektive des Systems und vor dem Hintergrund seiner (mutmaßlichen) Bedürfnisse bestimmen müsste. Durch diese Methode würde einerseits ein ‚internalistischer Syntaktizismus‘ im Stile Stichs³¹² vermieden, der zwar die Frage der materialen Realisierung und Kausalität von Geist beantworten zu können scheint, nicht jedoch die nach der semantischen Interpretation. Zugleich würde jedoch auch keinem Externalismus im Sinne Putnams oder Burges das Wort geredet, der zwar eine semantische Interpretation liefern, nicht jedoch die Kausalwirksamkeit des Geistes erklären könne. Dennett gibt den Externalisten zu, dass die internen ‚begrifflichen Zustände‘ hinsichtlich des Bezugs unterbestimmt sind. Dies ist für ihn jedoch kein Mangel, sondern ein Gewinn, da es schließlich darauf ankomme, das Verhalten des Systems erklären bzw. prognostizieren zu können und dafür seien Tatsachen, die für das Verhalten eines Systems keine Rolle spielen können als irrelevant anzusehen.

³¹¹ S. Dennett: ‚Beyond Belief‘ (1982), bes. S. 151 f.

³¹² Stich (1983).

Dennett entwirft in ‚Beyond Belief‘ einen beispielhaften Fall, um die Möglichkeit einer gleichsam ‚radikalen Konstruktion‘ einer ‚begrifflichen Welt‘ für ein intentionales System zu veranschaulichen: Stellen wir uns vor, wir wären durch irgendwelche Umstände mit einem tiefgefrorenen Organismus konfrontiert, von dem wir nicht wüssten, in welcher Umwelt er normalerweise tatsächlich lebt, dessen *interne physische* Struktur und Zusammensetzung uns aber vollständig und in allen Hinsichten bekannt ist. Nach Dennett sollte uns dies in die Lage versetzen zu bestimmen, wie dieser Organismus unter bestimmten Umwelteinflüssen reagieren würde. Die Beschreibung des Organismus und seiner Funktionen müsste sich dabei sukzessive ‚intentional‘ anreichern lassen, in dem wir, aufgrund unserer Kenntnisse der internen Physis, zu begründeten Aussagen über die Art der Sinnesorgane, der Verhaltensdispositionen und der systematischen Bedürfnisse gelangen, was uns wiederum über eine Form von Umwelt spekulieren ließe, an die dieser Organismus idealiter als angepasst erscheint.

Zwar könnten nach Dennett auf einer solchermaßen beschränkten epistemischen Ausgangsbasis, also ohne Kenntnisse etwa historischer Fakten, die Funktionen mancher Elemente für den Interpreten auf immer unterbestimmt bleiben. Ohnehin aber determiniere die subjektive begriffliche Perspektive eines Lebewesens grundsätzlich nicht eine einzig mögliche Welt: „For instance, variations in a world entirely beyond the ken, or interests, of a person would generate different possible worlds equally consistent with the maximal determination provided by the constitution of the person“ (1982, S. 158). Je größer jedoch die Komplexität eines Organismus, je ausgebildeter die Fähigkeit zu Lernen und zu Erinnern, desto spezifischer ließe sich dessen hypothetische ideale Umwelt beschreiben; – ja, Dennett glaubt sogar, dass sich im Fall von hochgradig angepassten Wesen, wie uns Menschen auf einer solchen Basis fiktionale Umweltspezifizierungen herausbekommen lassen, die der Komplexität unserer alltagssprachlichen Unterscheidungsmöglichkeiten entsprechen (wenngleich sich dabei nicht zwischen ‚Erde‘ und ‚Zwillings-Erde‘ im Sinne Putnams unterscheiden lasse).

Dennetts Konzeption begrifflicher Welten stellt eine ‚internalistische‘ Reaktion auf die oben diskutierten Einwände gegen ‚enge‘ Verständnisse propositionalen Gehaltes dar. Anders als etwa Fodor oder Dretske geht Dennett nicht den Weg einer naturalistisch gewendeten Externalisierung, sondern erklärt jegliche am *tatsächlichen Bezug* orientierte Frage nach dem richtigen Verständnis propositionalen Gehaltes als für die ‚Psychologie‘ intentionaler Systeme schlichtweg *irrelevant*. Es ist, wie ich meine, jedoch ziemlich unplausibel anzunehmen, dass auch kulturell bedingte Begriffe, wie Städtenamen (Dennetts Beispiel), Charaktereigenschaften, politische Ideen u. ä. m.

überhaupt unter den von Dennett genannten Beschränkungen zugeschrieben werden könnten. Dennett bleibt uns hier jedenfalls plausibilisierende Vorschläge der Anwendung seiner Strategie auf solche Fälle schuldig. In dieser Hinsicht zumindest ist Davidsons ‚Theorie der Interpretation‘ bei weitem nicht so ambitioniert. Insbesondere glaubt Davidson nicht, komplexe und hochgradig spezifizierte Überzeugungen ließen sich überhaupt ohne eine parallel laufende Sprachinterpretation zuschreiben (und er zieht daraus den Schluss, dass wir keinen Grund haben, nicht-sprachlichen Wesen überhaupt Gedanken zuzuschreiben³¹³). Für Dennett aber spielt die sprachliche Artikulierbarkeit von Gedanken keine besondere Rolle für die Frage, wie Interpretation möglich und ob sie überhaupt am Platze ist.

Aber auch unabhängig davon, ob nicht-sprachlichen Wesen im gleichen Sinne Gedanken zugeschrieben werden können, wie sprachlichen, leuchtet es grundsätzlich nicht ein, wie man mit Hilfe von Dennetts Vorschlag ‚begrifflicher Welten‘ die Referentialität und Wahrheitsbezogenheit von Überzeugungszuschreibungen umgehen können soll. Wir können einem zu interpretierenden Wesen doch nur dann Irrtümer zuschreiben, wenn wir einen (bis zu einem gewissen Grad) *unabhängigen gemeinsamen Bezug* zwischen *unseren* Überzeugungen und *seinen* herstellen. Ohne eine mögliche Unterscheidung von Überzeugungen in solche im *De-dicto*-Modus und solche im *De-re*-Modus, wäre es für einen Interpreten am Ende nicht nur unmöglich, Meinungsverschiedenheiten zwischen ihm und dem zu interpretierenden System zu spezifizieren; er könnte darüber hinaus nicht mal die aus der grundsätzlich bestehenden Perspektivendifferenz sich ergebenden *Unterschiede der Gegebenheitsweise* von Dingen, Zuständen oder Ereignissen *zueinander in Bezug setzen*. Eine berechtigte Zuschreibung von Einstellungen, bzw. das Verstehen eines Interpretandums setzt voraus, dass wir einen Zusammenhang zwischen unseren jeweiligen Perspektiven und Überzeugungen und seinen herstellen können, indem wir den gemeinsamen Bezug fixieren. Überzeugungen *de dicto* zu haben, wäre gewissermaßen ohne realen Gehalt, wären diese nicht intern mit Überzeugungen *de re* verknüpft, die den ‚Sinn‘ gleichsam ‚an die Welt koppeln‘. (Wir werden uns mit diesem Thema noch ausführlicher beschäftigen, da es auch in anderen Hinsichten für meine Kritik an naturalistischen Philosophien der Sprache und des Geistes zentral ist.)

³¹³ Zuerst in D. Davidson: ‚Denken und Reden‘ (1975).

2 Argumentationslinien gegen eine Reduzierbarkeit des ‚Intentionalen‘

Dennetts Vorschlag zur Arbeitsteilung zwischen einer idealisierenden, rationalen Psychologie (‚Kompetenztheorie‘), deren Geschäft eine Verwissenschaftlichung der ‚Alltagspsychologie‘ ist und einer materialistischen, subpersonalen Psychologie (‚Performanztheorie‘), die die *Möglichkeit* mikrostruktureller Implementierungen von kognitiven Prozessen untersuchen soll, macht deutlich, dass er nicht an eine *einheitliche Methode* glaubt, die *beiden* Aufgabenstellungen gerecht werden kann. – Ganz im Gegenteil: Die Vorgehensweise beider Formen von ‚Intentionaler Psychologie‘ müsste unterschieden werden, um deren jeweiliger spezifischer Explanativität gerecht zu werden. Dies bedeutet für Dennett aber keinen (Substanz-) Dualismus, da es am Ende *irgendeine* Form der Strukturierung von Materie sei, die den Geist (bzw. die Geiste) hervorbringe. *Wie* dies jedoch möglich ist sei eine Frage, die für eine Kompetenztheorie schlicht irrelevant sei. Würden wir aber auf eine solche Theorie verzichten wollen, verlören wir die Möglichkeit der systematischen Beschreibung eines *objektiv existierenden Musters*. Wie wir oben bereits sahen, gilt nach Dennett ähnliches ja auch für die ‚intentionalen Charakterisierungen‘ bzw. die teleofunktionalen Begriffe der Evolutionsbiologie.

Dennetts Argumente gegen starke reduktionistische Ansprüche bezüglich des Mentalen lassen sich natürlich zugleich auch als Kritik an einem ‚harten‘ Realismus, wie etwa dem Fodors, verstehen. Denn Dennetts sogenannter ‚instrumentalistischer‘ Ansatz unterstreicht nicht nur die funktionalistische These der multiplen Realisierbarkeit des Mentalen, sondern bestreitet ja zugleich auch, dass man auf der Ebene der Funktionen überhaupt *in gleicher Weise* von Fakten reden kann, wie in paradigmatischen Fällen physischer Gegenstände oder Ereignisse. Dennett unterstützt die quinesche These, nach der die Irreduzibilität ‚intentionaler‘ Idiome und die Unbestimmtheit der Übersetzung (bzw. der Zuschreibung propositionaler Einstellungen) zwei Seiten derselben Medaille sind: Weil es wegen des intrinsisch normativ-idealisierenden (bzw. -optimierenden) und holistischen Charakters der Zuschreibungsbedingungen von ‚Intentionalität‘ keine semantischen Fakten gibt, ist diese auch nichts, was sich im Sinne üblicher naturwissenschaftlicher Reduktionsweisen auf einfachere bzw. irgendwelche Formen physischer Fakten reduzieren ließe. Überzeugungen oder Wünsche könnten eben nicht, wie Fodor, Loar und andere behaupten, einfach als faktische interne Zustände oder Ereignisse verstanden werden, die sich unter kausale Gesetze subsumieren lassen. Dennett ist

nämlich, wie Schiffer, der Ansicht, dass solche ‚Gesetze‘ so unspezifisch wären, dass sie keinen explanativen Gehalt hätten³¹⁴.

Dennetts Kritik richtet sich auch gegen einen *reduktiv ambitionierten* ‚Turingmaschinenfunktionalismus‘, den er ja ansonsten selbst gern zur Plausibilisierung seiner materialistischen Theorie des Geistes heranzieht. Denn es stellt noch immer eine Form von Reduktionismus dar, wenn man glaubt, mentale Typen jeweils direkt mit ‚nicht-intentional‘ charakterisierten funktionalen Typen identifizieren zu können – auch wenn diese in multipler Weise physisch realisiert sein können. Wie wir oben bereits sahen, gerät der Funktionalismus aber in große Schwierigkeiten, wenn nicht nur die grundsätzliche Möglichkeit einer physischen *Simulation* kognitiver Prozesse aufgezeigt, sondern propositionale Einstellungen als mit kausalen Realisierungen identifizierbar plausibel gemacht werden soll. Auch Dennetts Kritik an solchen Reduktionsstrategien läuft vor allem darauf hinaus, dass die komplexen Identitätskriterien propositionaler Gehalte – mit denen wir ‚intentionale Zustände‘ oder ‚Ereignisse‘ individuieren – auf der Ebene einer ‚nicht-intentionalen‘ Beschreibung keine Entsprechung finden können. Er führt dabei unter anderem oben bereits vorgebrachte Argumente, wie etwa das Anzahlproblem ins Feld, also die Schwierigkeit, mit den unzähligen möglichen impliziten Überzeugungen eines Intentionalen Systems umzugehen, welche logische Konsequenzen seiner Überzeugungen darstellen. Sind all diese, so fragt sich auch Dennett, etwa irgendwie funktional bzw. physisch realisiert? Wenn dies unsinnig erscheint, wo zieht man eine Grenze? Vor allem aber gehört nach Dennett mehr zur Repräsentationalität von Zuständen oder Ereignissen als ‚pure Syntax‘, mehr als eine interne Beschreibung einer formale Struktur:

„There is no way to capture the semantic properties of things (word tokens, diagrams, nerve impulses, brain states) by a micro-reduction. Semantic properties are not just relational but, you might say, superrelational, for the relation a particular vehicle of content, or token, must bear in order to have content is not just a relation it bears to other similar things (e.g., other tokens, or causes of tokens, or sets of tokens, or causes of tokens) but a relation between a token and the whole life – and counterfactual life – of the organism it ‚serves‘ *and* that organism’s requirements for survival *and* its evolutionary ancestry.“ (Dennett 1981b, S. 65)

Dennett vertritt also, wie Millikan, Dretske und andere die Position, dass semantische Charakterisierungen von Zuständen oder Ereignissen auf einen teleologischen Funktionsbegriff angewiesen sind. Teleologische *Idiome* für ‚Intentionales‘ seien *in explanativer Hinsicht nicht substituierbar*, was aber nicht bedeute, dass die

³¹⁴ S. etwa Dennett (1981), S. 57.

zugrundeliegenden Mechanismen nicht *rein kausal* bzw. *physischer Natur* sein können (bzw. müssen). Die von Dennett angeratene Vorgehensweise hatten wir zum Teil bereits im Rahmen der Erörterung des Computermodells des Geistes dargelegt. Demnach sollten wir uns nicht mit dem aussichtslosen Unterfangen abgeben, vermeintliche ‚intentionale Fakten‘ direkt auf physische reduzieren zu wollen, sondern mit der Einsicht zufrieden geben, dass wir Maschinen entwickeln und deshalb in ihrem kausalen Zusammenhang vollkommen verstehen können, denen wir aus der Intentionalen Perspektive mit vollem Recht semantische Eigenschaften zusprechen können. Zwischen solchen Maschinen und uns gäbe es keinen prinzipiellen, sondern nur einen graduellen, die Komplexität betreffenden Unterschied. Auf die grundlegenden begrifflichen Schwierigkeiten von Dennetts Ansicht, dabei – anhand eines Homunkuli-Modells des Geistes – auch noch die Möglichkeit fließender Übergänge von der ‚intentionalen‘ zu einer physischen bzw. mechanistischen Redeweisen plausibel machen zu können (was nicht mit Typenreduzierbarkeit zu verwechseln ist!), bin ich oben bereits eingegangen.

Ungeachtet dessen können wir natürlich der *kritischen* Pointe Dennetts und anderer Teleofunktionalisten nur beipflichten: Wie genau man auch die kausalen Beziehungen zwischen den inneren und zwischen den inneren und äußeren Ereignissen Intentionaler Systeme auf einer Ebene rein physischer Beschreibungen untersucht, man wird dadurch nicht in die Lage versetzt, ‚intentionale Zustände‘ oder ‚Ereignisse‘ zu identifizieren. Extrem komplex strukturierte Formen von Materie mögen auch für sich betrachtet bereits einen Hinweis dafür liefern, dass wir es mit einem Intentionalen System zu tun haben könnten. Aber sie liefern *per se* nicht den geringsten Hinweis darauf, *welche* Gehalte oder Bedeutungen sie ‚tragen‘ – oder vielmehr, *sie repräsentieren oder bedeuten per se überhaupt nicht*.

Allerdings drängt sich auch mit Blick auf Dennetts Theorie Intentionaler Systeme der Einwand auf, dass dies zumindest noch nicht die *ganze* Geschichte zu den Bedingungen propositionaler Einstellungen und von ‚Intentionalität‘ ganz allgemein sein kann – mal ganz abgesehen davon, dass es, wie ich meine, auch nicht die *richtige* Geschichte ist. Dies lässt sich schon daraus ersehen, dass Zuschreibungen aus der Intentionalen Perspektive ja bereits entsprechende Kompetenzen seitens der *Subjekte* der Zuschreibung voraussetzen. Die *Möglichkeit* der *Einnahme* des *Intentional Stance* kann aber nicht ihrerseits einfach durch die Bedingungen erklärt werden, unter denen sich ‚intentionale Einstellungen‘ berechtigterweise zuschreiben lassen. Und nur letzteres ist ja Dennetts Thema. Auf dieses Defizit weist auch Brandom hin:

„Distinguishing *simple intentional systems*, which are merely intentionally interpretable, from *interpreting intentional systems*, systems toward which the

intentional stance can be adopted from systems that can adopt that stance toward others, is distinguishing instituting intentionality from instituted intentionality. Simple intentionality, which on this line is in the eye of the beholder, is for that reason *dependent* on and in an important sense *derivative* from the intentionality exhibited by interpreters.“ (Brandom 1994, S. 59 f.)

Zwar unterscheidet auch Dennett *interpretierende* Intentionale Systeme von nur *zu interpretierenden* als diejenige Teilklasse Intentionaler Systeme, die sich durch Sprachlichkeit und Kommunikation auszeichnet. Und er gesteht sogar zu, dass „without the talking intentional systems, of course, there would be no ascribing beliefs, no theorizing, no assuming rationality, no predicting“ (Dennett 1971, S. 17). Doch bringt ihn diese Einsicht in der Folge nicht dazu, die Sonderstellung interpretierender bzw. sprachlicher Intentionaler Systeme *in systematischer Weise* zu berücksichtigen. Dies wäre nur dann nachvollziehbar, wenn Dennett – entgegen der Deutung Brandoms und der meisten anderen Interpreten – sich doch soweit als ‚Realist‘ verstünde, dass er die jeweils *zugeschriebenen Muster* bzw. die Struktur der ‚intentionaler Eigenschaften‘ als in Intentionalen Systemen faktisch realisiert (wenngleich nur aus einer Intentionalen Perspektive epistemisch zugänglich) annehmen könnte, ganz gleich ob es sprachbegabte Wesen gäbe, die sie *tatsächlich* zuschreiben. Für eine solche Interpretation Dennetts lassen sich allerdings keine eindeutigen Anhaltspunkte finden³¹⁵.

Jedenfalls lässt sich gegenüber Dennett geltend machen, dass wir was ‚Intentionalität‘ ist oder ausmacht nicht allein dadurch erklären können, dass wir uns auf Fragen der Berechtigung und der Form ihrer Zuschreibung aus der Perspektive eines singulären Beobachters beschränken. Denn dies setzt ja voraus, dass wir bereits wissen was es heißt, ein solches Intentionales System zu sein, dass ‚Intentionalität‘ zuschreiben kann. Wir werden noch zu fragen haben, ob dieser Vorwurf nicht in ähnlicher Weise auch gegenüber Davidson erhoben werden muss. Zunächst müssen wir uns aber die wesentlichen Züge seiner Philosophie des Geistes vergegenwärtigen, insbesondere die Argumente gegen eine Reduzierbarkeit ‚psychologischer‘ Eigenschaften auf ‚nicht-psychologische‘.

Davidson charakterisiert seine Position zum Verhältnis zwischen Geistigem und Physischem durch den in der ‚analytischen Szene‘ mittlerweile einschlägigen Terminus ‚*Anomaler Monismus*‘³¹⁶. Davidsons Gründe für die damit verbundenen Thesen können

³¹⁵ Obwohl man vielleicht manche Stellen bei Dennett – etwa in Dennett (1991) – so lesen kann.

³¹⁶ Zuerst in Davidson (1970); vergl. auch ders.: ‚Donald Davidson‘ (1994a).

nur vor dem Hintergrund seiner ontologischen Voraussetzungen verstanden werden, nach denen was es gibt, Token von Ereignissen sind, die nicht durch Beschreibungen individuiert sind, sondern, wie Davidson zunächst meinte, durch ihre Position in einem Netz kausaler Beziehungen. ‚Monismus‘ bedeutet, dass es nur eine Art von Substanz gibt, nämlich die Physische und dass physische und geistige Zustände oder Ereignisse *tokenidentisch* sind. So etwas wie ‚Geiste‘, gibt es demnach nicht. Dagegen gibt es Personen, Zustände oder Ereignisse, die geistige Eigenschaften haben insofern entsprechende psychologische Prädikate von ihnen wahr sind. ‚Anomal‘ heißt nun für Davidson, dass es zwischen physischen und geistigen Ereignissen zwar kausale Beziehungen geben kann, diese Beziehungen aber keine (strikten oder ausnahmslosen) Gesetze instantiieren. Wichtig ist dabei, dass Davidson Gesetzesartigkeit, im Gegensatz zu Kausalität und Identität, nicht als eine extensionale Beziehung versteht, die zwischen individuellen Zuständen oder Ereignissen besteht, sondern als intensionale Beziehung, die zwischen Beschreibungen besteht. Strikte Gesetze könnten nur durch physische (genauer: physikalische) Ereignisse instantiiert werden. Die von Davidson behaupteten kausalen Beziehungen zwischen mentalen und physischen Ereignissen haben ihre Grundlage also in der Gesetzesartigkeit ihrer tokenidentischen Beziehung unter einer physikalischen Beschreibung³¹⁷. Davidsons These von der ‚Anomalie des Monismus‘ in Bezug auf geistige Eigenschaften läuft dann darauf hinaus, dass die Bereiche psychologischer und physischer Beschreibungen eine Form von *Heteronomie* aufweisen, die (strikte) *gesetzesartige Verknüpfungen* bzw. eine *Reduktion* durch ‚Brückengesetze‘, *unmöglich* machen – wobei sich seine Argumente hierfür im Laufe der Zeit allerdings etwas verschoben haben.

Ein durchgängiges Hauptmotiv Davidsons für die Heteronomiethese besteht sicherlich in dem von ihm immer wieder betonten besonderen, nämlich rationalen und holistischen Charakter des Mentalen. Weshalb dies zu einer nomologischen

³¹⁷ Davidson behauptet allerdings mittlerweile, dass der Kausalbegriff ein Merkmal von Handlungserklärungen bzw. des Geistigen ist, *weil* wir dort keine strikten Gesetze kennen. Wie Quine und Russell ist er nämlich zu der Ansicht gelangt, reife Naturwissenschaften seien gerade dadurch gekennzeichnet, dass ihre Erklärungen ohne den Kausalbegriff auskommen (s. etwa D. Davidson: ‚Subjektiv, Intersubjektiv, Objektiv‘ (1991)). Dies ist eine deutliche und in ihren Konsequenzen folgenreiche Verschiebung, zumal Davidson uns nun eine Analyse des Kausalbegriffs schuldig bleibt, die *nicht* auf der Gesetzesartigkeit entsprechender Erklärungen aufruhrt. Außerdem ist damit natürlich auch sein Individuationskriterium für Ereignisse hinfällig, die er nun, wie Quine, als durch ihre jeweilige Raum-Zeit-Stelle individuiert versteht (s. D. Davidson (1985). Allerdings hatte er selbst diese These ursprünglich abgelehnt, weil er es für möglich hielt, dass an der gleichen Raum-Zeit-Stelle durchaus verschiedene Ereignisse stattfinden können. Sollte es aber keine plausible Ereignisindividuation im Sinne Davidson geben, wäre natürlich zugleich auch Davidsons *Tokenidentitätsthese* von Physischem und Psychischem, die ja die Intransitität, bzw. *Beschreibungsunabhängigkeit* Ereignisindividuation abhängt, betroffen

Unvereinbarkeit zwischen psychologischen und physiologischen Vokabularen führen soll, wird von ihm nun aber zunächst (in ‚Mentale Ereignisse‘) damit begründet, dass allgemeine Aussagen, die das Geistige und das Physische miteinander verknüpften, *wegen der Unbestimmtheit* der Übersetzung, welche bei der Zuschreibung propositionaler Einstellungen – nach Maßgabe der Methode der Radikalen Interpretation – ins Spiel kommt, notwendigerweise *heteronom* sein müssten. Heteronom sind nach Davidson solche Verallgemeinerungen, die sich nicht, wie homonome, innerhalb der gleichen Art von Vokabular präzisieren lassen, weshalb es nur letztere ermöglichen, zu abgeschlossenen oder strikten Gesetzen (also solchen ohne *ceteris-paribus*-Charakter) zu gelangen.

Davidsons Heteronomie- bzw. Irreduzibilitätsthese wird in philosophischen Debatten häufig herangezogen, ohne dass dabei die Argumente, die für sie sprechen sollen wirklich hinlänglich klar werden. Und auch Davidson selbst ist, wie ich meine, in diesem Punkt nicht sehr luzide. In ‚Mentale Ereignisse‘ läuft seine Überlegung offenbar darauf hinaus, dass sich die Heteronomie des Verhältnisses zwischen Geistigem und Physischen aus der Unterschiedlichkeit der Art bzw. der ‚Härte‘ der Belege und ihres Verhältnisses zur jeweiligen Theorie ergeben. Wie Quine, so behauptet auch Davidson, dass es nicht *die* Interpretationstheorie geben kann, die einzig zu allen möglichen behavioralen Belegen passt. Unter der Maßgabe notwendiger Optimierungsprinzipien wäre eine Interpretationstheorie angesichts neuer Belege (Äußerungen in Situationen) einem prinzipiell nicht endenden Anpassungsdruck ausgesetzt. Anders als eine Theorie des physischen Naturgeschehens sei eine Interpretationstheorie damit – so verstehe ich Davidsons Pointe – quasi *intrinsisch unterbestimmt*.

Dieses Argument ist aber insofern fragwürdig, als es doch mittlerweile weitgehend unstrittig ist, dass *jede* Form von Theorie gegenüber tatsächlichen und möglichen Belegen unterbestimmt ist. Unterbestimmtheit kann also, wie Davidson inzwischen selbst einräumt³¹⁸, *per se* noch nicht dasjenige sein, was die Unterschiedlichkeit des Bereichs des Geistigen ausmacht. Ein wichtiger Unterschied besteht jedoch in der Tatsache, dass sich der Interpret nicht nur, wie etwa der Physiker, *in der Praxis der*

³¹⁸ Davidson: ‚Three Varieties of Knowledge‘ (1991a), S. 161 f. Interessanterweise sieht Davidson mittlerweile auch die für den Bereich des Mentalen (angeblich) typische *Verwendung kausaler Begriffe* als für die fragliche Heteronomie maßgeblich. Solche Begriffe verwendeten wir immer dann, wenn wir kein striktes Gesetz kennen. In den strikten Gesetzen der der modernen Physik würden kausale Begriffe dagegen keine Rolle mehr spielen (s. etwa Davidson (1991)). Allerdings räumt Davidson zugleich indirekt ein, dass es auch für die Biologie, Geologie, Meteorologie und andere Wissenschaften gälte, dass wir keine strikten Gesetze kennen, ohne dass wir deshalb meinen, dort den Bereich des Physischen verlassen zu haben. Damit scheint jedenfalls die Tatsache der Verwendung kausaler Begriffe als Unterscheidungskriterium nicht tauglich.

Beschreibung des Phänomenbereichs an den eigenen Normen der Rationalität orientiert, sondern darüber hinaus die Rationalität *in den Phänomenen selber* beschreiben muss. Die *Objektivitätsstandards* für Urteile bezüglich des Rationalen und Wahren lassen sich nach Davidson jedoch nicht im gleichen Sinne als *perspektivenübergreifend* verstehen, wie bei Urteilen bezüglich der Gegenstände der Naturwissenschaften:

„Here lies the source of the ultimate difference between the concepts we use to describe mental events and the concepts we use to describe physical events, the difference that rules out the existence of strict psychophysical laws. The physical world and the numbers we use to calibrate it are common property, the material and abstract objects and events that we can agree on and share. But it makes no sense to speak of comparing, or coming to agree on, ultimate common standards of rationality, since it is our own standards in each case to which we must turn in interpreting others. This should not be thought of as a failure of objectivity, but rather as the point at which ‚questions come to an end‘.“ (Davidson 1994, S. 232 f.)

Diese Passage ist in mancherlei Hinsicht aufschlussreich. Denn es scheint nun so, als ob Davidson die Andersartigkeit bzw. Nicht-Reduzierbarkeit des Geistigen letztendlich mit der Behauptung der *intrinsic Subjektivität* seiner Beschreibungsbedingungen begründen will. Dieses Argument lässt sich auch so reformulieren³¹⁹: Die Objektivität der Beschreibung physischer Dinge bzw. der Messung physischer Größen wird durch ‚intersubjektive‘ Übereinstimmung oder Übereinkunft ermöglicht, deren Grundlage sprachliche Kommunikation ist. Sprachliche Kommunikation und das ‚Erkennen des Fremdpsychischen‘ bzw. die wechselseitigen ‚Interpretationen‘ individueller Äußerungen und die Zuschreibung geistiger Zustände oder Ereignisse könne nun aber nicht ihrerseits – gleichsam ‚von außen‘ – auf ihre Objektivität hin befragt werden, da sie erst deren Grundlage bilden. Letztgültiger Maßstab für die Beurteilung der Bedeutung von Äußerungen und mentaler Eigenschaften seien die Urteile interpretierender Subjekte, wenngleich nur im Rahmen ‚sprachlicher Interaktion‘. Theorien über Physisches und Theorien über Bedeutung und Geist unterlägen demnach unterschiedlichen Gültigkeitsbedingungen. Dieses Argument wird allerdings erst auf der Grundlage von Davidsons ‚Triangulationsmodell‘ sprachlicher Kommunikation, dass in jüngeren Arbeiten Davidsons eine große Rolle spielt, besser verständlich. Wir werden es im Folgenden etwas ausführlicher diskutieren.

³¹⁹ S. auch Davidson (1991), S. 1012.

3 Davidson über Voraussetzungen für ‚Gedanken‘

Ein Punkt in dem sich Davidson von praktisch allen im Rahmen dieser Arbeit diskutierten naturalistischen Philosophen unterscheidet ist seine These, sprachliche Kommunikation sei eine Voraussetzung für ‚Gedanken‘. Dies ist insofern bemerkenswert, als Davidson ja auch einer intentionalistischen Theorie der Bedeutung sprachlicher Ausdrücke anhängt, die stark an gricesche Formulierungen erinnert. An Grice orientierte Analysen sprachlicher Bedeutung werden aber üblicherweise (etwa bei Bennett und Schiffer) mit der Voraussetzung verknüpft, begriffliche Kompetenz und der ‚Besitz‘ von Gedanken sei unabhängig von sprachlicher Verständigung möglich. Eine wesentliche Rolle in Davidsons gegenteiliger Argumentation übernimmt dabei mittlerweile das *Triangulationsmodell*, welches die intersubjektiven Konstitutionsbedingungen von Objektivität und Gehalt beschreiben soll. Nach diesem Modell soll, zumindest bei einfachen, wahrnehmungsnahen Überzeugungen, die *Gehalt bestimmende Ursache* nur durch die Überschneidung zweier unterschiedlicher Perspektiven zu fixieren sein.

Während ich nun der Stoßrichtung der von Davidson mit dem Triangulationsmodell vorgetragenen Kritik an individualistischen Konzeptionen des Geistigen grundsätzlich zustimme, will ich in der Folge aber auch plausibel machen, dass Davidson dabei Intersubjektivität auf einer zu schmalen, nämlich quasi-behavioristischen, Ausgangsbasis entwickelt – zu schmal jedenfalls, um ein angemessenes Verständnis von Intersubjektivität als einer Voraussetzung von Objektivität und Gehalt etablieren zu können. Darüber hinaus werde ich auch dafür argumentieren, dass Davidsons empiristisch-wahrheitssemantische Konzeption des Bedeutungswissens und der Verständigung nicht nur wegen den ihr zugrundeliegenden bedeutungsintentionalistischen Annahmen in Schwierigkeiten gerät, sondern eigentlich auch kaum Interessantes zum Verständnis der wesentlichen Eigenschaften *propositionaler* Verständigung beisteuert.

Die These, dass *Sprache eine Voraussetzungen für Gedanken* ist, genauer gesagt, dass nur ein sprachliches Wesen propositionale Einstellungen haben kann, hat Davidson erstmals ausführlich in ‚Rational Animals‘ (1982a) zu begründen versucht. Wenn wir etwa, unter Verwendung eines Beispiels von Norman Malcolm³²⁰, einem Hund die Überzeugung zuschreiben, dass die Katze auf die vor ihm befindliche Eiche geklettert ist, sollten wir uns nach Davidson hüten, solche Zuschreibungen für voll zu nehmen. Denn es ist ja völlig unklar, in welcher Weise wir jene Überzeugung im

³²⁰ In: N. Malcolm: ‚Thoughtless Brutes‘ (1972-73).

Einzelnen beschreiben könnten, von der wir sagen, dass der Hund sie habe: Etwa so, dass die Katze auf den ältesten Baum in der Umgebung geklettert ist? Oder so, dass sie auf denselben Baum geklettert ist wie gestern, als der Hund sie verfolgte? Anhand des *Verhaltens* des Hundes *allein* lässt sich dies offenbar nicht sinnvoll beantworten. Damit aber scheint das Kriterium der semantischen Undurchsichtigkeit des ‚Intentionalen‘ keine Anwendung finden zu können, demzufolge die Wahrheit propositionaler Zuschreibungen auch von der Art der Beschreibung des Gehaltes abhängt.

Außerdem stimmt Davidson mit der schon mehrfach zur Sprache gekommenen These überein, dass einzelne Gedanken vor allem auch durch ihre Position in einem dichten Netzwerk von anderen Gedanken bestimmt sind. Dann aber gelangen wir im Falle des Hundes schnell an die Grenzen des Plausiblen:

„If we really can intellegibly ascribe single beliefs to a dog, we must be able to imagine how we would decide whether the dog has many other beliefs of the kind necessary for making sense of the first. It seems to me that no matter where we start, we very soon come to beliefs such that we have no idea at all how to tell whether the dog has them, and yet such that without them, our confident first attribution looks shaky“ (Davidson 1982a, S. 321)

Auch wenn es, wie oben bereits argumentiert wurde, sicher keine eindeutige Liste von Überzeugungen gibt, die das inferentielle Beziehungsgeflecht eines Gedankens bestimmen, so scheint es jedenfalls eine gewisse Menge an unterschiedlichen wahren Überzeugungen zu sein, die jemand haben können muss, dem wir berechtigterweise einen bestimmten Gedanken im vollwertigen Sinne zuschreiben. Dazu gehören solche Überzeugungen, die sich schon rein logisch ergeben, aber auch solche, die mit empirischen Generalisierungen zu tun haben, wie etwa im vorliegenden Fall die, dass Katzen auf Bäume klettern können und situationsbezogene, wie dass dieselbe Katze, die eben noch flüchtete, nun im Baum versteckt ist. Dies bedeutet für Davidson, dass erst die Beobachtung eines ziemlich komplexen Verhaltensmusters die Zuschreibung eines einzelnen Gedanken rechtfertigen kann. Ein solch komplexes Muster könne es aber ohne Sprache nicht geben. Davidson präzisiert dies dahingehend, dass ein Wesen nur dann ein denkendes, rationales Wesen sein könne, wenn es viele Gedanken sprachlich *auszudrücken* könne und vor allem auch selbst in der Lage sei, *Interpret* von Äußerungen und Gedanken anderer zu sein. Allerdings wird diese Behauptung, wie Davidson selbst einräumt, durch die bisherigen Argumente zwar plausibel gemacht. Durch sie wird noch nicht wirklich nachgewiesen, dass Sprache eine notwendiger Bedingung für Gedanken ist. Um dies zu leisten, will Davidson zeigen, dass den *Begriff einer Überzeugung* zu haben notwendige Bedingung dafür ist, überhaupt eine

Überzeugung zu haben und dass eine *Sprache* zu haben notwendige Bedingung dafür ist, den Begriff einer Überzeugung zu haben.

Dazu erinnert uns Davidson zunächst an die unstrittige Tatsache, dass für (die meisten unserer) Überzeugungen gilt, dass wir *überrascht* werden können, indem wir feststellen, dass wir uns *geirrt* haben. Dieses Gewähr-Werden des Unterschieds hinsichtlich dessen, was man geglaubt haben und dem, von dem man nun überzeugt sind, impliziert aber eine Überzeugung *über* eine Überzeugung, denn es besteht unter anderem darin, dass man nun seine ursprüngliche Überzeugung für falsch hält. Damit aber habe man, so Davidson, nicht nur den Begriff einer Überzeugung, sondern auch den Begriff einer objektiven Wahrheit, also einer Realität, die von den je eigenen Überzeugungen unabhängig ist. Worin aber würde sich ein Verfügen über die Subjektiv-objektiv-Unterscheidung *im Verhalten zeigen* können? Davidsons Antwort lautet: In der *gelingenden linguistischen Kommunikation*. Wie wir bereits sahen, fordert ja bereits eine der Optimierungsannahmen, nämlich das ‚Korrespondenzprinzip‘, dass wir grundsätzlich von denselben Dingen überzeugt sein, also eine gemeinsame objektive Welt unterstellen müssen, wenn wir überhaupt eine Chance haben sollen, die Äußerungen eines Anderen zu verstehen. Diese gemeinsame objektive Welt sei, wie Davidson nun sagt, diejenige, die sich aus dem *Schnittpunkt der unterschiedlichen Perspektiven* zweier Wesen ergibt, welche die sprachlichen Äußerungen des jeweils anderen interpretieren.

Wie dies genauer zu verstehen sein kann, wird vielleicht deutlicher, wenn wir noch einen anderen Gesichtspunkt mit einbeziehen, unter dem Davidson diesen Gedanken ebenfalls entwickelt, nämlich seine Kritik an dem in empiristischen Kreisen üblichen ‚proximalen‘ Verständnis sinnlicher Belege von Überzeugungen (‚Sinnesdaten‘). Dieses Verständnis wird ja – wengleich in naturalistischem Gewand – in gewisser Weise auch noch von seinem Lehrer Quine vertreten. Gemeint ist die Vorstellung von einer empirischen Evidenzbasis, die dem Subjekt nicht äußerlich ist³²¹. Wie Davidson bemerkt, spielen im Falle Quines Nervenreizungen eine merkwürdige Zwitterrolle. Denn einerseits versteht Quine Nervenreizungen als in die Sphäre kausaler Zusammenhänge einzuordnendes ‚Eingangstor‘ der *Wirkungen* der Welt auf einen Organismus. Andererseits sollen sie zugleich so etwas wie eine vom Holismus ‚unverseuchte‘ *Überprüfungs- oder Belegbasis* von Überzeugungen bzw. Theorien eines Erkenntnissubjektes über die Welt bilden. Wie Davidson dem bereits in ‚A Coherence Theory of Truth and Knowledge‘ (1983) zu Recht entgegenhält, können sie dieser Doppelrolle begrifflich nicht gerecht werden. Denn es bleibt ja, wie schon

³²¹ Einschlägig für eine moderne Kritik *klassischer* (nicht-naturalistischer) Konzeptionen des ‚Gegebenen‘ als sinnlicher Evidenzbasis ist natürlich W. Sellars (1956).

mehrfach betont wurde, völlig unklar, wie man von Nerven-, bzw. Sinnesreizungen – welcher Art auch immer – zu Überzeugungen gelangen soll. Man beschreibt hier *entweder* kausale *oder* ‚logische‘ (inferentielle) Zusammenhänge – aber *nicht beides* zugleich. Mit der Unterscheidung zwischen *proximalen* und *distalen* Theorien der Erkenntnis lenkt Davidson die Pointe seiner Kritik an Quine in ‚Meaning, Truth and Evidence‘ (1990a) jedoch noch in eine andere Richtung. Nervenimpulsen, so lautet dort sein Haupteinwand, sehe man ihre *Ursache* nicht an. Wenn man, wie Quine und andere, davon rede, dass Gegenstände oder Aspekte der Welt in dem Sinne klassifiziert werden, dass Lebewesen manche Reize gegenüber anderen als ähnlich behandeln, dann müsse sich dies in der Ähnlichkeit ihrer Reaktionen manifestieren. Doch was bildet nun das Kriterium der Ähnlichkeit von Reaktionen? Ein solches Kriterium lässt sich schließlich nicht aus den Reaktionen des Individuums selbst ableiten. Es könnte sich, so Davidson, nur aus den *Reaktionen eines Beobachters auf diese Reaktionen* ergeben. Das Problem mit Stimuli sei nämlich folgendes: An der Reaktion eines Lebewesens auf einen Reiz könne man eigentlich nicht erkennen, worin der Reiz am Ende besteht. Reagiert der Hund auf das Klingeln der Türglocke oder auf die Vibrationen der Luft in der Nähe seines Ohres? Wenn wir auf irgendeine andere Weise die gleiche Art von Vibration der Luft in der Nähe des Hundehohres erzeugen könnten, würde der Hund sicherlich genauso reagieren, wie wenn die Ursache der Vibrationen die Türklingel ist.

Worauf Davidson hier hinweist, ist uns ja bereits aus der Diskussion um die Möglichkeit von Fehlrepräsentationen bei Dretske bekannt. Auch hier besteht die Schwierigkeit am Ende darin, dass es bei einer auf ein einzelnes Individuum beschränkten, kausalistischen Beschreibungen von Wahrnehmungen eigentlich keinen Grund gibt, eher die eine zuzuschreiben, als die andere, also etwa, dass die Klingel ertönt und nicht etwa, dass die Luft in der Nähe seines Ohres mit einer bestimmten Frequenz vibriert. Davidson glaubt, dass erst eine Dreiecksbeziehung, in der ein Beobachter die Reaktionen eines Sprechers auf Reize mit seinen eigenen Ähnlichkeitsklassifizierungen abgleicht, die Frage nach der Ursache von Gedanken – und damit nach deren Gehalt – sinnvoll beantworten lässt. Den Kern dieser Überlegung formuliert er folgendermaßen:

„(...) ehe das Dreieck, welches zwei Lebewesen untereinander und jedes dieser Lebewesen mit gemeinsamen Merkmalen der Welt verbindet, abgeschlossen ist, kann es keine Antwort geben auf die Frage, ob jemand beim Unterscheiden verschiedener Reize solche auseinanderhält, die sich an den Oberflächen der Sinnesorgane befinden, oder solche, die sich weiter draußen oder weiter drinnen befindet. Ohne diese Gemeinsamkeit der Reaktionen auf gemeinsame Reize hätte Denken und Reden keinen spezifischen Inhalt – das heißt, sie hätten gar keinen Inhalt. Um der Ursache eines Gedankens einen Ort zu zuschreiben, und so seinen Inhalt zu bestimmen, sind zwei Standpunkte nötig. Diesen Vorgang können wir

uns als eine Art Triangulation vorstellen: Jede der beiden Personen reagiert unterschiedlich auf Sinnesreize, die aus einer bestimmten Richtung heran strömen: Projizieren wir die herankommenden Linien nach außen, ist ihr Schnittpunkt die gemeinsame Ursache. Bemerkend die beiden Personen nun die Reaktionen des jeweils anderen (im Fall der Sprache: die verbalen Reaktionen), kann jeder von ihnen diese beobachteten Reaktionen zu den eigenen, von der Welt herkommenden Reizen in Beziehung setzen. Jetzt kann die gemeinsame Ursache den Inhalt einer Äußerung und eines Gedankens bestimmen.“ (Davidson 1991, S. 1006/7)

Gemäß Davidsons Dreieck laufen dessen Linien also von den sich ‚verständigenden‘ Individuen auf einen gemeinsamen Punkt (der gemeinsamen Ursache) zu bzw. verbinden diese Individuen miteinander. Diese Triangulation soll nun sowohl für Situationen der ‚Verständigung‘, als auch für solche des Spracherwerbs gelten, nur dass in ersteren Regelmäßigkeiten im Verhalten von Sprechern beobachtet, in letzteren dagegen *antrainiert* würden. Auf alle Fälle müssen den Ähnlichkeitsreaktionen des jeweiligen Gegenübers die entsprechenden eigenen zugeordnet werden, um ihnen einen *objektiven* Gehalt geben zu können.

Es bleibt allerdings unklar, wieso es für die Möglichkeit des Vergleichs der Ähnlichkeit von Reaktionen nach dem Triangulationsmodell überhaupt sprachlicher Kommunikation oder ‚Verständigung‘ bedarf. Einen Abgleich zwischen den Reaktionen eines anderen und den je eigenen könnte es doch auch – wenn man die behavioristischen Voraussetzungen mal soweit mitmacht – unterhalb der Ebene sprachlichen Verhaltens geben. Davidsons zunächst für den Zusammenhang von Sprachlichkeit und Geistigkeit vorgebrachtes Argument, nach dem wir einem Wesen Überzeugungen nur im Zusammenhang mit anderen Überzeugungen zuschreiben können, entsprechende komplexe Zusammenhänge aber nur auf der Grundlage *verbaler* Verhaltensbelege hergestellt werden könnten, sollte durch das Triangulationsmodell ja eigentlich argumentativ untermauert werden, nicht umgekehrt.

Allgemein scheint Davidsons Inanspruchnahme von Begriffen wie ‚Kommunikation‘ oder ‚Verständigung‘ unter den Voraussetzungen Radikaler Interpretation als etwas zu hoch gegriffen. Auch die Art und Weise, in der Davidson Kommunikation in einer Triangulationssituation beschreibt, gibt doch eigentlich nicht mehr her, als zwei Wesen, die wechselseitig ihre Geräuschproduktionen beobachten, dabei unterstellen, dass es sich um sprachliche Äußerungen handelt und auf der Grundlage eines Abgleichs ihrer (mutmaßlichen) Stimuli jeweils versuchen, dem jeweils anderen möglichst weitgehend die ‚eigenen Gedanken‘ unterzuschieben, um sich so einen Reim auf dessen Geplapper zu machen bzw. zu soliden Vorhersagen seines wahrscheinlichen Verhaltens zu kommen (oder, im Falle des Spracherwerbs, zwei Wesen, von welchen das eine das andere durch Plappern und begleitende

Konditionierungsmaßnahmen so hinbekommen muss, dass es überhaupt erst in die Lage versetzt wird, so zu plappern, dass man sich schließlich einen Reim darauf machen kann). Jeder der beiden könnte sozusagen auch ‚gegen die Wand reden‘, ohne gewahr zu werden, dass er hierbei von einem anderen interpretiert wird und ohne dass die einen grundsätzlichen Unterschied für den Interpreten machte. Mit ‚Verständigung‘ oder ‚Kommunikation‘ meinen wir aber üblicherweise mehr, als eine solche Form der wechselseitigen ‚Interpretation‘ (im Sinne Davidsons) von *nicht* notwendigerweise *an einen Hörer* (oder eine Hörerschaft) *gerichteten* sprachlichen Äußerungen aus einer Beobachterperspektive.

Zwar redet auch Davidson (wenngleich nicht bei der Erläuterung des Triangulationsmodells), an Grice anschließend, davon, dass es für erfolgreich *Kommunikation* darauf ankomme, einen Sprecher so zu ‚interpretieren‘, wie dieser nach seiner Absicht verstanden werden will. Der Schritt von der nicht notwendigerweise interaktiven ‚Interpretation‘ zur Kommunikation wird bei Davidson jedoch nach meiner Kenntnis nicht systematisch begründet. Außerdem sind die dabei zu unterstellenden kommunikativen Absichten, also – auf Seiten des Sprechers – in einer gewissen Weise verstanden werden zu wollen bzw. – auf Seiten des Hörers oder ‚Interpreten‘ – jemanden in dem Sinne zu ‚interpretieren‘, wie er/sie verstanden werden will, im Rahmen instrumentellen (bzw. ‚strategischen‘) Handelns zu analysieren. Eine solche Voraussetzung entspricht jedenfalls nicht einem zumindest möglich erscheinenden Verständnis von sprachlicher Interaktion als einem primär *verständigungsorientierten* – also auf *Einverständnis* über die Beurteilung erhobener Geltungsansprüche abzielenden – Handeln der Beteiligten³²². Darauf werden wir noch zurückkommen.

In diesem Zusammenhang bleibt auch unklar, in welchem Sinne Davidson die *Unabhängigkeit der Perspektiven* der triangulierenden Subjekte betont. Schwebt ihm vor, dass ein ‚Interpret‘ ein zu interpretierendes Gegenüber *als* seinerseits *interpretierendes Subjekt* mit ins Kalkül ziehen muss? Dafür gibt es in Davidsons Schriften eigentlich keinen Anhaltspunkt. Dass es für eine erfolgreiche Verständigung einerseits (auch) auf die Absicht des Sprechers ankommt, in einer bestimmten Weise verstanden zu werden, andererseits auf die Fähigkeit des Interpreten, diese Absichten zu erfassen, ist zwar unzweifelhaft, jedoch unvollständig, solange dies nicht durch ein wechselseitiges Wissen um die Symmetrie der virtuellen Gleichzeitigkeit von Sprecher- und Interpretenrolle ergänzt wird. Vor allem aber *ergänzen* sich die unterschiedlichen Perspektiven unter den Bedingungen des Triangulationsmodells nicht im Sinne

³²² S. etwa J. Habermas: ‚Handlungen, Sprechakte, sprachlich vermittelte Interaktionen und Lebenswelt‘ (1988).

wechselseitig in ihren Urteilen aufeinander bezogener Subjekte. Davidsons ‚Interpret‘ kann die *Unterschiedlichkeit der Perspektive des Anderen* nicht in der Weise zu seiner eigenen in Beziehung setzen, dass er sich von diesem hinsichtlich eigener Klassifizierungen *korrigieren* oder gar im Hinblick auf die Triftigkeit seiner Gründe *kritisieren* lassen könnte. Da Davidsons die Triangulationssituation im begrifflichen Rahmen von Reizen und Reaktionen beschreibt, können es eigentlich nur die *Reaktionen des beobachtenden Interpreten* selbst sein, die die Grundlage des Abgleichs seiner eigenen Reaktionen (bzw. ihrer auslösenden Reize) mit denen des jeweiligen Gegenüber bilden. Es werden also lediglich zwei aus unterschiedlichen Richtungen hereinkommende Reize gleichsam ‚mechanisch synthetisiert‘. Damit schrumpft die durch das Triangulationsmodell scheinbar etablierte ‚Intersubjektivität‘ schließlich auf die Perspektive eines ‚reagierenden Subjektes‘ zusammen, welches Beobachter eines anderen ‚reagierenden Subjektes‘ ist.

Davidson glaubt zwar, zumindest die *Fehlbarkeit* von Sprechern konzeptuell dadurch zu ermöglichen, dass ein ‚Interpret‘ das tatsächliche Sprachverhalten eines Sprechers dessen mutmaßlichen Mitteilungsabsichten gegenüberstellen kann. Wir werden uns noch zu fragen haben, ob diese Möglichkeit unter den von Davidson genannten Bedingungen wirklich verständlich gemacht werden kann. Ohnehin aber ist die Möglichkeit von Fehlern auf Seiten der Sprecher natürlich nicht dasselbe, wie die *Möglichkeit der Korrektur oder Kritik*. Diese ist bei Davidson offensichtlich gar nicht vorgesehen: Das ‚interpretierende‘ Subjekt erscheint als der Maßstab aller Dinge. Und es bleibt dabei in der methodologischen Rolle eines *Beobachters* von (mutmaßlich absichtlicherweise) Geräusche produzierenden Wesen, der versucht, sich auf solches Verhalten einen Reim zu machen, indem er den Geräuschen im Aufbau der Theorie wahre – bzw. *von ihm* für wahr gehaltene – Sätze zuordnet. Wir müssen uns natürlich grundsätzlich fragen, ob ein solcher Ansatz tatsächlich eine angemessene Analyse dessen liefern kann, was sprachliche Bedeutung und Verständigung im Wesentlichen ausmacht.

4 Geteilte Bedeutung? – Zum sozialen und normativen Charakter von Bedeutung

Die oben bereits angeführte Behauptung Davidsons, dass es so etwas wie eine gemeinsame Sprache eigentlich gar nicht gibt, wirkt *prima facie* genauso provokativ, wie kontraintuitiv. Worin besteht denn schließlich der Sinn von Lexika, warum kämpfen Minderheiten für das Recht, ihre Sprache – als Ausdruck ihrer kulturellen Identität – sprechen zu dürfen? Davidsons Position in dieser Frage ist allerdings nicht ganz so abwegig, wie es zunächst den Anschein haben mag. Und sie bildet eigentlich nur die naheliegende Konsequenz aus zwei seiner grundlegenden sprachphilosophischen Annahmen, nämlich (1.) *Kommunikation* auf der Grundlage eines Lockeschen bzw. bedeutungsintentionalistischen Modells zu verstehen; und (2.) *Bedeutungsverstehen* als der Kenntnis einer faktische Sprecheräußerungen beschreibenden wahren Wahrheitstheorie (im Stile Tarskis) entsprechend zu analysieren.

Ich möchte in der Folge noch zeigen, dass sowohl diese Annahmen, als auch die aus ihnen abgeleiteten Folgerungen teils explanativ unbefriedigend, teils schlicht unplausibel sind. Insbesondere erfahren wir zum einen praktisch nichts über die semantischen Eigenschaften (bzw. die kommunikativen Funktionen) der Propositionalstruktur; zum anderen werden die normativ-idealisierten Aspekte unserer sprachlichen Praxis in wichtigen Hinsichten völlig ausgeblendet, Aspekte, die, wie ich meine, für ein angemessenes Verständnis einiger der kognitiven und praktischen Funktionen von sprachlicher Bedeutung und Verständigung wesentlich sind. Dies gilt vor allem auch im Hinblick auf den – unverkürzt verstanden – *Verantwortlichkeit* implizierenden Handlungscharakter sprachlicher Äußerungen. Wir müssen aber zunächst noch genauer klären, in welchem Sinne die These, dass Sprache, Bedeutung und sprachliche Verständigung primär ‚sozialen‘ Charakters sind von Davidson überhaupt bestritten wird und in welchem nicht. Dabei werde ich mich auch auf Schriften seines Schülers Akeel Bilgrami stützen, dessen emphatische Verteidigung einer bedeutungsindividualistischen und anti-normativischen Position von Davidson ausdrücklich gutgeheißen wird³²³.

³²³ A. Bilgrami: ‚Norms and Meaning‘ (1993); ders.: *Belief and Meaning* (1992), und Davidson: ‚Reply to Akeel Bilgrami‘ (1993a).

Welcher Gemeinsamkeit bedarf Verstehen? – Und was ist eigentlich eine ‚Sprache‘?

Zunächst einmal wäre es ein Missverständnis, Davidson so zu deuten, als wolle er den *öffentlichen Charakter* sprachlicher Bedeutung bestreiten, also eine bedeutungstheoretische Position vertreten, nach der jeder Sprecher *prinzipiell nur selbst wirklich wissen* kann, was er meint bzw. was die von ihm benutzten Wörter und Sätze tatsächlich bedeuten – ganz im Gegenteil: Im Rahmen der Diskussion der Bedingungen Radikaler Interpretation, insbesondere der ‚Nachsichtigkeitsprinzipien‘, wurde ja schon deutlich, dass es für Davidson überhaupt nur dann sinnvoll ist, davon auszugehen, bestimmte Geräuschfolgen hätten eine Bedeutung (seien Sprachverhalten), wenn eine erfolgreiche ‚Interpretation‘ grundsätzlich möglich ist³²⁴. Was einem ‚Interpretieren‘ prinzipiell nicht zugänglich ist, könne für die Bedeutung keine Rolle spielen. Deshalb kann es nach Davidson auch *keine Privatsprache* geben, wenn man damit eine Sprache meint, die insofern nicht-öffentlich ist, als sie nicht ‚interpretierbar‘ ist.

Des weiteren will Davidson natürlich auch nicht die offensichtliche Tatsache bestreiten, dass ‚geteilte Sprachen‘ zumindest in dem Sinne existieren, als Sprecher *faktisch* meist dazu tendieren, so ähnlich zu reden bzw. dieselben Wörter zu gebrauchen, wie andere Leute in ihrer Umgebung, wenn sie dasselbe sagen wollen. Seine These ist, dass es *kein linguistisch signifikantes* und *philosophisch interessantes* Konzept von geteilter Sprache gibt³²⁵. Sprachphilosophisch interessant wäre ein solche Konzept nach Davidson nur dann, wenn es als *für die Möglichkeit von sprachlicher Verständigung notwendig* angesehen werden müsste. In ‚The Social Aspect of Language‘ (1994)³²⁶ beschreibt er diese von ihm abgelehnte Konzeption folgendermaßen: „in learning a language, a person acquires the ability to operate in accord with a precise and specifiable set of syntactic and semantic rules; verbal communication depends on speaker and hearer sharing such an ability (...)“. (S. 1). Es geht Davidson also vor allem darum, die Idee *geteilter Normen* der Sprachverwendung, sei in Form von (expliziten) Regeln, Konventionen oder schlicht in Form einer Praxis,

³²⁴ Vergl. etwa D. Davidson: ‚Was ist eigentlich ein Begriffsschema?‘ (1974), S. 264.

³²⁵ Für diese These hat Davidson starke Fürsprecher von Seiten der wissenschaftlichen Linguistik (s. u. a. N. Chomsky: *Regeln und Repräsentationen* (1980/1981) und ders.: ‚Language and Nature‘ (1995), sowie S. Pinker: *Der Sprachinstinkt* (1994/1996)). Meine Kritik richtet sich auch gegen die dort vertretenen, *philosophisch* ambitionierten Thesen zur Bedeutung.

³²⁶ Dieser Aufsatz Davidsons ist eine Reaktion auf eine entsprechende Kritik Michael Dummetts in ‚„Eine hübsche Unordnung von Epitaphen“. Bemerkungen zu Davidson und Hacking‘ (1986/1990).

als für die Erklärung des Wissens von Sprechern und Interpreten wesentlich zurückzuweisen.

Die Argumente, die Davidson und Bilgrami dafür ins Feld führen, sind sowohl empirischer, als auch konzeptueller Natur. So wird einerseits auf die – auch von Seiten der empirischen Linguistik gegen die ‚Regelfolgesemantiken‘ vorgebrachte – *Tatsache* verwiesen, dass es kaum zwei Leute geben dürfte, die in genau gleicher Weise sprechen. Es gibt, anders ausgedrückt, auch innerhalb von (politisch bestimmten) ‚Sprachgemeinschaften‘ wohl kaum zwei Sprecher, die soweit in ihren Sprachverwendungen übereinstimmen, dass man ihren Sprechweisen ein umfassendes gemeinsames Set semantischer und syntaktischer Regeln zuordnen könnte. Was wir nach dieser Sichtweise tatsächlich vorfinden, sind mehr oder minder große *Ähnlichkeiten zwischen Regelmäßigkeiten individuellen Sprachverhaltens*. Dies schade aber nicht unserer Fähigkeit, andere zu verstehen. Tatsächlich reden Leute ja häufig sogar bewusst ‚abweichend‘, etwa aus sprach-innovativen Motiven, um einen sprachlichen Witz zu machen (wie Davidson am Beispiel von Malapropismen zeigt³²⁷) oder um sich von anderen sozialen Gruppen auch dadurch abzusetzen, dass sie eben anders sprechen.

Auf der konzeptuellen Seite hängt die Plausibilität der Behauptungen Davidsons (und auch Bilgramis) unter anderem von der Annahme ab, dass ein sprachliches Verstehen im Sinne Radikaler Interpretation *möglich* ist und die Beschreibung ihrer Bedingungen alles hergibt, was für sprachliche Verständigung *notwendig* ist. Und dazu gehöre eben nicht, dass Sprecher und ‚Interpret‘ eine *Sprache teilen*. In ‚A Nice Derangement of Epitaphs‘ stellt Davidson die Sache so dar, dass wir im Falle gelingender Verständigung vom *Teilen* der ‚Übergangstheorien‘ von Sprecher und ‚Interpreten‘ reden sollten. Darunter versteht Davidson diejenige Theorie, „die der Interpret tatsächlich verwendet, um eine Äußerung zu deuten, und sie ist ebenfalls die Theorie, die der Interpret nach Ansicht des Sprechers verwenden soll“ (a.a.O., S. 219). Die Übergangstheorie wird nach Davidson auf der Basis einer ‚Ausgangstheorie‘ entwickelt, das heißt, einer Theorie, mit der ein Interpret jeweils zunächst – auf Grund seines allgemeinen sprachlichen und nicht-sprachlichen Wissens und einer Vorabanschätzung des Sprechers (etwa seines sozialen Hintergrunds) – glaubt, diesen ‚interpretieren‘ zu können und von der deshalb auch ein Sprecher glaubt, dass es die Theorie ist, die der ‚Interpret‘ verwenden wird. Im Verlaufe des Gesprächs gleichen sich dann sowohl Ausgangstheorien als auch Übergangstheorien von Sprecher und ‚Interpreten‘ immer mehr an, weil sich der ‚Interpret‘ zunehmend auf die jeweiligen Besonderheiten des Sprechers einstellt, bis seine Übergangstheorie *tendenziell* der

³²⁷ Davidson (1986/1990).

Theorie entspricht, die beschreibt, wie der Sprecher tatsächlich verstanden werden will. Eine derart situative und ‚flüchtige‘ Übergangstheorie entspreche aber offensichtlich nicht dem, was sinnvollerweise unter einer theoretisch beschreibbaren Verstehenskompetenz im Hinblick auf eine Sprache verstanden werden kann. Damit relativiert Davidson natürlich auch den Anspruch seines eigenen Projektes.

Mit dieser Infragestellung jedes sprachphilosophischen Projektes einer allgemeinen theoretischen Beschreibung von Verstehens- oder Verständigungskompetenz wird allerdings den bereits angeführten Argumenten gegen die Notwendigkeit einer geteilten oder gemeinsamen Sprache eigentlich nichts wesentlich Neues hinzugefügt. Und man kann, wie ich, Davidson Kritik an solchen Projekten (wie etwa auch Dummetts) unabhängig von seinen speziellen Ansichten über die *Art* der Beschreibung der Verständigungskompetenz und die ‚explanative Rangordnung‘ des Verhältnisses zwischen Bedeutung und Verstehen durchaus teilen. Denn für Davidson soll die erfolgreiche Übermittlung von Bedeutungsintentionen die Grundlage der Bedeutung sprachlicher Zeichen sein und nicht umgekehrt:

„Success in communicating propositional contents (...) is what we need to understand before we can ask about the nature of meaning or of language, for the concepts of a language or of meaning, like those of a sentence or a name or of reference or of truth, are concepts we can grasp and employ only when the communication of propositional contents is established. Meaning, in the special sense we are interested in when we talk of what an utterance literally means, gets its life from those situations in which someone intends (or assumes or expects) that his words will be understood in a certain way, and they are. (...) It is understanding that gives life to meaning, not the other way around“ (Davidson 1991a, S.11)

Für Davidson bildet also eine Analyse der Bedingungen des Verstehens bzw. des ‚erfolgreichen Kommunizierens propositionaler Gehalte‘ die Grundlage einer Analyse des Konzepts ‚Bedeutung‘ und verwandter Begriffe. ‚Bedeutung‘, ‚Sprache‘ und andere semantische oder linguistische Begriffe, wie ‚Satz‘, ‚Prädikat‘, ‚Referenz‘, ‚Wahrheit‘ u. ä. seien Abstrakta und stünden nicht für empirische Gegebenheiten³²⁸. Ihr Zweck sei es, Beschreibungen oder Erklärungen des Sprachverhaltens oder Wissens kompetenter Sprecher zu ermöglichen. Nach einem solchen Konzept von Sprache gibt es unendlich viele ‚Sprachen‘, die mit den (sich immer nur in einer begrenzten Anzahl von Äußerungen manifestierenden) tatsächlichen sprachlichen Dispositionen von Sprechern vereinbar sind.

³²⁸ Davidson: ‚The second Person‘ (1992), S. 256 f.

Die Verhältnisse in diese Weise zu beschreiben läuft darauf hinaus, eine andere Formulierung für Quines Unbestimmtheitsthese zu geben. Und auch das von Kripke aufgeworfene und vieldiskutierte ‚skeptische Problem‘, nach dem wir allein auf der Grundlage von Fakten bezüglich eines Sprechers nicht wissen könnten, was dieser meint³²⁹, entspricht nach Davidson eigentlich der Behauptung, dass wir auf Grund seiner sprachlichen Dispositionen nicht wissen können, *welche Sprache* er spricht. Und diese Frage sei, so Davidson, letztlich nur ein Spezialfall der Frage nach der Gültigkeit von induktiven Schlüssen.

Diese ‚Deflationierung‘ des Konzeptes einer Sprache im Sinne einer geteilten sprachlichen Praxis resultiert offensichtlich einerseits aus Davidsons empiristischem Nominalismus (oder ‚Regelskeptizismus‘) und seinem bedeutungsintentionalistischen, sprecherrelativen Begriff von Bedeutung andererseits: Was wir beobachten können, sind nicht Sprachen, Regeln oder Konventionen, sondern sprachliches ‚Äußerungsverhalten‘. Und die von einem ‚Interpreten‘ zu erfassende Bedeutung eines durch einen Sprecher verwendeten Ausdrucks soll ja diejenige sein, die er nach dessen Absicht haben soll. Wird ‚Sprache‘ dann, wie bei Davidson, als die durch die Theorie des Interpreten in Bezug auf einen Sprecher bestimmte Menge von sprachlichen Ausdrücken mit *identischer* Bedeutungen (und identischen Syntaxregeln) aufgefasst, muss jeder mögliche, durch Überzeugungs- oder Ausdrucksverwendungsabweichungen bedingte, *Bedeutungsunterschied* als *Sprachenwechsel* gedeutet werden. Und wenn es nicht *die* Referenzsprache gibt, auf die wir uns als etwas Geteiltes beziehen können, können wir uns natürlich nicht die Frage stellen, was *die* Bedeutung eines ihrer Ausdrücke sei.

Diese Ansicht teilt auch Bilgrami in *Belief and Meaning*, wo er sich ausführlich mit einigen Voraussetzungen und einigen der Konsequenzen eines (grundsätzlich) ‚Davidsonianischen‘ Ansatzes beschäftigt. Während Davidson das bedeutungstheoretische *Primat des Idiolekts* allerdings hauptsächlich durch die konzeptuelle Irrelevanz gleicher oder ähnlicher *Redeweisen* (also etwa der Ausdruckwahl) im Hinblick auf eine Theorie des Verstehens begründet, führt Bilgrami darüber hinaus auch Bedeutungsunterschiede im Sinne *begrifflicher Differenzen* ins Feld. Diese wichtige und systematisch folgenreiche Abweichung gegenüber Davidson lässt sich an Bilgramis Überlegungen zur Frage der konzeptuellen Notwendigkeit einer Inanspruchnahme geteilter Normen verdeutlichen. Wie Bilgrami zu Recht feststellt, sind hier nämlich Differenzierungen geboten, da sich ja auch ein Davidsonscher Interpret auf Normen der Rationalität beruft, die für Sprecher und ‚Interpreten‘ *gemeinsam* gelten müssen. Nach Bilgrami müssen wir zwischen ‚wirklichen Normen‘,

³²⁹ Kripke (1982/1987).

deren Gültigkeit von den beobachteten Verhaltensregularitäten *unabhängig* ist und solchen, für die dies nicht gilt (und die folglich die Bezeichnung „Norm“ eigentlich nicht verdienen), unterscheiden³³⁰. Erstere, zu denen zumindest die Regeln der Deduktion und der Entscheidungstheorie zählten, seien für die Möglichkeit von Interpretation bzw. Verständigung konstitutiv. *Solche* Normen *müssen* – so verstehe ich sowohl Davidson, als auch Bilgrami – in dem Sinne als gemeinsam oder geteilt unterstellt werden, als ein ‚Interpret‘ ihre Gültigkeit nicht nur in Bezug auf sich selbst, sondern auch für den zu ‚Interpretierenden‘ unterstellen muss. Dies gälte jedoch, wie Bilgrami (auch hier noch in Übereinstimmung mit Davidson) behauptet, nicht für *lexikalische Normen*, also solche, die die Bedeutung eines Terms etwa im Sinne von intersubjektiv gültigen Verwendungsregeln bestimmen sollen. Solche Normen seien zwar meist in einem pragmatisch-instrumentellen Sinne geboten, insofern ich sie befolgen sollte, wenn ich problemlos verstanden werden will. Sie seien jedoch im Hinblick auf die Möglichkeit der Verständigung ‚*extern*‘.

Bilgrami versteht also unter Normen, die wir im Prozess der Verständigung notwendigerweise als geteilt unterstellen müssen, nur solche, die für sprachliches Verstehen konstitutiv sind. Darunter fielen jedoch nicht nur keine lexikalischen Normen, sondern auch keine *material-inferentiellen Schlussregeln*:

„The idea of material inference is just the idea that one cannot have a concept (say, the concept of a table) without having some other concepts (say, the concept of a middle-sized object). But that merely says I cannot attribute *my* concept of table to someone who fails to have certain other beliefs or concepts. It does not require that any particular concept that *he* has must be the concept dictated by a norm set by my linguistic practice or by my social linguistic practice more generally.“ (Bilgrami 1993, S. 131)

Bilgrami teilt demnach nicht die Auffassung Davidsons, wonach erfolgreiche Verständigung auch eine weitgehende Überlappung der von Sprechern und ‚Interpreten‘ für gültig gehaltenen, material-inferentiellen Beziehungen bzw. der dadurch konstituierten Begriffen voraussetzt. Selbst wenn man, so Bilgrami, eine solche Notwendigkeit zugestehe, sei es ein Irrtum, zu glauben, dass entsprechenden ‚Normen‘ der gleiche Status zukäme, wie formal-inferenziellen oder praktisch-syllogistischen. Sie würden nämlich einen solchen Grad an Kontextabhängigkeit aufweisen, dass wir nicht mehr im üblichen Sinne von einer ‚Verletzung von Normen‘ reden könnten³³¹ (Damit radikalisiert Bilgrami den Idiolektbegriffs im Sinne eines

³³⁰ Darauf, dass die Bezeichnung „Normen“ allerdings auch im Hinblick auf Rationalitäts- oder Vernunftregeln nicht unproblematisch ist, wurde ja in der Einleitung bereits hingewiesen.

³³¹ Bilgrami (1993), S. 131 f.

individualistischen Konzeptes nicht nur der Sprachbedeutung, sondern auch der *begrifflichen Kompetenz*. Dieser starke Individualismus folgt nach seiner Auffassung letztlich aus der *Kombination* von *Holismus* in Bezug auf propositionale Einstellungen und *Bedeutungsintentionalismus*: Da es unter holistischen Voraussetzungen kaum wahrscheinlich wäre, dass je zwei Leute einen Begriff, geschweige denn dasselbe Begriffsrepertoire, teilen, sei auch die Rede von einer ‚gemeinsamen Sprache‘ im Sinne geteilter Bedeutungen substanzlos.

Ich hatten oben schon darauf hingewiesen, dass ein sprecherbezogener Holismus nicht nur die Gemeinsamkeit von Begriffen (begrifflicher Verständnisse) und Bedeutungsverständnissen hinsichtlich verschiedener Sprecher bedroht, sondern am Ende auch die Möglichkeit eines Konzeptes von ‚identischen‘ Begriffen oder Bedeutungen in Bezug auf *einzelne* Sprecher zu unterschiedlichen Zeitpunkten. Anstatt jedoch hieraus die, wie ich meine, naheliegende Konsequenz zu ziehen, die systematische Koppelung von sprachlicher Bedeutung an individuelle Sprecherüberzeugungen in Frage zu stellen – oder zumindest zu lockern, entschließt sich Bilgrami für die harte Tour und akzeptiert einen Bedeutungsindividualismus, der sich weit von vorthoretischen Intuitionen zu entfernen scheint. Er versucht uns zwar dadurch zu beruhigen, dass das Phänomen weitgehender Begriffsdifferenz lediglich auf der ‚*aggregativen*‘ Ebene von *Bedeutungstheorien* (für einen Sprecher) zum Tragen komme³³². Es spiele für *Handlungserklärungen* aber insofern keine Rolle, als hier der Gehalt von Einstellungen ‚*lokal*‘ zu bestimmen sei. Damit ist gemeint, dass in vielen Fällen einer Überzeugungszuschreibung gar nicht alle mit dieser jeweils material-inferentiell zusammenhängenden Überzeugungen in Betracht gezogen werden müssten, jedenfalls dann nicht, wenn es, wie Bilgrami glaubt, primär um Zuschreibungen im Rahmen *kausaler Handlungserklärungen* geht. Falls wir uns etwa fragten, warum jemand gerade Wasser trinkt, so bestünde die naheliegende Erklärung nach Bilgrami darin, dass er/sie Durst hatte und der Überzeugung ist, dass Wasser den Durst löscht. Die Kenntnis der chemischen Mikrostruktur des Wassers (Putnams Beispiel) sei dabei für die ‚psychologische Erklärung‘ gar nicht notwendig. Dazu reichten *lokal geteilte* Überzeugungsgelände, wie eben der, dass Wasser Durst löscht. Wenn ein ‚Interpret‘ und ein zu ‚Interpretierender‘ in einem solchen Fall *verschiedene Begriffe* von ‚Wasser‘ etwa in dem Sinne hätten, dass der eine die Mikrostruktur kennt und der andere nicht, sei dies für ‚psychologische Erklärungen‘ ohne Belang³³³.

³³² Bilgrami (1992), Kap. 4.

³³³ Damit könnte Bilgrami auch Fodors Hauptargument für dessen Verdammung des Holismus zurückweisen. Fodors ‚atomistische‘ Gegenposition fußt ja vor allem darauf, dass die hier geschilderten Konsequenzen des Holismus gesetzesartige Generalisierungen über Überzeugungen und damit kausale Handlungserklärungen, unmöglich machten (s. etwa Fodor

Bilgrami geht es bei dieser Modifikation des davidsonschen Ansatzes insbesondere auch darum, die ‚*Einheit des Gehalts*‘ zu retten bzw. einer Bifurkation in ‚engen‘ und ‚weiten‘ Gehalt zu entgehen. Denn einerseits kann er – auf der semantischen Seite – scheinbar den Argumenten der Externalisten Rechnung tragen und den Inhalt propositionaler Einstellungen als auch von der Umgebung des Subjektes abhängig begreifen. Externalist zu sein heißt für Bilgrami vor allem, dem öffentlichen Charakter von Bedeutung Rechnung zu tragen. Andererseits kann er – auf der Seite der Handlungs- oder Verhaltensklärung – offenbar auch der zentralen Intuition der Anhänger von Theorien ‚engen‘ Gehaltes gerecht werden, nach der dasjenige, wovon das Subjekt nichts weiß, für (kausale) Erklärungen seines Verhaltens irrelevant sein müsse. Damit aber gibt Bilgrami dem davidsonschen Programm einer ‚*Unified Theory of Meaning and Action*‘ einen durchaus andere Charakter: Der Anspruch an eine Bedeutungstheorie im Sinne einer Theorie des Verstehens wird depotenziert, womit zugleich Verstehen als ein möglicher eigenständiger Zweck der Interpretation zugunsten von Verhaltensklärung in den Hintergrund tritt:

„This meaning-theoretic notion of ‚concepts‘ has no psychological reality in the sense that it does not ever pull its weight in explanations. It is a trumped up theoretical posit which provides a pool of resources from which local concepts are selected to compose the contents which do explain behaviour in particular localities. Hence the entire idea of meaning-theory and the concepts of meanings it specifies is exhausted by its role in the following scheme. Contents attributed in particular localities are comprised of concepts, which are themselves cashed out in terms of a selection of beliefs, and the idea of a selection of beliefs innocuously requires the idea of some place from which the selection is made. The idea of a meaning-theory is merely a gesture acknowledging that a selection has to be a selection from somewhere. It has no further role than one of marking that place.“ (Bilgrami 1992, S. 12)

Bedeutungstheorien für einen bestimmten Sprecher sind nach Bilgrami idealisierte Konstruktionen im Sinne ideal vollständiger Spezifikationen ‚seiner Begriffe‘, die sich auf der Grundlage aller möglichen verfügbaren Evidenz aus Hypothesen über alle jene Überzeugungen ergeben, die ein Sprecher mit einem jeweiligen Ausdruck assoziiert. Solche Theorien seien aber für Verhaltensklärungen irrelevant, da es hierfür nur auf die *in der jeweiligen Situation kausal wirksamen* propositionalen Einstellungen ankomme. Die Begriffe, die in tatsächlich verhaltenserklärenden Einstellungen

(1987)). Bilgrami schränkt solche Generalisierungen auf ‚lokal‘ wirksame Realisierungen von Gehalt ein.

zugeschrieben werden, seien – gerade wegen ihrer explanativen Relevanz – notwendig kontextgebunden³³⁴.

Auf die Einzelheiten von Bilgramis ausführlicher Ausarbeitung seiner gewissermaßen radikalisierten Variante eines davidsonschen Ansatzes (und ihrer Konsequenzen – etwa für eine Wahrheitstheorie) kann ich hier nicht weiter eingehen. Doch schon auf der Grundlage des hier Referierten dürfte die Verkürzung ins Auge springen, unter der Bilgrami das Thema ‚Bedeutung‘ angeht. Zwar redet er mitunter so, als ginge es auch ihm um die Frage nach den Bedingungen sprachlicher Verständigung. Seine ganze Behandlung des Themas ‚Bedeutung‘ beschränkt sich jedoch letztlich auf jene eben erläuterte indirekte Rolle bei der Zuschreibung propositionaler Einstellungen, die ihrerseits lediglich als Mittel der Verhaltensvorhersage verstanden werden. Deshalb kann auch die Unterschiedlichkeit möglicher Formen und Zwecke sprachlicher Verständigung, etwa bei Aufforderungen oder Bitten, der Mitteilung subjektiver Eindrücke oder relevanter Sachverhalte oder bei konsensorientierten Diskursen über angemessene begriffliche Verständnisse, in Bilgramis Überlegungen keine systematisch folgenreiche Berücksichtigung finden. Auch in dieser Hinsicht ist Davidson zurückhaltender, da es ihm um eine Theorie des Verstehens der wörtlichen Sprecherbedeutung geht, die Raum lässt für weitergehende, wenngleich von Theorie nicht zu erfassende, pragmatische (bzw. illokutionäre) Zwecke sprachlicher Verständigung.

Ich möchte im Folgenden unterschiedliche, sich aber zum Teil wechselseitig ergänzende Einwände gegen eine davidsonsche Konzeption von Bedeutung bzw. Sprach- oder Verstehenskompetenz vorbringen, Einwände, die auch (oder erst recht) Bilgramis Version betreffen. Meine Kritik wird im Kern darauf hinauslaufen, dass eine solche Konzeption keine kategorial *angemessene* und sozusagen ‚vollwertige‘ Analyse der kognitiven Struktur unserer sprachlichen Verständigungspraxis und der notwendigen Fähigkeiten bietet, an dieser zu partizipieren. Dazu will ich zuerst – mit unterschiedlichen argumentativen Strategien – die These angehen, ein gemeinsamer

³³⁴ So glaubt Bilgrami auch Kripkes Problem scheinbar widersprüchlicher Überzeugungen innerhalb einer Person lösen zu können (vergl. Kripke (1979)). Die Möglichkeit einer solchen Widersprüchlichkeit schien sich ja aus Kripkes bekanntem Beispiel von Pierres Nichtwissen darüber zu ergeben, dass ‚seine‘ Begriffe, die er mit den Ausdrücken „Londres“ und „London“ verbindet, den gleichen Bezug haben. Da der (nach Maßgabe des Interpreten) tatsächliche Bezug für eine Erklärung von Pierres Verhalten jedoch irrelevant sei, könnten wir ihm, so Bilgrami, einfach *zwei* Begriffe (London₁ und London₂) zuschreiben, je nach dem, auf welchen Kontext sich die Interpretation zum Zwecke der Verhaltensvorhersage beziehen muss (Bilgrami (1992), S. 131 f.)

Sprachgebrauch bilde *keine* notwendige Bedingung der Möglichkeit sprachlicher Verständigung. Die Klärung dieser Frage wird allerdings weitgehend von der grundlegenden Frage nach einer angemessenen Konzeption von Verstehen und sprachlicher Verständigung, etwa im Sinne eines Kriteriums ihres Erfolgs, abhängen.

Zunächst möchte ich zur Begrenzung des wirklich Strittigen der genannten These zunächst schon mal in dem ziemlich trivialen Sinne zustimmen, dass Verständigungsteilnehmer, die der Sprache (im alltagsüblichen Sinne, also etwa Englisch) ihres jeweiligen Gegenübers mächtig sind, grundsätzlich nicht dieselbe Sprache *verwenden* müssen, um sich verständigen zu können (ich könnte mich mit jemandem natürlich auch dann verständigen, wenn er/sie in Englisch spricht, und ich in Deutsch). Es ist auch richtig, dass wir unter einer Sprache nicht so etwas verstehen sollten, wie ein präzises und umfassendes System von explizierbaren Ausdrucksverwendungsregeln, dessen alle Sprecher kompetent sein müssen, die sich in dieser Sprache verständigen wollen. Es gibt also auch keine klare Grenze zwischen dem Können und Nicht-Können bzw. Verstehen und Nicht-Verstehen einer Sprache. Was ich aber bestreiten möchte ist, dass die von Davidson und Bilgrami vorgebrachten Argumente einen ‚explanativen‘, bzw. die *philosophische Analyse betreffenden Vorrang*, einer *individualistischen Konzeptionen von Bedeutung* (bzw. von Ideolekten) *gegenüber dem Konzept einer als allgemein verbindlich verstandenen Ausdrucksbedeutung* (also einer Sprache) stützen³³⁵.

Auch der eingeschränkte Sinn, in dem ich der genannten These Davidsons zustimme, beruht offensichtlich auf einem Begriff von ‚Sprache‘, der nicht Davidsons ist. Für Davidson (nicht aber für Bilgrami) müssen Sprecher und ‚Interpret‘, soll Verstehen überhaupt möglich sein, eine ganze Menge an Überzeugungen bzw. Begriffen teilen. Eine Konsequenz von Davidsons und Bilgramis Bedeutungs- und Sprachbegriff ist aber, dass ‚Verständigung‘ (wechselseitige ‚Interpretation‘ oder ‚Erklärung‘) prinzipiell auch dann möglich sein muss, wenn zwei miteinander Sprechende jeweils *ständig* die *Art und Weise*, in der sie ihre *Mitteilungsabsichten sprachlich artikulieren*, variieren bzw. ‚die Sprache‘ (im Sinne Davidsons) wechseln. Das heißt, keine Ausdrucksverwendung *müsste* als gleich gemeint wiederholt werden.

Dies wie Davidson es tut³³⁶ auch nur als theoretisch – wenngleich nicht unbedingt praktisch – möglich hinnehmen zu wollen, erscheint mir aber aus mehreren Gründen unplausibel. Einer davon ist, dass die Annahme dieser Möglichkeit entweder mit der

³³⁵ In dieser Hinsicht übereinstimmend ist auch die Position Dummetts; s. etwa: ders: *The Logical Basis of Metaphysics* (1991), insbes. Kap. 4. Auf einige von Dummetts Argumenten werde ich unten noch zurückkommen.

³³⁶ Vergl. insbes.: D. Davidson: ‚The Social Aspect of Language‘ (1994).

Unterstellung einhergehen müsste, man könnte von symbolischen Token (welcher Art auch immer) unter Absehung von einer mit ihnen assoziierten Bedeutung reden bzw. als sie solche identifizieren. – Dass dies nicht einleuchtet, hatten wir ja schon gegen kausal-funktionalistische Theorien des Geistes und der Bedeutung eingewendet. Oder man räumt ein, dass sprachliche Zeichen zwar (auch) durch ihre Bedeutung identifiziert werden müssen, dies aber auch im Fall eines ständigen Bedeutungs- bzw. Sprachenwechsels gewährleistet sei, womit eine Ausdrucksbedeutung im Extremfall bereits durch eine einmalige Verwendung instantiiert wäre. Was aber sollte dies unter Davidsonschen Voraussetzungen eigentlich heißen? Davidson deutet die bedeutungsangebenden W-Sätze ja als kausale Gesetzesaussagen. Diese ‚Gesetze‘ müssten aber, unter den genannten Voraussetzungen, in ihrer Reichweite auf Sprecher *und Zeitpunkte* beschränkt werden. Ich hatte aber bereits im Zusammenhang mit der Kritik am Funktionalismus darauf hingewiesen, dass es wenig Sinn macht, auch dann noch von Gesetzen reden zu wollen, wenn deren Anwendungsbereich derart spezifisch sein soll. Denn wenn wir akzeptieren würden, Gesetze so zu verstehen, dass sie in dieser Weise auf Einzelfälle zugeschnitten sein könnten, gäbe es keine Beschreibung einer singulären Verursachungsbeziehung, die nicht zugleich als gesetzesartig – und nicht nur als *unter* irgendwelche allgemeinen Gesetze subsumierbar – verstanden werden könnte. Und in welchem Sinne könnten wir dann noch eine Unterscheidung zwischen einem *allgemeinen* Gesetz und einer *einzelnen* kausalen Beziehung treffen? Welchen Sinn könnten wir der Allgemeinheit und der Kontrafaktizität des Gesetzesbegriffs geben?

Doch ganz unabhängig von diesen spezifischen Schwierigkeiten eines Davidsonschen Ansatzes lässt sich, wie ich meine, die Möglichkeit eines ständigen Bedeutungswechsels auch aus sinnkritischer Perspektive nicht wirklich *verständlich* machen. Denn auch hier gilt, dass Unterschiedlichkeit konzeptuell nur vor dem Hintergrund von Gleichheit (in relevanter Hinsicht) möglich ist. Eine jede Konstatierung von Unterschieden oder Verschiebungen bedarf der Unterstellung eines Referenzsystems des (in gewisser Hinsicht) Gleichen, welches erst den Maßstab für Verschiedenheit abgibt. Dies behauptet ja auch Davidson selbst im Hinblick auf die Möglichkeit von Überzeugungsunterschieden. Davidsons ‚Korrespondenzprinzip‘, als ein konstitutives (bzw. transzendentes) Prinzip der Möglichkeit von Verstehen, stellt ja eine Art von *Idealisierung* dar. Warum aber sollte Analoges nicht auch im Hinblick auf die Bedeutung sprachlicher Ausdrücke gelten? Ohne die *notwendige* Unterstellung von Bedeutungsgleichheit von sprachlichen Ausdrücken durch verschiedene Instantiierungen ihres Gebrauchs würde doch auch der Begriff der Bedeutungsverschiedenheit – und damit der Begriff der Bedeutung generell – seinen

Sinn verlieren³³⁷. Von einer Bedeutung reden zu wollen, die sich auf eine singuläre Ausdrucksverwendung eines Sprechers beschränken könnte, macht einfach keinen Sinn.

Wenn dies zutrifft, ist für die Möglichkeit von sprachlicher Bedeutung *und* Sprache eine gewisse Bedeutungsstabilität sprachlicher Zeichen über verschiedene Gelegenheiten ihrer Verwendung hinweg konzeptuell *notwendig*. Dies ist allerdings noch kein Einwand gegen die Möglichkeit einer Verständigung *ohne geteilten* Sprachgebrauch. Ich hatte oben schon gesagt, in welchem Sinne ich diese Behauptung für unproblematisch halte. Davidsons (und Bilgramis) Behauptung lautet jedoch, ‚Verständigung‘ könne auch dann gelingen, wenn die Beteiligten eine je eigene Sprache im Sinne einer *Primärsprache* sprechen. Und damit ist etwas sehr viel Weitergehendes gemeint, als die einzuräumende Möglichkeit, dass es nur einen einzigen Sprecher einer Sprache (im üblichen Sinne) gibt, etwa dann, wenn wir es mit dem letzten Sprecher einer aussterbenden Sprache zu tun haben. Auch könnte ein sprachlich bereits kompetenter Sprecher natürlich eine ‚Kunstsprache‘ entwickeln, die einfach deshalb niemand mit ihm teilt, weil niemand sonst sie erlernt hat. Die ist jedoch nicht dasjenige, um was es hier geht. Denn im ersten Fall wird ja durch den Begriff der ‚aussterbenden Sprache‘ bereits ein Verständnis von Sprache vorausgesetzt, welches nicht Davidsons sein kann. Und in letzterem Fall wäre eine solche, nur von einem Individuum beherrschte, Sprache offenbar eine Zweitsprache, die in einer systematischen Beziehung zur Primär- oder Muttersprache des betreffenden Sprechers stehen müsste³³⁸. Es muss hier also einerseits um die Frage gehen, ob – und in welchem Sinne – sprachliche Verständigung möglich sein könnte, wenn die Primärsprachen der sich Verständigenden individuell einzigartig wären³³⁹. Wichtiger erscheinen mir aber grundsätzlichere Fragen, insbesondere, ob wir auf der Grundlage der theoretischen Annahmen des Davidsonschen Programms – inklusive der vom explanativen Primat des Ideolektbegriffs – überhaupt zu einer Analyse von Verstehen, Verständigung, und Bedeutung gelangen können, die zum einen *der ‚kategorialen Grammatik‘ unserer entsprechenden Begrifflichkeiten und Redeweisen entspricht*, zum anderen *wesentlichen kognitiven Aspekten bzw. Funktionen propositional-sprachlicher*

³³⁷ Und dies gilt auch unter der Voraussetzung einer bedingten hermeneutischen Offenheit unserer begrifflichen Verständnisse (im oben genannten Sinne).

³³⁸ Vergl. Dummett (1991), Kap. 4.

³³⁹ Dass Davidson einräumt, dass wir in der normalen Verständigungspraxis nicht nur von einem ähnlichen Sprachverhalten unseres Gegenüber ausgehen, sondern auch eine Menge anderen Wissens hinsichtlich dessen heranziehen, was wir von einem spezifischen Sprecher zu erwarten haben, tut hier nicht zur Sache. Denn die Fragestellung ist ja eine prinzipielle, also nach den *notwendigen Bedingungen*.

Verständigung und der selbstbezüglichen Dynamik intersubjektiver Verständnisklärungsprozesse gerecht wird.

Die Maßgeblichkeit dieser grundsätzlichen Fragen wird bereits an den Schwierigkeiten eines jeden Versuchs deutlich, sozusagen ‚freistehenden‘ zu klären, ob Verständigung grundsätzlich auch dann *möglich* ist, wenn die beteiligten Sprecher radikal abweichende Idiolekte (im primärsprachlichen Sinne) sprächen. Denn eine Antwort hierauf hängt offenbar davon ab, was wir unter ‚Verstehen‘ und ‚Verständigung‘ verstehen und eine solche Klärung bleibt uns Davidson am Ende schuldig. Er beschreibt ja lediglich die *Form* und die *Bedingungen der empirischen Anwendbarkeit* einer Theorie, die uns, kennten wir sie, in die Lage versetzte, fremde Äußerungen zu verstehen. Er sagt nicht, *worin* ein Verstehen eigentlich *besteht* bzw. welche Art von Fähigkeiten eine solche Theorie beschreibt. Eine entsprechende Theorie soll ein Verstehen von Äußerungen ermöglichen, wenn sie die Bedeutung der Äußerungen anhand der Angabe ihrer Wahrheitsbedingungen (also *qua* der Theoreme) gemäß den Mitteilungsabsichten der Sprecher ‚ausspuckt‘. Damit legt sich Davidson, anders als Bilgrami, zwar nicht auf eine Verkürzung der Art fest, Verstehen von sprachlichen Äußerungen lediglich unter dem Aspekt kausaler Handlungserklärung anzugehen. Dennoch muss natürlich die Frage erlaubt sein, ob es wirklich zufriedenstellen kann, wenn am Ende nicht mehr gesagt wird, als dass man jemanden dann versteht, wenn man seinen Äußerungen in der Weise Wahrheitsbedingungen zuordnen kann, dass – salopp gesagt – die eigenen Überzeugungen weitestgehend in den Verhaltensweisen des anderen unterzubringen sind und die Zuordnungen aus einer Theorie ableitbar sind, die durch empirische Belege auf der Grundlage situationsbezogener Sprecheräußerungen bestätigt werden. – Heißt verstehen wirklich vorrangig, die ‚eigenen Wahrheiten‘ im Anderen wiederzufinden? Dass dies eine unangemessene Darstellung insbesondere im Hinblick auf ‚echte‘ Verständigungsprozesse ist, werde ich unten noch zu zeigen versuchen. Außerdem bleibt – selbst unabhängig davon, ob wir nicht mehr dazu sagen können (und müssen), worin das Haben wahrer Überzeugungen und das Finden einer Übereinstimmung (in den Überzeugungen) eigentlich besteht – ziemlich unklar, was wir mit den sicherlich zahlreichen ‚Meinungsunterschieden‘ bzw. solchen Fällen anfangen sollen, in denen sich ein Sprecher gemäß unserem Urteil irrt oder in denen er etwas glaubt, dass offenbar jenseits unseres Kenntnishorizontes und damit jenseits unserer Wahr-Falsch-Beurteilungsmöglichkeit liegt. Denn fürs Verstehen kommt es doch auch wesentlich darauf an, irgendeinen Zugang gerade auch zu jenen Überzeugungen oder Begriffen anderer zu bekommen, die man nicht teilt.

Mit den vorstehenden Bemerkungen sollte bereits immerhin ein gewisses Problembewusstsein hinsichtlich Davidsons Konzeption von Verstehen begründet worden sein, dass wir im Hinterkopf behalten müssen, wenn es darum geht, ob sprachliche Bedeutung durch einen Begriff des Verstehens und der Verständigung analysiert werden kann (bzw. muss), der ohne die Notwendigkeit eines Konzept von geteilter Sprache auskommt. Und ich meine, dass dies – selbst wenn wir uns mal auf die Annahme einlassen, Verständigung oder wechselseitiges Verstehen gehe im gegenseitigen Herausbekommen der Mitteilungsabsichten von Sprechern auf – nur dann eindeutig richtig wäre, wenn man den von Davidson abgelehnten ‚platonistischen‘ Sprachbegriff als maßgebliche Alternative unterstellt. Wenn wir also Davidson in der Zurückweisung eines Konzeptes von gemeinsamer Sprache im Sinne einer ‚präzisen und spezifizierbaren Menge syntaktischer und semantischer Regeln‘, denen Sprecher gemeinsam folgen, zustimmen, heißt dies nicht zugleich, *jede* mögliche Konzeption einer gemeinsamen Sprache für philosophisch irrelevant zu erklären. Denn dies würde nur dann folgen, wenn man sich der von Davidson (wie auch von Quine und in ähnlicher Weise auch von anderen Naturalisten) als ausschließlich unterstellten Alternative: ‚Platonismus oder tatsächliche Verhaltensregularitäten‘ anschliesse. Diese Alternative halte ich jedoch, wie ich in der Folge noch ausführlicher darlegen will, für eine falsche. Ihre Unterstellung resultiert aus einer *kategorial verfehlten*, nämlich *deskriptiven* Verortung der Konzepte ‚Bedeutung‘, ‚Verstehen‘ und ‚Sprache‘. Ich möchte demgegenüber für eine der zentralen Thesen dieser Untersuchung argumentieren, nämlich dass der Zusammenhang zwischen Bedeutung, Verständigung und gemeinsamem Sprachgebrauch in einer Weise anzulegen ist, bei der das Konzept einer gemeinsamen Sprache zwar durchaus als Idealisierung verstanden werden muss, jedoch als Idealisierung im normativ-pragmatischen Sinne einer (intersubjektive und objektive) Verantwortbarkeiten ermöglichenden, gemeinschaftlichen Orientierung an, als – im Rahmen der Sprachgemeinschaft – allgemein verbindlich unterstellten, Standards der richtigen Verwendung sprachlicher Zeichen. Vereinfacht ausgedrückt: Sprachliche und begriffliche Kompetenz muss als normative, nicht als deskriptive Fähigkeiten analysiert werden.

Auf der Grundlage einer normativ-pragmatistischen Konzeption von geteilter Sprache müsste die zur Debatte stehende Frage dann etwa lauten, ob eine Verständigung im vollwertigen Sinne – über alles und auf jedem Niveau – tatsächlich *ohne* einen, an irgendeinem Punkt einsetzenden, Rückgriff auf *wechselseitig als gleichbedeutend unterstellte* sprachliche Zeichen möglich ist? Ich möchte nicht ganz ausschließen, dass es bei einfachen Formen der Kommunikation, die lediglich ‚beobachtungsnahe Gegenstände‘ thematisieren, tatsächlich möglich wäre, auf jede

sprachliche Gemeinsamkeit zu verzichten – obschon ich zu der Ansicht neige, dass jede Art von Objektivitätsverständnis am Ende auf die *Möglichkeit* der situativen Herstellung einer Übereinkunft im Zeichengebrauch angewiesen ist³⁴⁰. Doch wie dem auch sei, es ist jedenfalls *kaum vorstellbar*, ein Gespräch über zeitgenössische Literaturkritik zu führen, wenn wir davon ausgingen, dass alle Sprecher sich an vollkommen voneinander abweichenden sprachlichen Normen im Sinne je eigener Primärsprachen orientierten. Ihre Äußerungen könnten sich dann ja nicht nur auf relativ ‚untheoretisch‘ beschreibbare Umstände beziehen, sondern auch auf andere sprachliche (etwa schriftliche) Äußerungen, von denen, gemäß Davidson, unterstellt werden müsste, dass sie ihrerseits in einer anderen ‚Sprache‘ formuliert sein könnten, die unter den genannten Bedingungen richtig interpretiert werden müsste. Und diese Äußerungen könnten sich wiederum auf andere Äußerungen beziehen, für die das gleiche gilt usw. Allerdings kann ich die Vermutung, dass Kommunikation ohne eine gemeinsame Sprache (im von mir gemeinten Sinne) hier *nicht* möglich wäre, durch kein einfach zu formulierendes allgemeines Kriterium dafür, woran sich erfolgreiche Verständigung *bemessen* lässt, präzisieren bzw. erhärten (wobei uns Davidson allerdings auch keines gegeben hat, das von den spezifischen Annahmen seiner naturalistischen Sprachphilosophie unabhängig wäre³⁴¹. Wir müssen deshalb, im Sinne der oben genannten grundsätzlicheren Fragen, etwas weiter ausholen.

³⁴⁰ Vergl. hierzu W. Kamlah/P. Lorenzen: *Logische Propädeutik* (1973).

³⁴¹ Davidsons Überbewertung des Faktums, dass wir auch Idiosynkrasien, absichtliche Wortverdrehungen (Malapropismen) oder Ähnliches häufig ohne Schwierigkeiten verstehen, führt ihn dazu, den *Sonderfall* zum *Normalfall* deklarieren (in diesem Sinne etwa auch I. Hacking: ‚Eine Parodie der Konversation‘ (1986/1990)). Ich denke allerdings, dass das Verstehen in solchen *Sonderfällen* häufig nur vor dem Hintergrund der Unterstellung geteilter *Standardbedeutungen* möglich ist. Jedenfalls scheinen die Beispiele, die Davidson in ‚Eine hübsche Unordnung von Epitaphen‘ für Malapropismen gibt, doch gerade dies zu zeigen, da der ganze *Witz* solcher Äußerungen ja offenbar in dem Wechselspiel von unterstellten Standardbedeutungen und abweichenden (‚unkonventionellen‘) Ausdrucksverwendungen besteht. Und wir *verstehen* einen Malapropismus, wie „Verlegenheit macht Diebe“, wenn wir nicht nur seine wörtliche Bedeutung kennen, sondern auch die *Pointe* erfassen, also wissen, was da (beabsichtigterweise) ‚schiefläuft‘ bzw. welche Ausdrücke *eigentlich* – anstelle der tatsächlich verwendeten – als für die Formulierung der einschlägigen Redewendung *angebracht* gewesen wären und worin der Unterschied in der Bedeutung besteht.

Der institutionelle Charakter sprachlicher Bedeutung und der normative Status sprachlicher Handlungen

Da man bei Auseinandersetzungen mit *substantiell unterschiedlichen* philosophischen Positionen meist nicht davon ausgehen kann, die Sache entweder empirisch oder durch konzeptuelle ‚*knock-down-Argumente*‘ eindeutig zu entscheiden, bleibt auch mir nichts anderes übrig, als die argumentative Auseinandersetzung vor allem auch im Hinblick auf die *Angemessenheit der phänomenalen Darstellung* und/oder den *explanativen Wert der Theorie* bzw. *Analyse* zu führen. In unserem Fall muss es im Wesentlichen um die Fragen gehen, ob man tatsächlich zu einem *angemessenen Verständnis* von sprachlichem Verstehen und von Verständigungsprozessen und den dafür notwendigen Fähigkeiten gelangen kann, wenn man letztere als einem rein deskriptiven Wissen entsprechend beschreibt und ob man die *wesentlichen Bedingungen* des Funktionierens unserer tatsächlichen, komplexen Verständigungspraxis und die Rolle, die der Bedeutungsbegriff in ihr spielt, *aufklären* kann, wenn man vom Idiolekt als Basiskonzept ausgeht. Um nun meine Ansicht zu begründen, dass wir dies abschlägig beantworten müssen, möchte ich im Folgenden zunächst den Aspekt des *institutionellen Charakters interner Zusammenhänge zwischen Zeichenbedeutung und Handlungstypen* ins Spiel bringen.

Zur Klärung der Frontlinien sollte jedoch zuerst nochmal das Strittige vom Unstrittigen geschieden werden. So ist es zweifellos richtig, dass unterschiedliche Sprecher einer Sprache (im üblichen Sinne) sich in ihrem *faktischen Sprachverhalten* immer nur mehr oder weniger *ähneln*. In *dieser* Hinsicht ist die Frage, wie viel an Gemeinsamkeit in der Art zu Sprechen vorhanden sein muss, damit zwei Sprecher noch als die gleiche Sprache sprechend gelten können, am Ende (in Grenzfällen) entweder gar nicht oder nur aus anderen, als rein sprachlichen Gründen zu beantworten³⁴².

Wie ich bereits geltend gemacht habe, kommt es aber für ein Verständnis von geteiltem Sprachgebrauch, welches mir im Hinblick auf Fragen des ‚*explanativen Vorranges*‘ als angemessen erscheint, gar nicht so sehr darauf an, dass Mitglieder einer Sprachgemeinschaft in ihrem *tatsächlichen Sprachverhalten* weitgehend und ständig übereinstimmen. Wir müssen nicht verlangen, dass Sprecher einer Sprache *tatsächlich* ein mehr oder minder identisches Vokabular zum Einsatz bringen, den exakt gleichen Grammatikregeln folgen, dieselben Formulierungen präferieren, mit ihren Wörtern *immer* dasselbe meinen, die Wörter ähnlich aussprechen u. s. w., um dem Begriff der

³⁴² Dies lässt sich etwa an einem Beispiel Chomskys verdeutlichen, wonach es nahe der niederländischen Grenze auf deutscher Seite Dialekte gibt, die dem Niederländischen ähnlicher sind, als manche Dialekte, die ‚*offiziell*‘, d. h. nach sozio-politischen Maßstäben, dem Deutschen zugehören.

Sprache einen wichtigen philosophischen Stellenwert einräumen zu können. Gegenüber der Realität sprachlicher Verständigungspraktiken, in der es weder eine scharfe, gleichsam *natürliche* Grenze dafür gibt, welcher Dialekt oder Idiolekt noch zu einer Sprache gehört und welcher nicht, noch dafür, wann ein Sprecher oder ein Lernender als kompetent gilt oder unter welchen Umständen eine Ausdrucksverwendung noch als richtig angesehen werden kann, haben Sprachen, Sätze, Wörter, Bedeutungen und Regeln tatsächlich den Charakter *idealer Konstrukte*. Damit meine ich jedoch nicht die Form von Idealisierung, die Davidson im Auge hat, also theoretische *Abstraktionen*, die eine bessere systematische *Beschreibung* empirischer Ereignisse ermöglichen sollen, sondern eine *normative* Idealisierung im Sinne einer *als geteilt oder teilbar unterstellten Orientierungsgrundlage für objektive Richtig-falsch-Beurteilungsmöglichkeiten sprachlichen Handelns*.

Diejenigen sprachlichen Normen oder Standards, die sich verbal in Formen von Definitionen oder Regelformulierungen explizieren lassen, finden ihren Niederschlag in offiziell gültigen Lexika, Grammatikbüchern usw. Solche Kodifizierungen sollen den Sprachgebrauch *normieren* und *nicht* etwa bloß *beschreiben*. Allerdings besteht ein wechselseitiger Anpassungsdruck zwischen der Normierung einer Praxis und der Anpassung der entsprechenden Normexplikationen an die tatsächlichen Entwicklungen dieser Praxis. Ähnliches gilt grundsätzlich, wenngleich in informellerer Weise, für Dialekte oder gruppenspezifische Sprachgebräuche. Auch regional-, berufs-, alters-, geschlechts-, milieu- oder sonst wie gruppenspezifische Redeweisen orientieren sich an normativen Mustern (so wie sich heute die Zugehörigkeit zu einem bestimmten Jugendlichen-Millieu etwa darin manifestiert, dass man „Kommt fett“ sagen muss, wenn man ungefähr sagen will, dass etwas toll oder beeindruckend ist).

Nun möchte zwar auch Davidson die Existenz solcher Phänomene nicht bestreiten. Er behauptet jedoch, sie seien für Bedeutung und Verständigung ohne Belang:

„These pressures are social and they are very real. They do not, however, as far as I can devine, have anything to do with meaning or communication. Using a word in a non-standard way out of ignorance may be a faux pas, but it is a faux pas in the same way that using the wrong fork at the dinner party is, and it has as little to do with communication as using the wrong fork has to do with nourishing oneself, giving that the ‚wrong‘ word is understood and the fork works.“
(Davidson 1994, S. 5)

Diese Bemerkung zeigt zunächst einmal mehr, dass Davidson dasjenige, was kommuniziert bzw. verstanden werden soll, also ‚Gedanken‘, als etwas Verständigungsprozessen Vorgängiges und von jeder sprachlichen Normierung Unabhängiges unterstellen zu können glaubt. Der vereinheitlichende ‚normative

Zwang‘ gemäß den Regeln und Standards einer Sprachgemeinschaft zu reden, hat für Davidson aber eine rationale Grundlage, nämlich als eine Maxime instrumenteller Vernunft:

„Of course I don‘t mean that there is no reason why we are thought, and we learn, to speak more or less as others around do. Nothing could be more obvious: we want to be understood and others have an interest in understanding us; ease of communication is vastly promoted by such sharing. (...) In the case of our children, or certain poets or writers, we must or do make exceptions, but in general our tolerance of strongly deviant idiolects is limited by clear practical considerations. None of this creates a free-standing obligation, however. Any obligation we owe conformity is contingent on the desire to be understood. If we can make ourselves understood while deviating from the social norm, any further obligation has nothing to do with meaning or successful communication.“
(Davidson a. a. O.)

Jede Verpflichtung gegenüber geteilten Normen der Sprachverwendung ist also Davidson zufolge in *kontingenterweise von unserem Wunsch abhängig, verstanden* zu werden. Doch lässt sich unsere Orientierung an bzw. ‚Verpflichtung‘ gegenüber einem gemeinsamen Sprachgebrauch wirklich auf ein *Gebot instrumenteller Rationalität* im Hinblick auf den Zweck, verstanden zu werden, reduzieren? Selbst wenn wir mal davon absehen, dass unklar geblieben ist, woran sich erfolgreiche Verständigung bemessen lassen soll, unterschlägt eine solche Argumentation, wie ich denke, insbesondere auch den *institutionellen Charakter* von Sprachen bzw. sprachlicher Bedeutung. Sprachen werden nämlich, besonders im Falle der jeweils gültigen Amtssprachen, vor allem auch durch einen institutionell konstituierten normativen Rahmen (in Form von normativ wirksamen Gepflogenheiten, Recht oder Sitte) in ihrem Geltungsbereich begrenzt, innerhalb dessen eine *Standardbedeutung* von Zeichen und Zeichenkombinationen *als allgemein gültig unterstellt* wird. Und ich behaupte, dass diese Tatsache für sprachliches Verstehen und Verständigung wesentlich ist.

Natürlich besteht der *Vorteil* einer intersubjektiven Standardisierung der Bedeutung sprachlicher Symbole für die Beteiligten auch darin, dass so eine gewisse allgemeine Verlässlichkeit hinsichtlich der zu erwartenden Konsequenzen eigenen und fremden sprachlichen Tuns im Rahmen sozialer Interaktionen gegeben ist, die von der Mühe der Deutung je individueller Fälle entlastet ist. Aber dieser instrumentelle Aspekt ist nicht der Punkt, auf den es mir ankommt. Was ich behaupten möchte ist, dass die Unterstellung gemeinsamer sprachlicher Standards für die Verantwortbarkeit bzw. den Handlungscharakter sprachlichen Tuns konstitutiv ist. Ich hatte ja bereits meine Ansicht dargelegt, dass sich sprachliche Handlungstypen einerseits durch die Bedingungen individuieren, unter denen Personen zu ihrer Ausführung oder

Unterlassung (und entsprechende Konsequenzen) berechtigt, angehalten oder gar verpflichtet sind, andererseits durch ihre ‚logischen‘ Beziehungen und die normativen Konsequenzen, also Berechtigungen und Verpflichtungen, die ihre Aktualisierung seitens des Sprechers oder gegenüber Angesprochenen generieren kann. Eine entsprechende Zuschreibung oder Übernahme von Verantwortlichkeit bedarf jedoch eines objektiven Maßstabs, der, so meine These, letztlich nur in einem allgemein verbindlichen Sprachgebrauch bestehen kann – also nicht etwa von *Hypothesen* einzelner über die Intentionen von sprachlich Handelnden abhängt.

Klare Beispielfälle für den institutionellen Charakter von Sprachen und sprachlicher Bedeutung sind solche, wo Aktualisierungen sprachlicher Handlungstypen direkt mit ‚externen‘ Sanktionsandrohungen belegt sind oder wo Handlungen einen normativen Status haben, der in der Folge ein bestimmtes Tun (oder Unterlassen) ‚einklagbar‘ macht. Wer beispielsweise einen deutschen Polizisten mit „Hey Bulle“ anspricht, kann dafür zur Kasse gebeten werden – ohne das die Frage, was er mit dem Ausdruck „Bulle“ (in ‚seinem Idiolekt‘) gemeint haben könnte, eine besondere Berücksichtigung fände. Man denke aber auch an *performative* Äußerungen (im Sinne Austins³⁴³), wie etwa das „Ich schwöre“ bei Vereidigungen (etwa vor Gericht) oder das von Braut und Bräutigam an den Standesbeamten gerichtete „Ja, ich will“ und die darauffolgende Deklaration „Hiermit erkläre ich euch zu Mann und Frau“ durch den Standesbeamten. In solchen Fällen vollzieht man überhaupt nur dann eine entsprechende Handlung (mit all ihren institutionell geregelten Konsequenzen), wenn man es ziemlich genau *so* sagt. Es ziemlich genau *so* zu sagen, ist für die Handlung *konstitutiv*. Eine stärkere Abweichung vom jeweiligen Ausdrucksstandard böte hier nicht Anlass zu kniffligen Interpretationsbemühungen, sondern könnte schließlich nicht als Vollzug einer entsprechenden Handlung gewertet werden. (Dies wird vielleicht noch deutlicher, wenn wir uns den institutionellen Charakter bestimmter Formen *schriftlicher* Sprachverwendung vor Augen führen, wie etwa „Betreten verboten“.)

Grundsätzlich können Aktualisierungen sprachlicher Handlungstypen – *qua* einer Aktualisierung sprachlicher Ausdruckstypen – Verantwortlichkeiten *im intersubjektiven Sinne* auch in solchen Kontexten sprachlicher Interaktion erzeugen, in denen diese keine *eindeutigen* institutionell geregelten Konsequenzen haben. Wer (im deutschsprachigen Raum) etwa sagt, „Ich vertraue Dir“, von dem erwarten wir zwar keine *genau* bestimmte Art des Verhaltens gegenüber der angesprochenen Person, aber doch zumindest, dass er/sie diese danach nicht heimlich kontrolliert. Auch für Konstatierungen (wie etwa „Es gab kein Bier mehr“) können wir im *intersubjektiven* Sinne verantwortlich sein, vor allem dann, wenn es um etwas geht, das für einen Hörer,

³⁴³ J. L. Austin: *How to do Things with Words* (1960).

der sich auf unser Urteil verlässt, unmittelbar handlungsrelevant ist. Und es ist, wie ich meine, für ein Verstehen solcher Sprechakte durchaus *wesentlich*, dass man *weiß*, dass entsprechende *Verantwortlichkeiten durch Aktualisierungen von Ausdruckstypen mit einer Standardbedeutung entstehen*, auf deren gemeinsame Verbindlichkeit wir uns im Zweifelsfalle berufen können (müssen).

Der andersgeartete diskursive oder *objektive* Sinn, in dem sprachlichen Handlungen ein normativer Status zukommt, lässt sich, wie bereits dargelegt wurde, nicht umstandslos im Sinne des präskriptiven Charakters moralischer, sittlicher oder rechtlicher Handlungsnormen erläutern. Denn zumindest in einem gewissen Kernbereich entsprechen einerseits die ans sprachliche Handeln und ans Denken anzulegenden ‚logischen‘ Richtigkeitsmaßstäbe nicht Sollens-Geboten, sondern eher konstitutiven Regeln der Vernunft. Und andererseits können auch über diesen Kernbereich hinausgehende, durchaus als Sollens-Standards verstehbare, ‚Gebote der Vernunft‘ (etwa nach Objektivität oder Angemessenheit) nicht einfach als Anspruch anderer Personen uns gegenüber gedeutet werden, schon deshalb, weil unsere Selbstansprüche hier das Maß an Vernünftigkeit übersteigen können, welches andere von uns berechtigterweise einfordern könnten. Zudem können solche Gebote, anders als intersubjektive Handlungsnormen, *in keinem Fall* als allein auf Grund ‚positiver‘ Setzung, als lokal und bloß kontingenter Weise verpflichtend (oder berechtigend) verstanden werden. Ein wichtiger Zusammenhang zwischen objektiver und intersubjektiver Verantwortbarkeit besteht allerdings darin, dass andere von uns grundsätzlich erwarten, dass wir nach ‚bestem Wissen und Gewissen‘ urteilen und die inferentiellen bzw. rechtfertigungsmäßigen Zusammenhänge zwischen Sprachakten und zwischen diesen und sonstigen Handlungen, aber auch die kausalen Konsequenzen von Handlungen ganz allgemein, überschauen bzw. anerkennen. Die intersubjektive Verantwortlichkeit für bestimmte Handlungen erstreckt sich nämlich gegebenenfalls auch auf die genannten Zusammenhänge. Diese Erwartung nicht zu einem gewissen Maß erfüllen zu können ist gleichbedeutend damit, nicht verantwortlich gemacht und damit, in einem gewissen Sinne, für voll genommen zu werden. Wir sind nur dann berechtigt, jemanden für seine Handlungen im intersubjektiven Sinne verantwortlich zu machen, wenn wir ihm ein gewisses Mindestmaß an Kompetenz sowohl hinsichtlich der Beurteilung der Berechtigung von sprachlichen und nicht-sprachlichen Handlungen, als auch hinsichtlich der Beurteilung der Konsequenzen zuschreiben können.

Dies will ich hier allerdings nicht nochmal vertiefen. Worauf es mir im Hinblick auf die Frage nach dem konzeptuellen und dem explanativen Stellenwert eines gemeinsamen Sprachgebrauchs ankommt, ist der *interne Zusammenhang zwischen der*

Bedeutung sprachlicher Äußerungen und ihrem durch intersubjektive und objektive Verantwortbarkeiten bestimmten Handlungscharakter. Zwar sind auch sprachliche Normen oder Standards keine Handlungsnormen im engeren Sinne. Es gibt keine unmittelbare und allgemeine Verpflichtung, sich an sprachliche Normen oder Standards zu halten, welche *in der gleichen Weise* zu verstehen wäre, wie die allgemeine Verpflichtung, sich an geltende Handlungsnormen zu halten. Da jedoch sprachliche Normen oder Standards, im Rahmen der Institution einer Sprachgemeinschaften, in komplexer, sozusagen ‚kompositionaler‘ und kontextrelativer Weise, Handlungstypen konstituieren, gibt es eine zumindest *mittelbare* Verpflichtung, sich an sprachliche Normen oder Standards zu halten, wenn entsprechende Handlungen präskriptiven Normen unterliegen.

Der institutionelle Charakter von Sprache und Bedeutung besteht also im Kern darin, dass sich die Mitglieder einer Sprachgemeinschaft an standardisierten Bedeutungen sprachlicher Ausdruckstypen orientieren müssen bzw. dies in sprachlichen Interaktionen (vorab) *wechselseitig unterstellen* können. Zumindest im Hinblick auf diesen Aspekt sprachlicher Bedeutung ist die Frage, *was* ein Sprecher mit seiner Äußerung *meint* und *wie* ein Angesprochener diesen *versteht*, eine sekundäre. Es ist die Unterstellung von Standardbedeutungen, die die Verantwortbarkeit von sprachlichem – und wohl auch nicht-sprachlichem – Handeln ermöglicht, weil nur so die normativen Verhältnisse, die durch Aktualisierungen von Handlungstypen generiert werden, für handelnde Subjekte und Betroffene *antizipierbar* und *intersubjektiv kontrollierbar* bzw. *bewertbar* sein können. Die Zuschreibung von Verantwortung für sprachliche Äußerungen, die auch eine gegenüber dem Adressaten der Zuschreibung bestehende Erwartung der Berücksichtigung von deren sprachlichen und nicht-sprachlichen Konsequenzen beinhaltet, ist überhaupt nur dann *sinnvoll* und *legitim*, wenn sie als spezifische von den Beteiligten kognitiv überschaut, also situationsübergreifend antizipiert und intersubjektiv beurteilt und kontrolliert werden kann. Dies scheint aber nur dann möglich, wenn sprachliche Ausdruckstypen in allgemein standardisierter Form an Handlungstypen gekoppelt sind, wobei die Zusammenhänge zwischen sprachlichen Ausdruckstypen (Zeichen) und Handlungstypen natürlich durchaus komplexer Natur sind, nämlich nicht nur ‚kompositional‘ und kontextrelativ, sondern auch eine gewisse Variationsbreite zulassend.

Das vorstehende Argument für einen explanativen Vorrang eines Konzeptes von Sprache oder des allgemeinen Sprachgebrauchs vor dem Idiolekte bzw. der Ausdrucksbedeutung vor der Sprecherbedeutung lässt sich folgendermaßen zusammenfassen: Ich gehe davon aus, dass wir die philosophisch relevanten

Grundlagen von Bedeutung und Verständigung nur dann angemessen verstehen können, wenn wir die Tatsache ernstnehmen, dass sprachliche Äußerungen – im Normalfall – ein freiwilliges absichtsvolles Tun, also Handlungen sind. Nun steht der Begriff der Handlung, wie ich argumentiert habe, in einem *internen Zusammenhang* mit dem der *Verantwortbarkeit* von Subjekten für ihr Tun. Dies gilt für den Handlungscharakter sprachlicher Äußerungen allerdings in zweierlei Weise, nämlich in Form einer Verantwortbarkeit *gegenüber der Vernunft* und einer Verantwortbarkeit *gegenüber Personen* oder *Institutionen*. Diese beiden Formen von Verantwortbarkeit stehen in einem komplexen, nicht-reduktiven Wechselverhältnis. Der entscheidende, gegen individualistisch-naturalistische Konzeptionen von Sprache und Bedeutung gerichtete, Zug der Überlegung ist der, dass Verantwortbarkeit *Zuschreibungsstandards* bedarf, die als – zumindest zu einem gewissen Grade – *situationsübergreifend bekannt und verlässlich* und von den *individuellen sprachlichen Handlungsintentionen der Subjekte* der Verantwortung und den *individuellen Einschätzungen von Zuschreibenden unabhängig* verstanden werden können. Die Rolle solcher Standards können nur *als allgemein verbindlich geltende Bedeutungsverständnisse* im Rahmen eines als gemeinsamem unterstellten Sprachgebrauch übernehmen, nämlich dadurch, dass objektiv typisierbare sprachliche Symbole über intersubjektiv verbindliche Verwendungsstandards (insbesondere in Form paradigmatischer Beispiele) in komplexer, aber systematischer Weise, normative Status von Handlungen konstituieren. Die für das Verstehen sprachlicher Äußerungen – zumindest in vielen Fällen – wesentlichen Zusammenhänge intersubjektiver Verantwortbarkeit sind nur unter der Voraussetzung von Standardbedeutungen in einer für alle Beteiligten oder Betroffenen praktikablen Weise antizipierbar, erkennbar und kontrollierbar; die Zuschreibung intersubjektiver Verantwortlichkeit nur dadurch *legitim*.

Auf den institutionellen Charakter von Sprachen bzw. auf den sozialen Charakter sprachlicher Bedeutung verwiesen ja auch jene Argumente, die von externalistischer Seite gegen ‚enge‘ Semantiken vorgebracht wurden. Dies gilt jedenfalls für den Aspekt externalistischer Überlegungen, der durch Putnams Begriff der ‚linguistischen Arbeitsteilung‘ und Burges Argumente für einen ‚sozialen Externalismus‘ markiert wird. Aus der oben vorgeschlagenen Variation von Burges Arthritisbeispiel ging ja hervor, dass es für den Apotheker überhaupt keine Rolle spielen muss, wie der vom Arzt geschickte Patient seine Äußerung „Ich benötige ein Mittel gegen Arthritis“ *meint* (bzw. was er über Arthritis glaubt). Selbst wenn dem Patienten bewusst ist, dass er gar nicht genau weiß, was er meint, muss ihn das nicht von der Ausführung eines entsprechenden Sprechaktes abhalten. Worauf es hier für eine erfolgreiche

Verständigung ankommt, ist, dass es eine dem Apotheker und dem Arzt gleichermaßen bekannte Standardbedeutung des Ausdrucks „Arthritis“ gibt.

Genauso könnte (nach einem Beispiel von Dummett (1991)) jemand, der von einem Mechaniker eine entsprechende Auskunft erhalten hat, einem Kollegen erklären, dass er deshalb zu spät zur Arbeit gekommen sei, weil die Dichtungsmanschette an seinem Wagen defekt ist, auch wenn er, anders als der Kollege, nicht die leiseste Ahnung hat, was eine Dichtungsmanschette ist. Selbst wenn man nicht genau *weiß*, was eine Dichtungsmanschette ist, kann man *sagen*, dass die Dichtungsmanschette kaputt ist, auch wenn man dann nicht *glauben* kann, die Dichtungsmanschette sei kaputt, sondern eventuell nur, dass ein Teil des Autos, welches man „Dichtungsmanschette“ nennt, kaputt sei. Die Verständigung würde auch hier gelingen, weil der Sprecher davon ausgehen kann, dass es eine Standardbedeutung des Ausdrucks „Dichtungsmanschette“ gibt, die der Hörer kennt.

Ein Sprecher verfügt in den genannten Fällen entweder nur über eine sehr eingeschränkte (wie im Beispiel mit der Dichtungsmanschette) oder über eine nur partiell richtige, zum Teil aber auch falsche Begriffs- oder Bedeutungskompetenz³⁴⁴ (wie im Arthritis-Beispiel). Insofern wäre man für solche sprachlichen Handlungen bloß im Sinne der korrekten Übermittlung von bzw. Berufung auf entsprechende(n) Aussagen verantwortlich, jedenfalls dann, wenn es keinen Grund gab, an der Kompetenz der Person, auf die sie zurückgeht, zu zweifeln. Jedenfalls würden wir auch dann, wenn Sprecher nur über eine (etwa gegenüber dem Wissen von *Experten*) unvollkommene Kompetenz verfügen, sagen, dass sie die Bedeutung kennen. Dies gilt etwa für Putnams bekanntes Gold-Beispiel, wo es genügt, wenn ich ein ungefähres Wissen der wesentlichen phänomenalen und kulturspezifischen Eigenschaften von Gold habe, also, in Putnams³⁴⁵ Worten, entsprechende *Stereotype* kenne, wie etwa ‚gelbliches Metall‘, ‚schwer‘, ‚wertvoll‘, ‚häufig für Schmuck verwendet‘, – auch wenn ich hinsichtlich möglicher Entscheidungen darüber, ob etwas Gold ist, leicht in die Irre zu führen wäre und gegebenenfalls auf das Wissen von Experten zurückgreifen müsste. Ein noch drastischeres Beispiel ist das Wort „Temperatur“, da hier zwar ein Alltagsgebrauch besteht, „Temperatur“ jedoch gleichzeitig auch ein technischer Term der Physik ist. Und obwohl die entsprechenden physikalischen Kenntnisse das übliche, vortheoretische Wissen über den Gebrauch des Ausdrucks „Temperatur“ weit übersteigen, würden wir dennoch sagen, es handle sich um dasselbe Wort und dasselbe Phänomen. Vielleicht kann man ganz grundsätzlich sagen, dass sprachlichen

³⁴⁴ Dies geht natürlich nicht notwendig zusammen. Jemand kann über eine begriffliche Kompetenz verfügen, aber der Bedeutung eines bestimmten Wortes, das diesen Begriff zum Ausdruck bringt, nicht mächtig sein, weil er die entsprechende Sprache nicht beherrscht.

³⁴⁵ Putnam (1975).

Kompetenz auch in dem Wissen besteht, wie man im Bedarfsfalle Wissenslücken schließen oder ausgleichen könnte – also etwa durch Befragung eines Experten. Jedenfalls ist diese, in der alltäglichen Kommunikation nicht unübliche, Praxis der *partiellen Delegierung von Bedeutungskompetenz* nicht leicht auf der Grundlage eines idiolektischen Sprachbegriffs verständlich zu machen³⁴⁶. Davidson sagt, er sei von solchen Argumenten nicht beeindruckt³⁴⁷. Nun, ich bin meinerseits nicht davon beeindruckt, dass Davidson unbeeindruckt bleibt.

Noch offensichtlicher wird der institutionelle Charakter von Sprache und Bedeutung, wie Dummett (der vom ‚sozialen‘ Charakter redet) verdeutlicht, im Fall der Ortsnamen:

„Place names are the best example. The employment of a word of this kind rests on a complex of social practices. It depends, primarily, on our ability to get to the place it names and to know when we have arrived; and this is embedded in the practices of making and reading maps and the operations of our various systems of transport. Knowing where one is has to do with recognising landmarks, with being able to read roadsigns or names of railwaystations, but also, where there are people living, with where they say one is. This involves a system of established correspondences between names in different languages (...).“ (Dummett 1991, S. 85 f.)

Kompetent in Bezug auf Ortsnamen zu sein heißt also unter anderem, ungefähr zu wissen, wie man zu den jeweils bezeichneten Orten gelangen könnte. Dabei ‚lebt‘ die Institution der Ortsnamen (bzw. der Namen überhaupt, aber auch der Zeitmaße u. ä. m.) geradezu von der Unterstellung gemeinsamer Standardbedeutungen, ja sogar von Standardübersetzungen zwischen Sprachen. Grundsätzlich scheint es in Bezug auf Eigennamen doch geradezu absurd, von individuellen ‚Sprecherbedeutungen‘ ausgehen zu wollen.

³⁴⁶ Es ist naheliegend, dass mit der durch die partielle Delegierbarkeit begrifflicher Kompetenz gegebenen Graduierung auch eine Graduierung von Verantwortbarkeit einhergeht. Wir relativieren die Verantwortlichkeit, die durch die Verwendung eines entsprechenden Wortes entsteht, auf die unterstellten Kenntnisse eines Sprechers hin. So sind wir im Falle der Falsifikation einer Aussage „Dies ist Gold“ gegenüber einem Juwelier weniger nachsichtig, als gegenüber Tante Hertha.

³⁴⁷ Davidson (1994a), S. 6. Er argumentiert dort, dass es zur Erklärung solcher Phänomene nur darauf ankäme, dass Sprecher *glauben*, dass es Experten gibt, *nicht*, dass es sie *wirklich gibt*. Dies leuchtet mir nicht ein. Denn was ihre Überzeugungen ‚wahr macht‘, ist die *Existenz* von Experten.

5 Bedeutungsverstehen, Wahrheitsfähigkeit und Propositionalstruktur

Ziel der bisherigen Überlegungen war es vor allem zu zeigen, dass es für das, was jemand sagt bzw. was die von ihm geäußerten Worte bedeuten, primär weder auf das *individuelle* kausale Zustandekommen der Überzeugungen eines Sprechers ankommt, noch primär darauf, was seine Äußerungen *seiner Ansicht nach* bedeuten – obwohl solche Erwägungen natürlich eine sekundäre Rolle spielen können. Eine eher grundsätzliche Abwägung zwischen naturalistischen Positionen einerseits und einer hier verteidigten normativ-pragmatistischen Herangehensweise kann auch dadurch unterstützt werden, dass wir sowohl deren jeweiligen explanativen Ertrag, als auch deren jeweilige ‚phänomenologische Angemessenheit‘ gegeneinander abwägen. Ich möchte deshalb in diesem abschließenden Teil noch einige konstruktive Überlegungen zum Thema Bedeutung, Verständigung und Geist anstellen, die durch kritische Bezüge zu den hier diskutierten naturalistischer Theorien ergänzt werden.

Einige der problematischen Konsequenzen von naturalistischen Theorien, die Gehalte über irgendwelche Formen kausaler Beziehungen zu Ereignissen in der Welt individuieren oder erklären wollen, hatten wir ja bereits in der Auseinandersetzung mit den jeweiligen Positionen diskutiert. So hatte ich im Rahmen der Kritik an den korrespondenztheoretischen Annahmen Millikans unter anderem eingewendet, dass die ganze Vorstellung des Geistes als eines *Organs*, eine dessen wesentlicher Funktion in der richtigen Abbildung der Tatsachen bestünde, die durch unseren jeweiligen Sprachgebrauch bedingte ‚Gemachtheit von Tatsachen‘ – im Sinne ihrer Möglichkeit – unterschlägt. Es ist nicht nur der konventionelle Charakter sprachlicher Unterscheidungen, es ist die Verwobenheit sprachlicher Praktiken mit nicht-sprachlichen Praktiken (Techniken, Bräuchen u. ä. m.), mit Handlungsnormen (sittlichen, moralischen oder positiv-rechtlichen) und anderen ‚sinnstiftenden‘ Gegebenheiten (der evolutionsbedingten Kontingenz unserer wahrnehmungsmäßigen und kognitiven Möglichkeiten, unseren Gefühlsdispositionen, kulturell überlieferten Deutungsmustern u.a.) unseres Daseins, der unsere begrifflichen *Verständnisse* – und damit mögliche Tatsachen – wesentlich bedingt. Die entsprechenden Praxis- und Relevanzbeziehungen unseres Sprachgebrauchs *erschließen* erst den *substantiellen Sinn* einer Welt möglicher Tatsachen, wie wir im Anschluss an einen Term Heideggers sagen können³⁴⁸. Was wir beispielsweise unter einer Tatsachenbehauptung verstehen können, dass etwas ein Hammer ist, hängt vom *Sinn* her von einem verstehenden

³⁴⁸ M Heidegger: *Sein und Zeit* (1926), insbes. § 44.

Zugang zur Funktion des Hammers bzw. der Praxis des Hämmerns ab, *extensional* aber auch davon, *was wir* noch als einen Hammer zählen – womit ich natürlich nicht sagen will, wir hätten damit bereits den Begriff der Tatsache bzw. die Wahrheitsfähigkeit sprachlicher Äußerungen ‚eingeholt‘ (dazu komme ich noch).

Es trägt meiner Ansicht nach auch nichts Wesentliches zum sozusagen ‚formalen‘ Verständnis von ‚Geist‘ (im Sinne von Begrifflichkeit und Propositionalität – bzw. beurteilbaren Gehalten) bei, wenn wir von individuellen Realisierungen propositionaler Einstellungen ausgehen und dann Theorien aufstellen, die zeigen sollen, wie deren Gehalt etwa durch ihre –in der ein oder anderen Weise zu spezifizierenden – Ursachen bzw. ihre normalen Verursachungsbedingungen oder ihre *Funktion*, die Tatsachen richtig abzubilden, bestimmt werde. Nach der in dieser Arbeit vertretene Sicht der Dinge kommen wir nur dann solchen Begriffe wie ‚Gehalt‘, ‚Tatsache‘ oder ‚Wahrheit‘ auf die Schliche, wenn wir danach fragen, *was wir eigentlich tun* bzw. *mindestens wissen und tun können müssen*, wenn wir eine Aussage machen (eine Behauptung aufstellen) oder etwas als wahr oder als falsch beurteilen. Mit anderen Worten: ‚Geist‘(-igkeit) können wir philosophisch nur dann in angemessener Weise aufklären, wenn wir eine *rational-rekonstruktive Perspektive* auf unsere sprachliche Praxis und unser Vermögen, an ihr zu partizipieren, einnehmen. Und aus einer solchen Perspektive geht es nicht um eine Erklärung von mentalen Zuständen oder Ereignissen durch bestimmte Formen der Beziehungen zu Tatsachen (Zuständen Ereignissen), ja es geht dabei eigentlich gar nicht um die *Beschreibung von Faktischem* (etwa im Sinne der Frage, unter welchen Bedingungen ein physischer Zustand $Z(p)$ zugleich ein geistiger Zustand $Z(g)$ ist). Es geht, wie ich im Folgenden noch weiter plausibel machen will, um ein angemessenes Verständnis der *normativen Struktur* unserer Praxis assertorischen, imperativischer oder volitiver Sprachgebrauchs – wobei ich mich, aus bereits genannten Gründen, auf die assertorische Rede konzentrieren werde. Erst auf der Grundlage einer entsprechend verfassten sprachlichen Praxis erhält, so behaupte ich, die Zuschreibung propositionaler Einstellungen ihren Sinn – allerdings nicht den der *Beschreibung* von Zuständen oder Ereignissen.

Eine *rational-rekonstruktive Perspektive* verstehe ich als eine *normativ-pragmatistische*, wenn die Beschreibung dessen, *was wir können oder tun müssen*, damit diese Praxis möglich ist bzw. wir an ihr erfolgreich teilnehmen können, in einer Begrifflichkeit von ‚Orientierungen‘, ‚Standards‘, ‚Beurteilungen‘ oder ‚Stellungnahmen‘ vorgenommen wird. Aus einer solchen Perspektive lässt sich, wie ich denke, *mehr* über Verständigung oder die ‚Übermitteln von Gedanken‘ sagen, als eine rekursive Wahrheitstheorie für sprachliche Äußerungen im Sinne Davidsons hergibt. Davidsons ‚einheitliche Theorie der Bedeutung und des Handelns‘ setzt zwar

Sprachlichkeit als Bedingung von Geist methodisch voraus; doch er bleibt dabei an einer wesentlich *deskriptiv* verstandenen Kompetenz von Interpreten orientiert. Die augenscheinliche *Trivialität*, die Davidsons W-Sätzen besonders in Fällen anhaftet, in denen sie die Bedeutung der Äußerungen eines Sprechers *innerhalb* einer Sprache angeben, also die ‚Objektsprache‘ und die Metasprache dieselbe sind, ist dabei kein Zufall: Ein davidsonscher Interpret bewegt sich nämlich schon auf der Grundlage dessen, was eine philosophische Theorie der Bedeutung und des Verstehens gerade aufzuklären hat. Wenn wir mehr darüber erfahren wollen, *worin* die Bedeutung von Worten oder Äußerungen besteht oder was eine Bedeutungskompetenz ausmacht, die uns überhaupt erst in die Lage versetzt, (wahre) Bikonditionale für Sprecher aufzustellen oder uns zu verständigen, scheint es sinnvoller, sich an Wittgensteins Formulierung zu orientieren, demgemäß Bedeutung das ist, was die Erklärung der Bedeutung erklärt³⁴⁹. (Ob man so noch eine systematische Theorie des Sprachverstehens zustande bringen kann, wie nicht nur Davidson im Auge hat, sondern – unter der Maßgabe eines epistemisch gedeuteten Wahrheits- bzw. Bedeutungsbegriffs – etwa auch Dummett, ist allerdings eine andere Frage. Eine zugleich *substantielle und systematische Theorie* des Verstehens scheint mir jedenfalls aus Gründen, die wir unter anderem durch Wittgenstein kennen, unmöglich³⁵⁰).

Ich schlage also vor, unter der Bedeutung eines sprachlichen Ausdruckes einfach erst mal das verstehen, was wir jemandem erklären müssten, *der den entsprechenden Begriff noch gar nicht beherrscht* bzw. über einen äquivalenten Ausdruck in seiner Muttersprache nicht verfügt. Damit mochte ich natürlich nicht behaupten, wir könnten jemanden durch entsprechende Erklärungen Wort für Wort eine Primärsprache lehren. Es geht hier um Fragen der rationalen Rekonstruktion. Aber es scheint klar, dass solche Erklärungen nicht zur Gänze rein verbalen Charakters sein können, sondern darüber hinaus eines unmittelbaren Welt- oder Praxisbezugs bedürfen, etwa indem auf etwas gezeigt wird oder indem *vorgeführt* wird, wie etwas zu tun ist. Wem ich den Begriff des Hammers erläutern will, dem zeige ich ein entsprechendes Werkzeug und sage etwa „Dies ist ein Hammer. Damit kann man Nägel in die Wand schlagen“. Eine

³⁴⁹ L. Wittgenstein: *Philosophische Grammatik*, § 23.

³⁵⁰ McDowells Kritik an Dummett, der es für möglich hält, eine *Bedeutungstheorie* einer Sprache „in the full blooded sense“ aufzustellen, scheint mir, jedenfalls in diesem Sinne, berechtigt (J. McDowell: ‚In Defence of Modesty‘ (1987)). Das heißt allerdings nicht, dass, wie McDowell offensichtlich folgert, jede Theorie der Bedeutung, die sowohl Behaviorismus, als auch Psychologismus vermeiden will, auf der Ebene der Angabe von Wahrheitsbedingungen stehen bleiben muss, da sie sonst Gefahr liefe, Gehalt „as from the outside of content“ beschreiben zu wollen. Dass es keine systematische Bedeutungstheorie ‚as from the outside of content‘ geben kann, heißt nicht, dass sich jede philosophische Reflexion über den Bedeutungsbegriff an einem nicht weiter zu hinterfragenden Begriff von Wahrheitsbedingungen zu orientieren hat.

praktische Ergänzung einer solchen Erklärung würde dabei darin bestehen, dass man zeigt, wie es gemacht wird. Das Konzept der *Bedeutungserklärung* (im engeren, verbalen Sinne) hat ohnehin darin seine Grenze, dass nicht alle Begriffe durch sprachliche Erklärung eingeführt werden können. Dies gilt natürlich auch in lernpsychologischer Hinsicht: Bei Kleinkindern, die noch über gar keine sprachliche Verstehenskompetenz verfügen, muss es ja vor jeder *Erklärung* so etwas wie ein ‚Abrichten‘ bzw. irgendeine Form von repetitiver Lernphase im Hinblick auf die Einführung bzw. den Erwerb eines Basisvokabulars und einer Basisgrammatik anhand von Standardbeispielen geben, damit spätere Erklärungen überhaupt eine Grundlage haben. Auch die Funktionsweise der Propositionalstruktur und der illokutiven Kraft von Äußerungen kann demjenigen nicht einfach *erklärt* werden, der über sprachliche Kompetenz nicht zumindest rudimentär verfügt. Ihrer rationalen Rekonstruktion mit sprachlichen Mitteln steht dagegen nichts im Wege.

Interessante Überlegungen im Hinblick auf eine Rekonstruktion des ‚Aufbaus‘ der unterschiedlichen semantischen und syntaktischen Kategorien während der kindlichen Sprachentwicklung hat Quine – unter den Prämissen seines gleichsam ‚naturalisierten Empirismus‘ – in *Die Wurzeln der Referenz* (1974/1976) angestellt. Quine will zeigen, wie sich auf der dünnen, behavioristischen Grundlage von angeborenen Mechanismen zur Gruppierung ähnlicher Reize, einer äußerlichen ‚Verstärkung‘ dieser Mechanismen und der (unter behavioristischen Voraussetzungen nicht ganz unproblematischen) Inanspruchnahme hinweisender Erklärungen die Möglichkeit der Bildung der unterschiedlichen semantischen Kategorien, also ‚normaler‘ genereller Termini (z.B. ‚ist ein Hund‘), Stofftermini (z. B. ‚ist Wasser‘; ‚ist gelb‘) und des Individuations- und Referenzapparates, verständlich machen lässt. Und weil für Quine das Bezug nehmen auf Gegenstände im Mittelpunkt unseres wissenschaftlichen Weltbildes steht, hält er es für wünschenswert, insbesondere dessen Wurzeln im Spracherwerb aufzuspüren. Es geht für Quine also darum zu zeigen, wie man, ausgehend vom naturalistischen Substitut für Wahrnehmungen oder Empfindungen, nämlich (‚Oberflächen-‘) Sinnesreizungen, zum kognitiven Aufbau einer Welt von Körpern und Dingen und damit schließlich zur Möglichkeit abstrakter wissenschaftlicher Theoriebildung gelangen kann³⁵¹.

Wie wir bereits sahen, besteht für Quine die *sprachliche* Basis in – als subsententiell unstrukturiert vorgestellten – Beobachtungssätzen (als dem ‚Eingangstor zur Sprache‘). Er sieht „bodies materializing, ontologically speaking: as ideal nodes at the foci of intersecting observation sentences“ (1992, S. 24). Als *rationales Motiv* für die

³⁵¹ Man könnte fast sagen, Quine wolle hier Carnaps Programm eines ‚logischen Aufbaus der Welt‘ (Carnap 1928/1961) unter naturalistischen Vorzeichen durchführen.

Postulierung einer ‚Dingwelt‘ sieht Quine die dadurch ermöglichte Systematisierbarkeit und Prognostizierbarkeit von Sinnesreizungen. Ein entsprechender ‚kognitiver Vorteil‘ bleibt jedoch grundsätzlich auf die singuläre Perspektive eines erkennenden Subjektes beschränkt. Dabei entgeht Quine, wie ich zeigen will, der wesentliche, *als intersubjektiv angelegt zu analysierende, kognitive und praktische Kern der Propositionalstruktur*.

Allerdings können wir Quine darin zustimmen, dass man sich bei dem Versuch, eine Kompetenz angemessen zu verstehen, nicht auf die Frage beschränken darf, wie man diese tatsächlich erwirbt. Eine andersgeartete, rationale Rekonstruktion tut dabei aber, wie ich meine, auch gut daran, zu fragen, welche *neuen kognitiven und praktischen Möglichkeiten* sich durch diese Kompetenz bzw. durch eine entsprechend komplex konstituierte Praxis gegenüber niedrigstufigeren Formen von Fähigkeiten bzw. Praktiken der gleichen Art ergeben. Das heißt für den uns interessierenden Fall, eine rationale Rekonstruktion der wesentlichen Bedingungen unserer sprachlichen Praxis im Hinblick darauf vorzunehmen, dass deren *möglichkeitserweiternden kognitiven Struktureigenschaften* deutlich werden. Durch ein solches Vorgehen wird es, so denke ich, einfacher, sich darüber auseinanderzusetzen, welche Form oder Ebene der Begrifflichkeit für eine angemessene Rekonstruktion mindestens zum Einsatz zu bringen ist. Denn angemessen ist sie dann, wenn sie nicht wesentliche kognitive und praktische Möglichkeiten unserer sprachlichen Praxis entweder implizit lässt oder schlichtweg unterschlägt.

Nun scheint es für solche rekonstruktiven Bemühungen naheliegend, bei der Untersuchung der Voraussetzungen und Eigenschaften einfacherer Formen von Kommunikation oder ‚Sprache‘ anzusetzen. Solche offensichtlich einfacheren Formen ‚sprachlich‘ vermittelter Verhaltenskoordinierung sind ja bekanntermaßen bereits auf der Stufe tierischer Gemeinschaften, etwa bei Bienen oder Affen, vorhanden. Es geht hier jedoch nicht eigentlich um die Frage, welches Niveau von Kommunikation bei bestimmten nicht-menschlichen Tieren *tatsächlich* erreicht wird. Worauf es ankommt ist, ob die Wahl einer bestimmten Begrifflichkeit, die möglicherweise für die Beschreibung oder Analyse bestimmter tierischer Kommunikationsformen hinreicht, auch dem kognitiven und praktischen Potential unserer sprachlichen Praxis angemessen sein kann. Naturalistische Philosophen sind nämlich – wie in dieser Arbeit deutlich geworden sein sollte – häufig geneigt, eine Beschreibungsform zu wählen, die das Verhältnis zwischen den bekannten tierischen und den menschlichen Kommunikationsformen als eines der *graduellen* Komplexitätssteigerung begreifbar machen soll. Aus naturalismus-kritischer Perspektive wäre es daher instruktiv, (mutmaßliche) Defizienzen von Kommunikations- oder Sprachformen, deren Potential

tatsächlich (noch) in einer Form von naturalistischer Begrifflichkeit beschreibbar zu sein scheint, gegenüber jenem Niveau sprachlicher Verständigung herauszuarbeiten, welches unsere sprachliche Praxis auszeichnet³⁵².

Dies hat insbesondere Ernst Tugendhat in einer, wie ich meine, besonders aufschlussreichen Weise getan³⁵³. Ein wichtiger Schritt in seiner Rekonstruktion ist ein Gedankenexperiment, bei dem er ein Niveau von sprachlicher Kommunikation analysiert, welches er mit dem Begriff ‚*Quasiprädikatensprachen*‘ kennzeichnet. Von diesem aus nähert er sich dann schrittweise dem kommunikativen Niveau unserer Sprache und Verständigungspraxis an, wobei die ursprüngliche Beschreibung schrittweise angereichert wird. Mit dem Term ‚*Quasiprädikatensprache*‘ will Tugendhat die formalen Eigenschaften einer Sprache benennen, deren Bedeutungen über Konditionalregeln der Form: „Immer wenn *X* (Bedingung *B* gegeben ist), dann *Y* (wird Ausdruck *A* verwendet)“ bestimmbar sind. Um keine präjudizierenden Vorannahmen zu machen, können wir es offenhalten, ob solche Sprachen in kausalen Begriffen analysiert werden können. Jedenfalls lassen sich auf diesem Niveau möglicherweise frühkindliche Äußerungen beschreiben, etwa das „Wauwau“, wenn ein Kind einen Hund sieht. Das Eigentümliche solcher Sprachen besteht nun darin, dass sie vollkommen *situationsabhängig* sind. Was damit gemeint ist, werde ich gleich noch genauer erläutern. Darüber hinaus geht solchen Äußerungsformen aber auch noch eine kommunikative Funktion bzw. ein adressierender Charakter ab.

Wir können uns zunächst auf jenes letztere Defizit beschränken und es zu beheben versuchen, indem wir mögliche Reaktionen eines Hörers solcher Äußerungen in der Analyse berücksichtigen. Wir könnten hier etwa an Wittgensteins Konzeption einer primitiven Form von Sprache denken, wie sie in den *Philosophischen Untersuchungen* (§ 2 ff.) anhand des bekannten ‚Baustellenbeispiels‘ beschrieben wird. Für eine solche Form der Kommunikation soll nicht nur gelten, dass ein Sprecher einen Ausdruckstyp (ein Zeichen) unter bestimmten Umständen verwendet, sondern auch, dass ein Hörer, an den sich die Äußerung richtet, darauf mit einer bestimmten Tätigkeit reagiert. Auch die Bedeutung eines Ausdruckstyps bzw. einer Äußerung *für den Hörer* ließe sich dann in Form einer Konditionalregel beschreiben, nur dass der Anlass keine nicht-sprachliche Situation, sondern die Äußerung ist. In Wittgensteins Beispiel benötigt ein Sprecher *S* einen bestimmten Baustein, doch statt selbst die passende ‚Handlung‘ auszuführen, vollzieht er eine ‚Ersatzhandlung‘, nämlich eine Äußerung, die ihrerseits einen Hörer *H* dazu veranlasst, jene ‚Handlung‘ auszuführen. Der Begriff ‚veranlassen‘ soll hier wieder eine gewisse Neutralität zwischen normativ-begründungsmäßiger und

³⁵² Diese Strategie verfolgt etwa auch J. Bennett in: *Rationalität - Versuch einer Analyse* (1964/1967), sehr ausführlich am Beispiel der Bienensprache.

³⁵³ Tugendhat (1976); s. bes. Kap. 12 u. 13.

kausalistischer Beschreibungsweise zum Ausdruck bringen, die bis hierher auch noch als durchaus möglich erscheint. Naturalistische Theorien wählen hier natürlich eine Form von kausalistischer oder ‚subsumtionstheoretischer‘ Beschreibung. Dafür stehen insbesondere auch jene Analysen von Bedeutung, wie sie in klassischen Varianten des linguistischen Behaviorismus im Stile Bloomfields³⁵⁴ vertreten wurden. Solche Theorien analysieren sprachliche Bedeutungen grundsätzlich auf der Ebene von *Signalsprachen*. Die kommunikative Funktion von Signalen besteht – in behavioristischer Terminologie – darin, Reaktionen auf ‚Stellvertreterreize‘ hervorzurufen. So wäre etwa der Ausruf „Feuer!“ signalsprachlich zu deuten, wenn er lediglich dazu bestimmt wäre, unmittelbar die gleiche Reaktion hervorzurufen, wie ein Feuer.

Nun entsteht, wie Tugendhat zeigt, ein erstes Problem für signalsprachlich bzw. quasiprädicativ ansetzende Theorien dadurch, dass die Bedeutung einer Äußerung für ‚Emittenten‘ und ‚Rezipienten‘ verschieden ist, da die Umstände, die den Sprecher zur Äußerung veranlassen, ja nicht mit den Reaktionen übereinstimmen, die die Äußerung beim Hörer hervorrufen soll. Dies entspricht jedoch nicht unserer vorthoretischen Intuition, nach der die Bedeutung sprachlicher Zeichen nicht grundsätzlich in eine Sprecher- und eine Hörerfunktion zerfällt³⁵⁵. Eine weitere, allgemeine Schwierigkeit solcher Ansätze besteht darin, dass sie offensichtlich der Differenz von imperativischem und assertorischem Sprachgebrauch nicht gerecht werden können. Eine Äußerung, wie „Feuer!“, kann signalsprachlich sowohl ‚quasiindikativisch‘ als auch ‚quasiimperativisch‘ gedeutet werden, also etwa sowohl der Übermittlung von Informationen, als auch der Aufforderung zur Flucht dienend. Entsprechendes gilt beispielsweise für die Interpretationsmöglichkeiten von Bientänzen. Denn auch die Reaktionen von Bienen auf die Tänze ihrer Artgenossen können mit gleichem Recht sowohl im Sinne der ‚Befolgung einer Aufforderung‘, als auch im Sinne der Umsetzung einer Information gedeutet werden. Die Unterscheidung von indikativischem und imperativischem Modus scheint aber nicht nur die Grenzen von Quasiprädicaten- oder Signalsprachen zu sprengen, sondern den Rahmen

³⁵⁴ L. Bloomfield: *Language* (1935).

³⁵⁵ G. H. Mead hat in *Geist, Selbst und Identität* (1934/1968) auf der Grundlage eines behavioristischen Vokabulars u.a. versucht, diese Schwäche des klassischen linguistischen Behaviorismus durch interessante Überlegungen, die er um Begriffe wie ‚Reizverzögerung‘, ‚gleichzeitige Selbststimulation‘, ‚virtuellen Mitvollzug‘ und ‚ideale Rollenübernahme‘ entwickelt, zu überwinden. Seine Analyse oszilliert jedoch zwischen einer eher behavioristischen und einer eher normativ-idealisierenden Betrachtungsweise und ist im Hinblick auf den Zusammenhang von Bedeutung und Handlung programmatisch geblieben. Eine angemessene Diskussion der Überlegungen Meads würde aber wohl ein weiteres, eigenständiges Kapitel erfordern, worauf ich hier verzichten muss.

behavioristischer Analysen überhaupt, da es offenbar keine Möglichkeit gibt, diese Unterscheidung im Rahmen einer Rede von *Reaktionsweisen* verständlich zu machen.

Die Schwierigkeiten signalsprachlich oder quasiprädikativ ansetzender Analysen verschärfen sich allerdings noch, wenn man berücksichtigt, dass diese nicht über eine *situationsgebundene* Form der Bedeutung hinauskommen können. Die Bedeutung von Signalen im Sinne des klassischen Behaviorismus bleibt ja daran gebunden, *unmittelbar eine bestimmte Reaktion in der jeweils gegebenen Situation hervorzurufen*. Dies aber hat zur Folge, dass wir eigentlich keinen Grundlage dafür hätten, einen Ausruf, wie „Feuer!“ (,quasi-indikativisch‘ gedeutet) als „Es brennt“, bzw. – was dem gleich kommt –, als „Es brennt jetzt hier“ oder gar in der Form „X ist brennend“ zu interpretieren. Damit solche Interpretationen angemessen sein können, müssten wir Gründe haben, der betreffenden Kommunikationsform das *sprachliche Potential zu einer situationsunabhängigen Verständigung*, also einer *intersubjektiven Identifizierbarkeit und Reidentifizierbarkeit* einer Situation oder eines Dinges, die bzw. das wir durch eine Eigenschaft (wie die des Brennens) charakterisieren, zu unterstellen. Für eine solche Unterstellung aber bieten Analysen auf der Ebene von Signalsprachen keinen Anlass, weil über Differenzierungen von Reaktionsweisen nicht diejenigen semantischen Ressourcen mobilisiert werden können, die man braucht, um die Ebene situationsübergreifender Kommunikation von Gehalten zu erreichen. Dass es *jetzt hier* brennt, kann man ja nur dann wirklich sagen (oder verstehen), wenn man *auch* sagen (bzw. Anhaltspunkte für ein Verstehen in dem Sinne haben) könnte, dass es zu einem anderen Zeitpunkt und/oder an einem anderen Ort brennt. Wie aber sollte sich eine solche Kompetenz anhand der unmittelbaren Reaktionsweisen von Sprechern auf Situationen bzw. Hörern auf Äußerungen ausweisen lassen? Was wäre die typische Reaktion auf die Mitteilung, dass es gestern an einem anderen Ort gebrannt hat? Und wie könnte sich die, unterschiedlichen Situationen oder Dingen gemeinsame, Eigenschaft zu brennen in den jeweiligen Reaktionsweisen festmachen lassen, je nachdem, ob es jetzt hier oder ob es gestern an einem anderen Ort gebrannt hat?

Die offensichtliche Aussichtslosigkeit des Versuchs, hierauf plausible Antworten zu geben, liefert ein starkes Argument dafür, unserer sprachliche Praxis und ihr kommunikatives Potential auf einer anderen Ebene, als der von Konditionalregeln für situationsrelative Verwendungen von bzw. Reaktionsweisen auf Äußerungen zu rekonstruieren. Durch den Begriff ‚*Quasiprädikatensprache*‘ soll ja gerade zum Ausdruck gebracht werden, dass, wenn die propositionalstrukturelle, *funktionale Differenzierung von Identifikation und Prädikation* (und mit ihr die semantischen Unterscheidbarkeit von singulären und generellen Termen) nicht gegeben ist, selbst von Prädikation eigentlich keine Rede sein kann. Denn durch einen Singulärterm wird ja

erst dasjenige bestimmt, auf was sich eine Prädikation bezieht. Eine Quasiprädikation, also eine ‚Prädikation‘ ohne eine ergänzende Identifikationsfunktion, wäre sozusagen ohne Ort, Zeit und Gegenstand. ‚Echte‘ Prädikate sind, wie ja bereits Frege betont hatte, wesentlich ergänzungsbedürftig. Eine Eigenschaft kann man nicht irgendwie in notwendig situationsgebundener Weise ‚konstatieren‘, da eine solche ‚Eigenschaft‘ sonst gleichsam ohne jede Beschränkung wäre, das heißt, es würde eigentlich gar nichts *ausgesagt* und von einer Eigenschaft könnte keine Rede sein. Es ist immer *etwas*, dem eine Eigenschaft zukommt (bzw. etwas und etwas anderes, zwischen denen eine Relation besteht). Erst durch das Zusammenspiel der – in Tugendhatscher Terminologie – ‚Charakterisierungsfunktion‘ des generellen Terms und der *Identifikationsfunktion* des Singulärterms ergibt sich ein propositionaler Gehalt, der ausgesagt (bzw. behauptet, gewünscht, befohlen usw.) werden kann³⁵⁶

Nun präjudiziert diese hier übernommene – und noch genauer zu erläuternde – Redeweise von der Identifikationsfunktion singulärer Termini³⁵⁷ bzw. der Charakterisierungsfunktion genereller Termini zugegebenermaßen bereits eine bestimmte Auffassung davon, wie eine angemessene Analyse der Grundlagen unserer Bedeutungskompetenz auf der Ebene propositional-sprachlicher Verständigung auszusehen hätte. Insbesondere entspricht sie nicht der, konzeptuell weitgehend in der Tradition *repräsentationaler Modelle des Geistes* stehenden, formal-semantischen Terminologie des *Bezeichnens von-* oder der *Bezugnahme auf* Gegenstände(n) bzw. Klassen von Gegenständen als *explanativer Basisbegriffe* einer Theorie der Bedeutung. Bevor ich aber weiter für die von mir präferierte Art von Begrifflichkeit argumentiere,

³⁵⁶ Nun könnte man natürlich fragen, warum ich hier Positionen, wie den *klassischen* linguistischen Behaviorismus angehe, die heute ohnehin von keinem ernstzunehmenden Theoretiker mehr vertreten werden. Ich will deshalb an dieser Stelle noch einmal darauf hinweisen, dass der Hinweis auf die, für ein Verstehen der Propositionalstruktur zentrale, Funktion der Singulärterme auch im Hinblick auf moderne naturalistische Positionen, wie die hier diskutierten, aktuell ist, insofern diese leichtfertig voraussetzungsreiche Begriffe wie ‚Wahrheit‘, bzw. ‚Wahrheitsbedingungen‘ oder ‚Tatsachen‘ ins Spiel bringen. Dies hatten wir ja zuletzt auch an Quines Konzept von Beobachtungssätzen (als sowohl quasi-epistemischer als auch genealogischer Ausgangsbasis der Sprache) gesehen, da Quine solche Sätze als wahrheitsfähig betrachtet, obwohl der ‚Aufbau eines referenziellen Apparates‘ (in beiderlei Hinsicht) nachgeordnet sein soll. Aber auch Davidson und andere hier diskutierte Naturalisten reden von ‚Wahrheit‘ und ‚Tatsachen‘, ohne uns mehr darüber zu sagen, was es heißt, dass etwas wahr ist, bzw. den Tatsachen entspricht. Es handelt sich dabei jedoch um ein begriffliches Repertoire, welches wir, wie ich meine, nur im Rahmen einer rational-rekonstruktiven Erläuterung der kommunikativen Funktion der Propositionalstruktur *explizit* verständlich machen können. Erst vor diesem Hintergrund lässt sich beurteilen, ob eine aus naturalistischer Perspektive entwickelte Begrifflichkeit (die ich als *kategorial* deskriptiv bezeichne) für eine philosophische Aufklärung von Bedeutung und Gehalt adäquat sein kann.

³⁵⁷ Der Vorschlag, den Blick von der Bezugnahme von Singulärtermen auf deren identifizierende Funktion zu lenken, geht auf P. F. Strawson (in: ders.: ‚On Referring‘ (1950a)) zurück.

möchte ich daran erinnern, dass das semantische Niveau von Quines ‚Beobachtungssätzen‘ (auch nach dessen eigenem Verständnis) dem einer ‚Quasiprädikatensprache‘ entspricht. Beobachtungssätze sollen ja auf einer semantischen Stufe stehen, die unterhalb jeder Gegenstandsindividuation liegt. Wir können nun auf den oben schon angedeuteten Einwand zurückkommen, dass Quine solche Sätze als wahrheitsfähig ansieht, obwohl er sie nicht als propositionalsprachlich strukturiert verstehen will. Dies kann schon deshalb nicht einleuchten, weil man, um einen Ausdruck oder eine Äußerung berechtigterweise im Sinne einer Aussage (bzw. als Feststellung, Behauptung, Aussagesatz, Beschreibung o.ä.) interpretieren zu können, also als etwas, das *wahr oder falsch sein* kann, offenbar zumindest eine *semantische Struktur* $(F)a$ unterstellt werden muss. Nun teilen Quine und auch Davidson zwar die hier vertretene Ansicht, dass eine Behauptung, dass $(F)a$ nicht so analysiert werden darf, als wäre sie dann wahr, wenn ihr Gehalt in irgendeiner Entsprechungsrelation zu Tatsachen steht. Dass ‚substantiell‘ gemeinte Korrespondenztheorien in die Irre gehen, liegt aber nur mittelbar an der Unsinnigkeit ihrer ‚Ontologisierung‘ von Tatsachen. Der *grundlegende* Fehler solcher Theorien ist nämlich, die assertive Satzform $(F)a$ so zu deuten, als würde durch sie ausgesagt, *dass* $(F)a$ (in irgendeinem Sinne) *wirklich ist* – und nicht, *dass a wirklich F ist*. Was es allgemein heißt, sagen oder verstehen zu können, dass ein Ding X wirklich die Eigenschaft Y hat, müssen wir erklären, um einen der zentralen Aspekte (vielleicht sogar *den* zentralen Aspekt) des Wahrheitsbegriff zu erläutern.

So will mir auch nicht recht einleuchten, warum Davidson meint, den Wahrheitsbegriff (und damit indirekt auch den Tatsachenbegriff) als ‚analytisch primitiven‘, also nicht weiter analysierbaren Grundbegriff verwenden zu können, dem er am Ende aber die ganze explanative Last seiner Theorie des semantischen Verstehens aufbürdet. Zu fragen wäre doch, worin ein Wissen(-können), dass a wirklich F ist, eigentlich besteht. Grundsätzlich liegt ja der Verdacht nahe, dass naturalistische Theoretiker, anstatt uns eine erhellende Analyse unserer Verstehens- und Verständigungskompetenz zu bieten, am Ende eigentlich nur den Versuch unternehmen, alle ‚intentionalen‘ Phänomene irgendwie in eine deskriptiv-kausale bzw. naturalistische Terminologie hineinzuzwängen und dabei dasjenige außen vor lassen müssen, was sich nicht ohne weiteres in ihrem Sinne reformulieren lässt. Auch Davidsons Theorie ist, gemessen an ihrem grundlegenden Anspruch, eigentlich eine herbe Enttäuschung. Wenn uns zur Bedeutung, bzw. über die Bedingungen des Verstehens eines singulären prädikativen Satzes, wie „Davidson ist ein Philosoph“, am Ende nicht wesentlich mehr gesagt wird, als dass man anhand beobachtbaren Sprecherverhaltens herausbekommen können muss, dass dasjenige, für das „Davidson“

steht – nämlich Davidson – unter die Klasse derjenigen Dinge fällt, die Philosophen sind und es diese ist, für das „- ist ein Philosoph“ steht, erscheint mir der Erkenntnisgewinn, vorsichtig formuliert, eher dürftig. Und daran ändert sich prinzipiell auch dadurch nichts, dass hinzugefügt wird, dass nicht-triviale der Konzeption Davidsons bestünde darin, den formalen Apparat Tarskis (*qua* Optimierungsprinzipien) im Sinne einer empirischen Theorie der ‚Interpretation‘ umzusetzen. Denn die Bedingungen der Möglichkeit Radikaler Interpretation könnten uns ja bestenfalls angeben, was ein Interpret, der bereits über ein Wissen verfügt, welches sich als Befähigung zur Erstellung einer zutreffenden ‚Wahrheitstheorie‘ für eine Sprache *beschreiben* lässt, tun muss, um eine entsprechende Theorie für eine *ihm unbekannte* Sprache aufstellen zu können. Und es bleibt offen, *was es überhaupt heißen kann, über die Befähigung zur Erstellung einer Wahrheitstheorie zu verfügen bzw. was es heißt, zu sagen, zu verstehen und beurteilen zu können, dass etwas wahr ist.*

Zur kommunikativen Struktur propositional-sprachlicher Verständigung

Anders als Davidson verstehe ich den Wahrheitsbegriff nicht als *deskriptiv* zu verwendenden, *analytisch primitiven Begriff*, sondern (mit der Tradition des Pragmatismus) primär als *Stellungnahme-* bzw. *Zustimmungsbegriff*, dessen kompetente Verwendung allerdings *voraussetzungsreich* und einer rationalen Rekonstruktion zugänglich ist. Um zu einer wirklich erhellenden Analyse unseres Verständnisses dessen zu gelangen, was es heißt, etwas für wahr zu halten (bzw. als wahr zu beurteilen oder zu behaupten) und was es heißt, Wahrheitsbedingungen zu kennen (bzw. sie mentalen Einstellungen, Sätzen oder Äußerungen zuzuordnen), müssen wir bei der einfachsten Form von Wahr-falsch-Beurteilungen ansetzen, nämlich solche von singulären, nicht-relationale Behauptungen. Wir müssen eine Form von Erklärung dafür anbieten, *worin* unser Verstehen einer singulären prädikativen Behauptung (der behauptenden Verwendung eines Satzes) besteht. Mein teils an Strawson und Tugendhat, teils an der pragmatistischen Tradition orientierter Vorschlag lässt sich etwa wie folgt zusammenfassen: *Wer eine als wahr oder falsch beurteilbare sprachlichen Äußerung versteht, der weiß – im einfachsten Fall³⁵⁸ – , wie dasjenige, was durch das Prädikat charakterisiert (bzw. klassifiziert und unterschieden) wird innerhalb aller möglichen Gegenstände der entsprechenden Prädikation ausfindig zu*

³⁵⁸ Diese Einschränkung soll nur darauf hinweisen, dass es natürlich kompliziertere Fälle gibt, etwa wenn Quantoren, wie ‚einige‘ und ‚alle‘ ins Spiel kommen oder auch dann, wenn es um abstraktere, nicht raum-zeitliche Formen der Identifikation (oder ‚Spezifikation‘) geht.

machen wäre (wenn es existiert); und der weiß um mögliche Unterschiede im Hinblick auf sprachliche und nicht-sprachliche (normative und praktische) Konsequenzen, je nachdem, ob er oder ob andere die Äußerung für wahr oder für falsch halten.

Die Fähigkeit, den mit einer Behauptung gemeinten Gegenstand unter allen anderen ‚herausgreifen‘ und zugleich beurteilen zu können, ob er in relevanter Hinsicht denjenigen gleicht, denen (bisher) typischerweise die dem Prädikat entsprechende Eigenschaft zugesprochen wird (wurde), entspricht einer *pragmatistisch verstandenen Kenntnis der Verifikationsbedingungen* eines Satzes. Die unterschiedlichen Möglichkeiten einer *Identifikation* und *Charakterisierung* (bzw. Individuierung, Unterscheidung und Klassifizierung) verstehe ich als das *Sinnpotential* einer Sprache. Dabei sind die formalen Bedingungen der Charakterisierungsfunktion von prädikativ verwendeten Ausdrücken die relativ einfacher zu bestimmenden, insofern sich die dazu notwendigen Fähigkeiten grundsätzlich als normative Kompetenz hinsichtlich ihrer in relevanter Hinsicht gleichen Verwendung erläutern lassen. Wer die Bedeutung eines generellen Terms kennt, verfügt über eine bestimmte begriffliche Unterscheidungs- bzw. Klassifizierungskompetenz.

Einen Begriff beherrschen oder teilen zu können setzt zwar sicherlich voraus, dass wir über gewisse naturgegebene – mehr oder minder ähnliche – perzeptiv-diskriminatorische Fähigkeiten verfügen³⁵⁹: Und die faktische weitgehende Übereinstimmung dürfte wohl tatsächlich in einer geteilten evolutionären *Naturgeschichte* begründet sein. Diese Fähigkeiten bilden aber nur die Grundlage unserer begrifflichen Kompetenz. Ohnehin macht eine diskriminatorische Fähigkeit, wie ja auch Davidson und andere Naturalisten koinzidieren, noch keinen Begriff. Was bei begrifflicher Kompetenz zum bloßen diskriminatorischen Aspekt hinzutritt, ist die nur im propositionalen Zusammenhang erfüllbare, also *wesentlich ergänzungsbedürftige, prädikative Rolle* einerseits *und* ihre, nur im Zusammenhang mit anderen Begriffen bzw. Sätzen gegebene, *inferentielle Rolle* andererseits. Die Ergänzungsbedürftigkeit der Prädikation, auf die uns zuerst Frege eindringlich hinwies, wurde ja schon im Zusammenhang mit der Analyse von Verständigung auf dem kognitiven Niveau von Quasiprädikatsprachen deutlich. Und die Relevanz der inferenziellen Rolle wird, wenn gleich kausalistisch reformuliert, von den meisten

³⁵⁹ Mit dieser Bemerkung schlage ich mich – jedenfalls in diesem Punkt – auf die Seite von Gareth Evans, der in *The Varieties of Reference* (1982) so etwas wie ‚informationsgestützte kognitive Zustände‘ unterhalb der Ebene des Begrifflichen postuliert und dafür von McDowell (1994) kritisiert wird. Solche Wahrnehmungs-Zustände sollten wir tatsächlich als kausal veranlasst verstehen. Sie bilden die *natürliche Grundlage* von Unterscheidungsmöglichkeiten, die durch unsere Begriffe nur zum Teil erschöpft wird. Dagegen halte ich es für verfehlt, solche Zustände, wie Evans, an ‚Gedanken‘ bzw. propositionale Einstellungen zu assimilieren.

naturalistischen Philosophen nicht bestritten (obwohl es auch erklärte ‚Widerständler‘, wie etwa Fodor, gibt).

Das Beherrschen der begrifflichen Rolle eines prädikativ verwendeten Ausdrucks umfasst darüber hinaus auch ein normativ-idealisiertes Potential, welches über die Regelmäßigkeiten tatsächlicher Instantiierungen seiner richtigen Verwendungen hinausweist. Begriffliche Kompetenz schließt eine Beurteilbarkeit *relevanter Gleichheit* in stets *neuen Fällen* ein, wobei auch ein gewisses Maß an gleichsam ‚kreativer Urteilskraft‘ gefordert ist. Die Frage, woher wir wissen, worin die Gleichheit bisheriger Anwendungen und je neuer tatsächlich besteht, bzw. woher wir wissen, wie ‚der Regel zu folgen ist‘ oder um was eine Regel es sich eigentlich handelt, entspricht natürlich dem viel diskutierten ‚Problem des Regelbefolgens‘. Wittgensteins einschlägige Bemerkungen deute ich so, dass wir hier am Ende nicht mehr sinnvoll nach einem explizierbaren Wissen fragen können, da wir in letzter Konsequenz über kein anderes Kriterium verfügen, als das des *Gelingens sprachlicher Übereinstimmung im Rahmen einer gemeinsamen Praxis*. Wer – in kartesischem Geiste – mehr verlangt, der müsste zeigen, welche Art von *Begründung* hier überhaupt noch möglich oder hilfreich sein kann. Aus dem Aufweis der Unsinnigkeit des Versuchs, Begründungsansprüche auch noch dort anmelden zu wollen, wo wir uns bereits auf der Ebene der zu unterstellenden Voraussetzungen möglicher Gründe befinden³⁶⁰, folgert Wittgenstein allerdings nicht, dass begriffliche- bzw. Regelkompetenz naturalistisch gedeutet werden sollte, da eine solche Deutung die ‚Grammatik‘ des zur Diskussion stehenden begrifflichen Feldes missachtet bzw. fehldeutet und deshalb Formen der ‚Erklärung‘ anbieten will, die gar nicht zur Fragestellung passen.

Was nun die formal komplexere Komponente der Propositionalität, die Identifikationsfunktion, betrifft, gilt es genauer zu erläutern, *wie* ein singulärer Terminus denjenigen von allen möglichen Gegenständen (eines bestimmten Bereichs) der Rede bestimmt (‚herausgreift‘), welcher gemeint ist. Ich werde mich im Folgenden auf Fälle direkt wahrnehmbarer Gegenstände beschränken. Denn einerseits bilden diese ja die von Naturalisten üblicherweise diskutierten Beispiele; andererseits spielen singuläre Aussagen, die sich durch direkt wahrnehmbare Gegenstände überprüfen lassen, tatsächlich eine in mancher Hinsicht *begründungslogisch grundlegende* Rolle.

Ich hatte gesagt, dass wir keinen Grund haben, situationsgebundene Äußerungen bzw. Quasiprädikate als wahr oder falsch – anstatt lediglich als richtig oder unrichtig – beurteilbar zu unterstellen. Denn das, was eine Äußerung oder ein Satz aussagt, *als wahr zu beurteilen*, bedeutet ja zunächst vor allem, *zuzustimmen*, dass das durch den Singulärterm bestimmte *etwas* die durch den generellen Term bestimmte Eigenschaft

³⁶⁰ S. insbes. auch Wittgenstein: *Über Gewißheit* (1984).

hat. Es kommt nun darauf an, genauer zu erläutern, inwiefern der Ausdruck „etwas“ eine semantische Kategorie markiert, die sich dadurch auszeichnet, dass eine *situationsübergreifende Identifizierbarkeit* möglich ist. Klassische Theorien, die die Bedeutung von Singulärterminen vor allem nach dem Paradigma von Namen erklären wollen, wie etwa die J. S. Mills³⁶¹, setzen voraus, dass die Redeweise von Dingen, für die Namen stehen (bzw. die durch diese bezeichnet werden), elementar in dem Sinne ist, dass unsere Rekonstruktionsbemühungen bei der Unterstellung enden müssen, wir verstünden einfach, was es heißt, dass ein Name für einen Gegenstand steht. Moderne, wahrheitssemantisch orientierte Theorien zerfallen dagegen im Wesentlichen in zwei Gruppen: Die eine analysiert Namen entweder (wie Frege) vollständig, oder doch zumindest (wie Russell³⁶²) weitgehend als Kennzeichnungen, also als nur mittels einer Beschreibung Bezug nehmend. Die andere Gruppe beruft sich, im Anschluss an Kripke, doch wieder auf die klassische Theorie, die dann allerdings in der ein oder anderen Weise (etwa im Sinne einer ‚historischen‘ bzw. kausalen Deutung) modifiziert wird. Beiden Spielarten der Wahrheitssemantik, wie sie auch für naturalistische Theorien maßgeblich sind, ist es jedoch zu eigen, nicht nach der spezifischen, *die situationsunabhängige Kommunikation zwischen Subjekten* ermöglichenden, *Funktionsweise* singulärer Termini zu fragen. Eine pragmatistisch orientierte, rationale Rekonstruktion dieser Funktionsweise sieht die Möglichkeit der Rede von Gegenständen oder Dingen – und damit auch die Möglichkeit der Bezugnahme oder des Bezeichnens – als *abhängig* von der Möglichkeit der *Identifikation und Reidentifikation durch ein komplexes Zusammenspiel verschiedener Formen singulärer Termini und bestimmter Rahmenbedingungen*³⁶³. Durch die Verwendung eines singulären Terminus bestimmen wir, welcher unter allen möglichen Gegenständen der Welt gemeint ist. Und zumindest in jener wichtigen Klasse von Fällen, auf die wir uns hier beschränken, nämlich den raum-zeitlich lokalisierbaren, direkt wahrnehmbaren Gegenständen, *bedeutet eben ein Gegenstand zu sein, perspektiven- und situationsübergreifend identifizierbar und ‚wiederauffindbar‘ zu sein.*

Wie und unter welchen Bedingungen gelingen also konkrete Identifikationen? Im einfachsten Fall kann ein Sprecher, der sich auf einen für den Hörer situativ wahrnehmbaren Gegenstand beziehen will, diesen häufig bereits durch eine hinweisende – und durch eine Zeigegeste begleitete – Äußerung, wie „dies“ oder „das

³⁶¹ J. S. Mill: *A System of Logic* (1843), Kap. 2.

³⁶² Russell: ‚On Denoting‘ (1905). Für Russell sind ‚echte‘ bzw. ‚logische‘ Namen nur „dies“ und „Ich“, da nur hier die Bezugnahme epistemisch gesichert sei.

³⁶³ Wobei darüber hinaus allerdings auch die von Brandom (1994, Kap. 6 u. 7) beschriebene Ermöglichung der Bildung inferentiell relevanter – diskursive Berechtigungen und Verpflichtungen erhaltender – ‚anaphorischer Ketten‘ eine wichtige Rolle spielen mag.

da“ identifizieren. Das gilt vor allem in Fällen, wo der zu identifizierende Gegenstand schon durch seine situative Relevanz oder Prägnanz hervorsticht. Ebenso kann ein Sprecher sich in einer gemeinsamen Verständigungssituation durch „ich“, den (oder die) Angesprochenen durch das Pronomen „du“ (bzw. „ihr“) und andere Beteiligte durch „er“, „sie“ oder „es“ identifizieren. Auf einen nicht-wahrnehmbaren Gegenstand können wir durch Pronomina, wie „jenes“, Bezug nehmen, die durch den Hinweis auf eine Wahrnehmungssituation spezifizierbar sind, in der der Gegenstand durch „dies“ identifiziert wurde. Allerdings kann eine hinweisende Identifizierung durch „dies“ unklar bleiben. So könnte jemand sowohl den Tisch oder die Tischplatte, aber auch das Material meinen, aus dem er gefertigt ist. Dann muss auf sogenannte ‚sortale Prädikate‘ wie etwa „Tischplatte“ zurückgegriffen werden. Ein solches Sortal bestimmt, anders als etwa ein Farbprädikat, Kriterien der räumlichen oder zeitlichen Abgrenzung (und damit der Zählbarkeit) des gemeinten Gegenstandes.

Nun hatte ich, in Anlehnung an Tugendhat (1976), behauptet, dass von Gegenstandsbezug und wahrheitsfähigen Aussagen überhaupt nur dann die Rede sein kann, wenn der Gegenstand der Prädikation *situations- und perspektivenübergreifend* identifizierbar bzw. reidentifizierbar ist. Das heißt zugleich, dass sich sprachliche *Bedeutung* auf der Ebene *wahrheitsfähiger Gehalte* gegenüber situationsgebundenen Kommunikationsformen durch eine Art *objektivierender Abstraktion* auszeichnet. Wenn etwa Paul in einer bestimmten Situation sagt, er habe Zahnschmerzen, indem er äußert: „Ich habe Zahnschmerzen“, muss er, ebenso wie Anwesende, die den Satz verstehen, also wissen, dass sie *situationsgebunden*, aber *perspektivenübergreifend*, die *gleiche Aussage* durch den Satz „Du hast Zahnschmerzen“ oder den Satz „Er hat Zahnschmerzen“ zum Ausdruck bringen können, in der Lage sein, *dies* auch einem Dritten, der sich nicht in dieser Situation befindet, mitzuteilen, etwa indem Paul am Tag danach sagt „Ich hatte gestern Zahnschmerzen“ oder indem die bei der ursprünglichen Äußerung Anwesenden später sagen: „Paul hatte am Dienstag Zahnschmerzen“. Das heißt, sie müssten Ausdrücke bzw. Worte zu wählen in der Lage sein, die es einem Dritten ermöglichen, den *identischen Gehalt* der Aussage situationsübergreifend – bzw., bei Nennung eines Datums, situationsunabhängig – zu verstehen. Verständigungsteilnehmer müssen einen identischen Gehalt durch die *richtige Substitution von Singularterminen* zum Ausdruck bringen und verstehen können, um die jeweilige Subjektivität von Perspektiven und Singularität von Situationen objektivierend zu ‚überbrücken‘. Diese Objektivierung vermittelt also nicht nur zwischen den verschiedenen Perspektiven von Subjekten, sondern stellt auch den Bezug zwischen verschiedenen Situationen untereinander und zu einer ursprünglichen oder möglichen *Verifikationssituation* her.

Diese Anforderung an die Identifikationsfunktion von Singularterminen wird allerdings erst durch ein ziemlich komplexes Wechselspiel von Bedingungen erfüllt. Eine wesentliche Rolle für die Möglichkeit einer Identifikation spielen vor allem auch Kennzeichnungen. Diese lassen sich in *kontextunabhängige* oder eindeutige und in *kontextabhängige* oder nicht-eindeutige unterscheiden (erstere sind jene, die Russell „*definite descriptions*“ genannt hat). So wird durch Kennzeichnungen, wie „der gegenwärtige Oberbürgermeister von München“ oder „der höchste Berg Pakistans“ der Gegenstand einer möglichen Prädikation eindeutig identifiziert, wogegen eine Kennzeichnung, wie „mein Auto“ einen Gegenstand nur in Abhängigkeit vom Äußerungskontext, eine Kennzeichnung, wie „der blonde Kerl mit der Lederjacke“ einen Gegenstand darüber hinaus nur unter spezifischen Bedingungen (vor allem der, dass es nicht mehrer relevante blonde Lederjackenträger gibt) eindeutig identifiziert. Auch die traditionellerweise als für Bezug nehmende Ausdrücke prototypisch angesehenen *Eigennamen*, wie etwa „Anneliese Rothenberger“, „Frankfurt am Main“ oder „K 2“, verweisen in ihrer identifizierenden Funktion auf *Kennzeichnungen* bzw. Bündel von Kennzeichnungen. Wie Tugendhat betont, gehört jedoch zum Verständnis der identifizierenden Funktion solcher Kennzeichnungen am Ende auch das Wissen um die Möglichkeit eines Übergangs zu Kennzeichnungen, die eine *endgültige, objektiv lokalisierende* Identifikation ermöglichen. Denn wenn ich etwa frage: „Welcher Gegenstand ist es, den Du mit ‚Frankfurt‘ meinst“ und jemand antwortet mit einer Kennzeichnung, wie „Es ist die größte hessische Stadt“, kann ich natürlich immer noch fragen, *welche* denn nun die größte hessische Stadt ist. Eine entsprechende, objektiv lokalisierende Identifizierung ist dann durch eine Kombination lokalisierender Kennzeichnungen und sortaler Prädikate möglich, anhand derer ein Gegenstand räumlich abgegrenzt bzw. unterscheidbar ist. Eine als sortales Prädikat fungierende Kennzeichnung für Frankfurt am Main wäre natürlich „Stadt“ oder etwa „große Stadt“. Eine objektiv lokalisierende Identifizierung kann hier etwa durch die Kennzeichnung „ungefähr vierzig Kilometer von der Mündung des Mains in den Rhein Flussaufwärts gelegen“ gelingen. Allerdings muss eine solche Kennzeichnung natürlich ihrerseits verstanden werden. Am Ende müssen wir den Gegenstand in einer unmittelbaren Wahrnehmungssituation mit Hilfe eines Sortals *gemeinsam ostensiv* identifizieren können.

Auch zum Verständnis von Namen im semantischen Sinne von *Eigennamen* gehört es ganz allgemein, dass man weiß, *wie* sie ihre Funktion erfüllen (können), einen Gegenstand in eindeutiger Weise aus der Gesamtheit von Gegenständen herauszugreifen. Dazu gehört auch das Wissen um jene ‚Starrheit‘, von der schon die Rede war, also dass der *Bezug eines Namens über eine kommunikative Kette seiner*

historischen Verwendungen auf eine ursprüngliche Benennungs- oder ‚Taufsituation‘ zurückführbar ist³⁶⁴. Wir können dann offenbar sagen, dass man sich etwa mit der Verwendung eines Personennamens auch dann auf die richtige Person beziehen kann, wenn die *meisten* Kennzeichnungen, die man mit dem Namen assoziiert, falsch sind. Man kann also die Bedeutung eines Eigennamens in einer wichtigen Hinsicht auch dann kennen, wenn man außer denjenigen Kennzeichnungen, die für eine endgültige Identifikation des auf diesen Namen getauften ‚Gegenstandes‘ (bzw. der auf diesen Namen ‚getauften‘ Person) erforderlich sind, nichts über den Träger weiß bzw. alle darüber hinausgehenden Überzeugungen falsch sind³⁶⁵. Dazu aber muss man das ganze Repertoire der eine endgültige Identifikation ermöglichenden singulären Termini im Griff haben, das es uns erst ermöglicht, das durch einen Namen Bezeichnete an irgendeiner Stelle in der Raum-Zeit zu finden (bzw. wiederzufinden) und endgültig zu identifizieren. Ohnehin muss ja auch in einer ‚Taufsituation‘ der beabsichtigte Namensträger zunächst einmal durch ein hinweisend verwendetes Demonstrativpronomen und ein sortales Prädikat identifiziert werden können³⁶⁶.

Strawson hatte in *Individuals* (1959) bereits die Wichtigkeit von dauerhaften materiellen Gegenständen als raum-zeitlicher Fixpunkte für eine situationsübergreifende objektive Lokalisierung betont. Dieser Gedanke wird durch Tugendhat komplementär ergänzt: „So wie durch materielle Gegenstände Raumstellen markierbar werden, werden durch Veränderungen Zeitstellen markierbar“ (Tugendhat 1976, S. 457). Entsprechende Fixpunkte für räumliche Lokalisierungen bestehen meist in markanten dauerhafte Gegenständen. Die Möglichkeit *objektiver* zeitlicher Lokalisierungen beruht dagegen auf dem Faktum regelmäßig wiederkehrender natürlicher (etwa astronomischer) Ereignisse. Einen wesentliche Objektivierung und Präzisierung raum-zeitlicher Lokalisierbarkeit wurde natürlich durch die Einführung des Koordinatensystems von Längen- und Breitengraden und die Entwicklung einer standardisierten und mechanisierten Entfernungs- und Zeitmesstechnik ermöglicht. Bekanntermaßen sind deren willkürlich gewählte ‚Null-Fixpunkte‘ einerseits Greenwich (jedenfalls für die Verhältnisse auf der Erde), andererseits (jedenfalls im

³⁶⁴ Eine hervorragende, ausführliche Darstellung der Besonderheit der Semantik von Eigennamen findet sich bei Evans (1982), insbes. Kap. 11.

³⁶⁵ Zu einer umfassenderen Kompetenz hinsichtlich des Verstehens von Eigennamen gehört allerdings, wie wir oben an Dummetts (1991) Beispiel der Institution der Ortsnamen gesehen hatten, einiges mehr, wie etwa – im genannten Fall – ein Wissen um Reisemöglichkeiten und gegebenenfalls (bei berühmten Städten) auch über historische Hintergründe.

³⁶⁶ Was übrigens auch bedeutet, dass auf der basalsten Ebene der Funktionsweise singulärer Termini, nämlich der unmittelbaren raum-zeitlichen Identifikation, Donnellans – ansonsten berechnete – Unterscheidung zwischen der attributiven und der referentiellen Verwendung von Kennzeichnungen, von der oben schon die Rede war, keine Rolle mehr spielt.

westlich beeinflussten Kulturkreis) jenes astronomisch bestimmte Datum, welches für Christi Geburt stehen soll.

Etwas identifizieren setzt also voraus, dass wir unsere jeweiligen *subjektiven Perspektiven* zu jenen im vorigen Absatz beschriebenen, mehr oder minder starren *Referenzsysteme objektiver Lokalisierbarkeit in Beziehung setzen können*. Denn etwas räumlich lokalisieren zu können, bedeutet *für ein Subjekt*, zu wissen, *wo* es sich *relativ zu seinem Standpunkt* im Raum befindet oder es zu finden wäre; und etwas zeitlich lokalisieren zu können bedeutet, bestimmen zu können, *wann* es sich *relativ zum eigenen Jetzt* ereignet hat oder ereignen wird. Dabei muss dann natürlich auch die Beziehung der je meinigen Perspektive zum objektivierenden Referenzsystem mit den je subjektiven Perspektiven Anderer ins Verhältnis gesetzt werden können. Jemand versteht eine assertorische Äußerung eines anderen nur dann, wenn er in der Lage ist, die Weise, in der der Sprecher den Gegenstand aus seiner Perspektive (Position) identifiziert zu seiner eigenen Perspektive (Position) in Bezug zu setzen. Erst wenn die Subjektivität jeweiliger Perspektiven bzw. die Indexikalität der Situationen, in der wir uns jeweils befinden, also ‚ich‘, ‚hier‘ und ‚jetzt‘, in *praktisch folgenreicher Weise* nicht nur in Relation zu ‚du‘, ‚früher‘, ‚später‘, ‚dort‘, sondern auch zu einem objektivierenden Referenzsystemen gesetzt werden kann, ist ein Verstehen bzw. Beurteilen propositionaler Gehalte möglich.

Materielle Gegenstände (auch Personen) haben eine *raum-zeitliche Geschichte* mit, je nach Gegenstandsart, unterschiedlichen Identitätskriterien und mehr oder minder bestimmbarem Beginn und Ende. Das heißt, die kompetente Verwendung singulärer Termini und deren Verstehen setzt auch eine hinreichende Kenntnis und Urteilsfähigkeit hinsichtlich der jeweiligen raum- und zeitübergreifenden Identitätskriterien voraus. Dies ist natürlich ein Thema für sich, auf das ich mich hier aber nicht mehr einlassen möchte. Jedenfalls können wir festhalten, dass es eine der wesentlichen Eigenschaften *propositional-sprachlicher* Kommunikation ist, etwas durch eine *situationsrelative Verwendung von Ausdrücken situations- und individuen-übergreifend so zum Ausdruck bringen* zu können, dass es *als Identisches verstanden bzw. beurteilt*, werden kann. Bedeutungskompetenz besteht also wesentlich in der Fähigkeit, sowohl etwas Identisches mit unterschiedlichen (sprachlichen) Mitteln, als auch mit gleichen Mitteln Unterschiedliches sagen zu können – und die entsprechenden Zusammenhänge zu kennen. Und dieses Identische zu verstehen, heißt nicht nur, über eine entsprechende Identifikations- und Charakterisierungskompetenz zu verfügen. Es heißt, wie ich oben argumentiert habe, auch, die unterschiedlichen normativen (diskursiven und praktischen) ‚Räume‘ von Verantwortlichkeit zu überblicken, innerhalb derer man sich durch eine Äußerung *positioniert*.

Der produktive Charakter von Verstehens- und Verständigungsbemühungen

Die Rede vom ‚Identischen‘ im Sinne geteilter propositionaler Gehalte, der ich mich im letzten Abschnitt bedient habe, stellt natürlich eine Idealisierung dar. Für sich genommen lässt sich daraus aber kaum ein Vorwurf ableiten, denn in gewissem Sinne stellt *jeder Begriff* von Identität oder Gleichheit eine *Idealisierung* dar. Problematisch werden Idealisierungen nur, wenn sie hypostasierend zu irgendwelchen abstrakten Entitäten mit eigentümlicher Daseinsberechtigung hochstilisiert werden. Solche Deutungen mögen in Bezug auf formale oder strukturelle Zusammenhänge, wie Mathematik oder Logik, sogar etwas für sich haben. In Bezug auf *materiale* Gehalte oder ‚Gedanken‘ scheinen sie mir jedoch völlig unangebracht, und ich hatte es ja auch vorgezogen, von notwendigen, handlungs- oder urteilsorientierenden Unterstellungen zu sprechen.

Eine bestimmte Art von zu Missverständnissen Anlass gebender ‚Idealisierung‘ spielt allerdings unter der Hand auch in der philosophischen Tradition eine wichtige Rolle, die sich am Modell des weltrepräsentierenden singulären Geistes orientiert und zu der – in der ein oder anderen Weise – auch die hier kritisierten naturalistischen Positionen zählen. Wie wir gesehen haben, liegt ja auch naturalistischen Theorien meist die Vorstellung zugrunde, Wörter und Sätze seien lediglich Mittel des ‚öffentlichen Transports‘ von etwas (Repräsentationen, Gedanken, Einstellungen), das ‚in den Köpfen‘ der ‚Verständigungsteilnehmer‘ bereits *vorhanden* ist, also gleichsam *fertig vorliegt* – in welcher Form auch immer. Diese Vorstellung ist aber vor allem auch deshalb unangemessen, weil wir unsere sprachlichen Verständnisse bzw. die Bedeutung unserer Wörter und die entsprechenden, idealisierend als identisch unterstellten, gemeinsamen Gehalte, im *Prozess der Verständigung* durch den *konkreten Gebrauch* von Sätzen – jedenfalls ein Stück weit – ‚erzeugen‘.

Häufig ist nämlich, wie ich oben schon angedeutet hatte, das Sinnpotential von Sätzen gegenüber jeweiligen Verwendungssituation ‚bedingt hermeneutisch‘ offen. Das heißt aber, dass geäußerten Sätzen ein Sinn in unmittelbaren Verständigungssituationen mitunter erst gegeben werden muss. Diesen ‚Herstellungscharakter‘ von Sinn in Verständigungssituationen beschreibt Kambartel³⁶⁷ folgendermaßen:

„Der genaue Sinn der Sätze ist in den geschilderten Situationen weder als eine wohlbestimmte subjektive Intention richtig begriffen noch als eine Erscheinung, welche Sätzen gleichsam ‚objektiv‘, aufgrund allgemeiner den Sprachgebrauch tragender Vereinbarung anhaftet. Der Gebrauch, den wir von einem Satz in einem

³⁶⁷ F. Kambartel: ‚Ein Versuch über das Verstehen‘ (1991).

bestimmten Falle machen, ist eben im Allgemeinen nicht *vor* seiner gelingenden Verwendung bereits völlig vorhanden und fertig.

Was wir mit dem Gesagten meinen, *stellen* wir ein Stück weit *selbst erst fest*, wenn wir auf Verständnisfragen antworten, gemeinsam mit anderen an der Deutung unserer Sätze arbeiten oder schließlich in der von solchen Deutungen angeleiteten Praxis unsere Übereinstimmung oder ihr Ausbleiben sehen.“ (S. 127)

Kambartel nennt ein entsprechendes hermeneutisches Bemühen die „dialogisch-praktische Einbettung“ eines Satzes. Diese ist gegebenenfalls auch deshalb erforderlich, weil unser Verständnis von Sätzen nicht alle *möglichen* Handlungssituationen bzw. Verständigungsprobleme umfassen kann. Man könnte hier von einer dynamischen wechselseitigen Beziehung zwischen einer – aktuellen Verständigungsbemühungen je vorgängigen – Sinnstruktur (die wir vorab als Standardbedeutungen eines allgemeinen Sprachgebrauchs unterstellen) und der verständigungsorientierten, ‚dialogisch-praktischen Einbettung‘ eines Satzes in einer konkreten Verwendungssituation sprechen.

Hermeneutischen Fragen spielen natürlich eine umso größere Rolle, je weniger an (‚doxastischen‘ oder kulturellen) Gemeinsamkeiten wir bei unserem Verständigungspartner – auch innerhalb der Institution einer Sprachgemeinschaft – unterstellen können. Dominierend werde sie aber, wenn die Lebenspraktiken derer, die versuchen sich zu verstehen, stark abweichend sind. Wir müssen uns dann zunächst auf sehr basale Gemeinsamkeiten im Sinne kulturübergreifender ‚anthropologischer Grundkonstanten‘ stützen, um überhaupt ein Verstehen bzw. eine Verständigung *aufbauen* zu können. Auch im gadamerschen Begriff von der ‚Horizontverschmelzung‘ steckt ja, dass wir, ausgehend von unseren jeweiligen ‚Vor-Verständnissen‘, einen zum Verstehen notwendigen, *geteilten Sinn* erst produktiv herstellen müssen.

Ein solches Bemühen etwa im Sinne der Erstellung einer Wahrheitstheorie durch einen davidsonischen ‚Radikalen Interpretieren‘ deuten zu wollen, ist völlig unangemessen³⁶⁸. Eine solche Beschreibung von Situationen des Verstehens unterschlägt vollständig die Produktivität einer ‚überbrückenden‘ Annäherung an unterschiedliche begriffliche Verständnisse. Beispielsweise können wir den Begriff der Demokratie, wie er im antiken Athen verstanden wurde, nicht einfach durch den das moderne westliche (Standard-) Verständnis artikulierenden Term „Demokratie“ beschreiben. In solchen Fällen, in denen eine gemeinsame Verständigung nicht mehr möglich ist, bedarf es zumindest einer ‚dichten Erläuterung‘, die überlieferte Äußerungen bzw. Texte in Kontexte kulturspezifischer Maßstäbe einer Richtig-falsch-

³⁶⁸ Hierzu – und zu anderen Defiziten *repräsentationalistisch* verkürzter Verständnisse der Möglichkeiten sprachlicher Artikulation (mit einigen anschaulichen Beispielen): C. Taylor: ‚Bedeutungstheorien‘ (1982/1988).

Beurteilung einbettet und zugleich einen Bezug zu jenen Maßstäben vernünftigen Urteilens und Beurteilens herzustellen vermag, von denen her wir uns jeweils einer angemessenen Interpretation annähern müssen. Doch auch im Hinblick auf Fälle *möglicher* Verständigung sehe ich keinen Grund dafür, warum wir uns beim Nachdenken über Bedeutung, Verstehen und Verständigung auf die Unterstellung von Situationen Radikaler Interpretation einlassen sollten, außer dem, dass man einfach glaubt, alles, was sich sinnvollerweise über sprachliche Bedeutung sagen lasse, müsse den Bedingungen einer unter den entsprechenden Vorgaben zu erstellenden, wahrheitssemantisch verfassten empirischen ‚Theorie des Verstehens‘ genügen.

LITERATURVERZEICHNIS

- Anscombe, G. E. M.: *Intention*, 2. Aufl. (1957/1963), Oxford.
- Austin, J. L.: *How to do Things with Words* (1960), Oxford.
- Beauchamp, T. / Rosenberg, A.: *Hume and the Problem of Causation* (1981), New York/Oxford.
- Beckermann, A.: ‚Handeln und Handlungserklärungen‘ (1977), Kronberg/Ts.
- ‚Handeln und Handlungserklärungen‘ (1977a), Frankfurt/M.
 - *Analytische Einführung in die Philosophie des Geistes* (1999), Berlin/New York.
- Bennett, J.: *Linguistic Behaviour* (1976), Indianapolis.
- Bilgrami, A.: *Belief and Meaning* (1992), Cambridge/Ma.
- ‚Norms and Meaning‘ (1993), in: Ralf Stoecker (Hg.): *Reflecting Davidson* (1993), Berlin/New York.
- Block, N.: ‚Troubles with Functionalism‘ (1978), in: ders. (Hg.): *Readings in the Philosophy of Psychology* (1980), Cambridge/Ma.
- ‚Advertisements for a Semantics for Psychology‘ (1986), in: *Midwest Studies in Philosophy X*.
- Bloomfield, L.: *Language* (1935), London.
- Boghossian, P.: ‚The Rule-Following Considerations‘ (1989), in: *Mind* 98.
- ‚Naturalizing Content‘ (1991), in: B. Loewer/G. Rey (Hg.): *Meaning in Mind. Fodor and his Critics*, Cambridge/Ma.
 - ‚Analyticity Reconsidered‘ (1996), in: *Journal of Philosophy*, in: *Nous* 30.
- Boorse, C.: ‚Wright on Functions‘ (1976), in: *The Philosophical Review* LXXXV, 1.
- Brandom, R.: *Making in Explicit* (1994), Cambridge/Ma.
- Burge, T.: ‚Individualism and the Mental‘ (1979), in: *Midwest Studies in Philosophy IV*,
- Carnap, R.: *Der logische Aufbau der Welt* (1928/1961), Hamburg.
- ‚Überwindung der Metaphysik durch logische Analyse der Sprache‘ (1931/1932), in: *Erkenntnis* 2.
 - *Meaning and Necessity* (1947/1956), Chicago.
- Chalmers, A. F.: *Wege der Wissenschaft*, 5. Aufl. (2001), Berlin/Heidelberg/New York.
- Chomsky, N.: ‚Review of B. F. Skinner’s *Verbal Behaviour*‘ (1959), wiederabgedr. in: Block (Hg.) (1980).
- *Regeln und Repräsentationen* (1980/1981), Frankfurt/M.
 - ‚Language and Nature‘ (1995), in: *Mind* 104.
- Churchland, P. A. : ‚Der logische Status von Handlungserklärungen‘ (1970/1977), in: A. Beckermann (Hg.) (1977).

- *A Neurocomputational Perspektive* (1989), Cambridge/Ma.
- Collins, A.: ‚Action, Causality and Teleology‘ (1984), in: *Midwest Studies in Philosophy IX*.
- Cummins, R.: ‚Functional Analysis‘ (1975), in: *Journal of Philosophy* 52.
- *The Nature of Psychological Explanation* (1983), Cambridge/Ma.
- Davidson, D.: - ‚Handlungen, Gründe und Ursachen‘ (1963), in: ders. (1980/1985).
- ‚Wahrheit und Bedeutung‘ (1967), in: ders. (1984/1986).
- ‚Kausale Beziehungen‘ (1967a), in: ders. (1980/1985).
- ‚Mentale Ereignisse‘ (1970), in: ders. (1980/1985).
- ‚Radikale Interpretation‘ (1973), in: ders. (1984/1986)
- ‚Was ist eigentlich ein Begriffsschema?‘ (1974), in: ders. (1984/1986).
- ‚Denken und Reden‘ (1975), in: ders. (1984/1986).
- ‚Die Unerforschlichkeit der Bezugnahme‘ (1979), in: ders.: (1984/1986)
- *Handlung und Ereignis* (1980/1985, Frankfurt/M.
- ‚Kommunikation und Konvention‘ (1982), in: ders. (1984/1986).
- ‚Reply to Quine on Events‘ (1985), in: E. Le Pore/B. Mc Laughlin (Hg.): *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, Oxford/New York 1985.
- ‚A Coherence Theory of Truth and Knowledge‘ (1986), in: E. LePore (1986)
- ‚Eine hübsche Unordnung von Epitaphen‘ (1986a), in: E. Picardi/J. Schulte (Hg.): *Die Wahrheit der Interpretation. Beiträge zur Philosophie Donald Davidsons* (1990), Frankfurt/M.
- ‚Knowing one’s own Mind‘ (1987), in: *Proceedings and Addresses of the American Philosophical Association* 1987.
- ‚The Structure and Content of Truth‘ (1990), in: *Journal of Philosophy* 87.
- ‚Subjektiv, Intersubjektiv, Objektiv‘ (1991), in: *Merkur* 512.
- ‚Three Varieties of Knowledge‘ (1991a), in: A. Phillips Griffiths (Hg.): *A. J. Ayer: Memorial Essays*, Cambridge/Ma.
- ‚Thinking Causes‘ (1993), in: J. Heil/A. Mele (Hg.): *Mental Causation*, Oxford.
- ‚Reply to Akeel Bilgrami‘ (1993a)
- ‚The Social Aspect of Language‘ (1994), in: B. Mc Guinness/G. Oliveri (Hg.): *The Philosophy of Michael Dummett*, Dordrecht.
- ‚Donald Davidson‘ (1994a), in: S. Guttenplan (Hg.): *A Companion to the Philosophy of Mind*, Cambridge/Ma.
- ‚Unbestimmtheit und Anti-Realismus‘ (1998), in: T. Metzinger (Hg).
- Dennett, D.: ‚Intentional Systems‘ (1971), in: ders. (1978a).
- *Brainstorms: Philosophical Essays in Mind and Psychology* (1978a), Montgomery/Vt.

- ‚Artificial Intelligence as Philosophy and as Psychology‘ (1978b), in: ders. (1978a).
 - ‚Three Kinds of Intentional Psychology‘ (1981), in: ders. (1987).
 - ‚True Believers‘ (1981a), in: ders. (1987).
 - ‚Making Sense of Ourselves‘ (1981b), in: ders. (1987).
 - ‚Beyond Belief‘ (1982), in: ders. (1987).
 - ‚Evolution, Error, and Intentionality‘ (1987a), in: ders. (1987)
 - *The Intentional Stance* (1987), Cambridge/Ma.
 - ‚Real Patterns‘ (1990), in: *The Journal of Philosophy* 88.
- von Ditfurth, H.: *Im Anfang war der Wasserstoff* (1972), München.
- Donnellan, K.: ‚Reference and Definite Descriptions‘ (1966), in: *Philosophical Review* 75
- Dretske, F.: *Knowledge and the Flow of Information* (1981), Cambridge/Ma.
- ‚Misrepresentations‘ (1986), in: R. Bogdan (Hg.): *Belief*, Oxford.
 - *Explaining Behaviour* (1988), Cambridge/Ma.
 - ‚Does Meaning Matter?‘ (1990), in: E. Villanueva (Hg.): *Information, Semantics, and Epistemology*, Cambridge/Ma.
 - ‚Dretske’s Replies‘ (1991), *Dretske and his Critics*, Cambridge/Ma.
- Dreyfus, H.: *Die Grenzen künstlicher Intelligenz. Was Computer nicht können* (1972/1985), Königstein/Ts.
- Dummett, M.: ‚Was ist eine Bedeutungstheorie?‘ (1975), in: ders.: *Wahrheit* (1982), Stuttgart.
- ‚„Eine hübsche Unordnung von Epitaphen“. Bemerkungen zu Davidson und Hacking‘ (1986/1990), in: Picardi/Schulte (1990).
 - *The Logical Basis of Metaphysics* (1991), Cambridge/Ma.
- Ellis, B.: ‚What Science Aims to Do‘ (1985), in: P. Churchland/C. Hooker (Hg.): *Images of Science*, Chicago/London.
- Evans, G.: ‚Reply: Semantic Theory and tacit Knowledge‘ (1981), in: C. Leich/S. Holzman (Hg.): *Wittgenstein: To Follow a Rule*, London.
- *The Varieties of Reference* (1982) (hrsg. v. J. McDowell), Oxford.
- Evans-Pritchard, E. E.: *Hexerei, Orakel und Magie bei den Zande* (1937/1978), Frankfurt/M.
- Field, H.: ‚Mental Representation‘ (1978), in: *Erkenntnis* 13.
- Fodor, J.: *Psychological Explanation* (1968), Cambridge/Ma.
- *The Language of Thought* (1975), Cambridge/Ma.
 - ‚Methodological solipsism considered as a research strategy in cognitive science‘ (1980), in: ders.: *Representations* (1981), Brighton.
 - ‚Semantics, Wisconsin Style‘ (1984), in: ders. (1990)

- ‚Fodor’s Guide to Mental Representation‘ (1985), in: ders. (1990)
 - *Psychosemantics* (1987), Cambridge/Ma.
 - ‚A Theory of Content‘ (I+II), Cambridge/Ma.
 - *The Elm and the Expert* (1994a), Cambridge/Ma.
- Fodor, J./Lepore, E.: *Holism: A Shopper’s Guide* (1993), Cambridge/Ma.
- Føllesdal, D.: ‚Essentialism and Reference‘ (1986), in: L. Hahn/P. Schilpp (Hg.): *The Philosophy of W. v. Quine* (1986), La Salle.
- ‚Indeterminacy and Mental States‘ (1990), in: R. Barrett/R. Gibson (hg.): *Perspectives on Quine*, Cambridge/Ma.
- Frege, G.: ‚Über Sinn und Bedeutung‘ (1892/1986), in: ders.: *Funktion, Begriff und Bedeutung* (1986) (hrsg. v. G. Patzig), Göttingen.
- ‚Der Gedanke‘ (1918), in: ders.: *Logische Untersuchungen* (1986) (hrsg. v. G. Patzig), Göttingen.
- Gadamer, H. G.: *Wahrheit und Methode* (1960/1990), Tübingen.
- Gauker, C.: *Thinking Out Loud* (1994), Princeton/NJ.
- Geach, P.: *Mental Acts* (1957), London
- Geertz, C.: *Dichte Beschreibung* (1987), Frankfurt/M.
- Glüer, K.:
- ‚Bedeutung zwischen Norm und Naturgesetz‘ (2000), in: *Deutsche Zeitschrift für Philosophie* 49, 5, Berlin.
- Goldmann, G.: *A Theory of Human Action* (1970), Englewood Cliffs/NY.
- *Epistemology and Cognition* (1986), Cambridge/Ma.
- Gould, S./ Lewontin, R.: ‚The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptionist Programme‘ (1978), in: *Proceedings of the Royal Society*, B 205.
- Goodman, N.: *Tatsache, Fiktion, Voraussage* (1954/1988), Frankfurt/M.
- Grice, P.: ‚Meaning‘ (1957), *Philosophical Review* 66.
- Grice, P./Strawson, P. F.: ‚In Defence of a Dogma‘ (1956), *Philosophical Review* 65.
- Habermas, J.: ‚Theorie der Gesellschaft oder Sozialtechnologie. Eine Auseinandersetzung mit Niklas Luhman‘ (1971), in: J. Habermas/N. Luhman (Hg.): *Theorie der Gesellschaft oder Sozialtechnologie* (1985), Frankfurt/M.
- ‚Intentionalistische Semantik‘ (1975/1976), in: ders. *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns* (1984), Frankfurt/M.
 - *Theorie des kommunikativen Handelns* (1981) (4. Aufl. 1984), Frankfurt/M.
 - ‚Handlungen, Sprechakte, sprachlich vermittelte Interaktionen und Lebenswelt‘ (1988) in: ders.: *Nachmetaphysisches Denken*, Frankfurt/M..

- ‚Von Kant zu Hegel. Zu Robert Brandoms Sprachpragmatik‘ (1999), in: ders.: *Wahrheit und Rechtfertigung*, Frankfurt/M.
- Hacking, I.: ‚Eine Parodie der Konversation‘ (1986/1990), in: E. Picardi/J. Schulte (1990).
- Hardcastle, V. G.: ‚Understanding Functions: A Pragmatic Approach‘ (1999), in: dies. (Hg.): *Where Biology Meets Psychology: Philosophical Essays*, Cambridge/Ma.
- Hart, H. L. A.: ‚The Ascription of Responsibility and Rights‘ (1948/1949), in: *Proceedings of the Aristotelian Society* 49.
- Haugeland, J.: *Artificial Intelligence: The very Idea* (1985), Cambridge/Ma.
- Heckmann, H.-D.: *Mentales Leben und Materielle Welt* (1994), Berlin/New York.
- Heidegger, M.: *Sein und Zeit* (1926/1986), Tübingen.
- Hempel, C. G.: *Aspekte wissenschaftlicher Erklärung* (1965/1977), Berlin/New York.
- Husserl, E.: *Logische Untersuchungen* (1900/1993), Tübingen.
- Jeffrey, R.: *The Logic of Decision*, 2. Aufl. (1983), Chicago.
- Kambartel, F.: ‚Ein Versuch über das Verstehen‘ (1991), in: B. McGuinness u. a. (Hg.): *Der Löwe spricht...und wir können ihn nicht verstehen*, Frankfurt/M. 1991.
- Kamlah, W./Lorenzen, P.: *Logische Propädeutik* (2. Aufl.) (1973), Mannheim.
- Kaplan, D.: ‚Bob and Carol and Ted and Alice‘ (1973), in: J. Hintikka u. a. (Hg.): *Approaches to Natural Language* (1973), Dordrecht.
- ‚Dthat‘ (1978), in: P. Cole (Hg.): *Syntax and Semantics* (1978), New York.
- Keil, G.: *Kritik des Naturalismus* (1993), Berlin/New York..
- *Handeln und Verursachen* (2000), Frankfurt/M.
- Keil, G./Schnädelbach, H.: ‚Naturalismus‘ (2000), in: dies. (Hg.): *Naturalismus - Philosophische Beiträge*, Frankfurt/M.
- Kenny, A. J. P.: *Action, Emotion and Will* (1963), London.
- Kim, J.: ‚Epiphenomenal and Supervenient Causation‘ (1984), in: *Southern Journal of Philosophy* 22 (Suppl.).
- *Philosophy of Mind* (1996), Boulder/Co.
- Kripke, S.: *Name und Notwendigkeit* (1972/1981), Frankfurt/M.
- ‚A Puzzle about Belief‘ (1976), In: A. Margalit (Hg.): *Meaning and Use* (1979), Dordrecht.
- *Wittgenstein über Regeln und Privatsprache* (1982/1987), Frankfurt/M.
- Küppers, B.-O.: *Der Ursprung biologischer Information* (1986), München.
- Lafont, C.: *Sprache und Welterschließung* (1994), Frankfurt/M.
- ‚Ist Objektivität perspektivisch? Ein Vergleich zwischen Brandoms und Habermas‘ Konzeption von Objektivität‘ (2001), in: L. Wingert//K. Günther (Hg.): *Die*

Öffentlichkeit der Vernunft und die Vernunft der Öffentlichkeit – Festschrift für Jürgen Habermas, (2001), Frankfurt/M.

Lewis, D.: ‚An Argument for the Identity Theory‘ (1966), in: ders. (1983)

- ‚How to define theoretical Terms‘ (1970), in: ders. (1983)

- *Conventions* (1972a), Cambridge/Ma.

- ‚Psychophysical and Theoretical Identifications‘ (1972b), in: N. Block (Hg.) (1980).

- *Philosophical Papers Vol. I* (1983), Oxford.

Lewontin, R.: ‚Darwin’s revolution‘ (1983), in: *New York Review of Books* 30 (6).

Loar, B.: *Mind and Meaning* (1981), Cambridge.

Lorenz, K.: *Die Rückseite des Spiegels – Versuch einer Naturgeschichte menschlichen Erkennens* (1973), München.

Lycan, W.: *Judgement and Justification* (1988), Cambridge UP.

Lyons, W.: *Approaches to Intentionality* (1995), Oxford 1995.

Mackie, J.: ‚Causes and Conditions‘ (1965), in: Sosa/Tooley (1993).

Malcolm, N.: ‚Thoughtless Brutes‘ (1972-73), *Proceedings and Addresses of the American Philosophical Association* 46.

McDowell, J.: ‚In Defence of Modesty‘ (1987), in: Taylor (Hg.): *Michael Dummett – Contributions to Philosophy* (1987), Dordrecht.

- *Mind and World* (1994), Cambridge/Ma.

McGinn, C.: ‚The Structure of Content‘ (1982), in: A. Woodfield (Hg.): *Thought and Object* (1982), Oxford.

- *Mental Content* (1989), Oxford.

Mead, G. H.: *Geist, Selbst und Identität* (1934/1968), Frankfurt/M.

Mill, J. S.: *A System of Logic* (1843), London.

Millikan, R. G.: *Language, Thought and Other Biological Categories* (1984), Cambridge/Ma.

- ‚Biosemantics‘ (1989), in: dies. (1993).

- ‚In Defense of Proper Functions‘ (1989a), in: dies. (1993).

- *White Queen Psychology and other Essays for Alice* (1993), Cambridge/Ma.

- ‚Biofunctions: Two Paradigms‘ (1999), in: *Journal of Philosophy* 86.

Neander, N.: ‚What does Natural Selection Explain? A Correction to Sober‘ (1988), in: *Philosophy of Science* 55.

- ‚The Teleological Notion of ‚Function‘‘ (1991a), in: *Australian Journal of Philosophy* 69.

- ‚Functions as Selected Effects: The Conceptual Analyst’s Defence‘ (1991b), *Philosophy of Science* 58.

Papineau, D.: *Philosophical Naturalism* (1993), Oxford.

- Perry, J.: ‚Frege on Demonstratives‘ (1977), in: *Philosophical Review* 86.
- Pinker, S.: *Der Sprachinstinkt* (1994/1996), München.
- Putnam, H.: ‚Minds and Machines‘ (1960), in: ders. (1975).
- ‚The meaning of „Meaning“‘ (1975a), in: ders. (1975).
 - *Mind, Language and Reality – Philosophical Papers Vol. 2* (1975), Cambridge UP.
 - *Meaning and the Moral Sciences* (1978), London.
 - *Vernunft, Wahrheit und Geschichte* (1981/1982), Frankfurt/M.
 - ‚Meaning Holism‘ (1986), in: ders.: *Realism with a Human Face* (1990), Cambridge/Ma.
 - *Repräsentation und Realität* (1988/1991), Frankfurt/M.
- Pylyshyn, Z.: *Computation and Cognition: Toward a Foundation for Cognitive Science* (1984), Cambridge/Ma.
- Quine, W. O. v.: ‚Was es gibt‘ (1948/1979), in: ders. (1961/1979).
- *Word and Object* (1960), Cambridge/Ma.
 - *Ontological Relativity and other Essays* (1969), New York.
 - *Die Wurzeln der Referenz* (1974/1976), Frankfurt/M.
 - *Theorien und Dinge* (1981/1985), Frankfurt/M.
 - ‚Reply to Dagfinn Føllesdal‘ (1986), in: in: Hahn/Schilpp (1986).
 - *Pursuit of Truth* (Rev. Ed.) (1992), Cambridge/Ma.
- Roth, G.: *Das Gehirn und seine Wirklichkeit* (1994), Frankfurt/M.
- Rorty, R.: *Der Spiegel der Natur* (1979/1981), Frankfurt/M.
- ‚Pragmatismus, Davidson und der Wahrheitsbegriff‘ (1986/1990), in: Picardi/Schulte (1990).
- Russell, B.: ‚On Denoting‘ (1905), in: *Mind* 14.
- Ryle, G.: *The Concept of Mind* (1949), London.
- Schiffer, S.: *Meaning* (1972), Oxford.
- *Remnants of Meaning* (1987), Cambridge/Ma.
- Schnädelbach, H.: ‚Rationalität und Normativität‘ (1990), in: ders.: *Zur Rehabilitierung des animal rationale* (1992), Frankfurt/M.
- Schütz, A./Luckmann, T.: *Strukturen der Lebenswelt – Bd. 1* (1979) und *Bd. 2* (1984), Frankfurt/M.
- Searle, J.: ‚Proper Names‘ (1958), in: *Mind* 67.
- *Speech Acts* (1969), London.
 - ‚Mind, brains and programs‘ (1980), in: *Behavioral and Brain Sciences* 3 (3).
 - *Intentionalität. Eine Abhandlung zur Philosophie des Geistes* (1983/1987), Frankfurt/M.
 - *Geist, Hirn und Wissenschaft* (1984/1986), Frankfurt/M.

- ‚Indeterminacy, Empiricism and the First Person‘ (1987), in: *Journal of Philosophy* LXXXIV, 3.
- *The Rediscovery of the Mind* (1992), Cambridge/Ma.
- Sellars, W.: ‚Philosophy and the Scientific Image of Man‘ (1960), in: ders. (1963/1991)
- ‚Empiricism and the Philosophy of Mind‘ (1956), in: ders. (1963/1991)
- ‚Some Reflections on Language Games‘ (1954), in: ders. (1963/1991)
- *Science, Perception and Reality*, Atascadero/Cal.
- Smart, J. J. C.: ‚Sensations and Brain Prozesses‘ (1959), in: *Philosophical Review* 58.
- Sober, E.: *The Nature of Selection* (1984), Cambridge/Ma.
- Sosa, E./ Tooley, M. (Hg.), *Causation* (1993), Oxford.
- Stalnaker, R.: *Inquiry* (1984), Cambridge/Ma.
- Stegmüller, W.: *Wissenschaftlicher Erklärung und Begründung*, 2. Aufl. (1972), Berlin/Heidelberg/New York.
- Stich, S.: ‚Dennett on Intentional Systems‘ (1981), *Philosophical Topics* 12.
- *From Folk Psychology to Cognitive Science – The Case Against Belief* (1983), Cambridge/Ma.
- Strawson, P. F.: ‚Truth‘ (1950), in: ders. (1971).
- ‚On Referring‘ (1950a), in: ders. (1971)
- *Einzelding und logisches Subjekt* (engl.: *Individuals*) (1959/1972), Stuttgart.
- ‚Freedom and Resentment‘ (1962), in: ders.: *Freedom and Resentment and other Essays* (1974), London/NewYork.
- *Logico-Linguistic Papers* (1971), London.
- *Naturalismus und Skeptizismus* (1985/1987), Frankfurt/M.
- Tarski, A.: ‚Der Wahrheitsbegriff in formalisierten Sprachen‘ (1935), *Studia philosophica* 1.
- ‚Die semantische Konzeption der Wahrheit und die Grundlagen der Semantik‘ (1944), in: Skirbekk, G. (Hg.): *Wahrheitstheorien* (1980), Frankfurt/M.
- Taylor, C.: ‚Bedeutungstheorien‘ (1982/1988), in: ders.: *Negative Freiheit?* (1988), Frankfurt/M.
- Tugendhat, E.: *Vorlesungen zur Einführung in die sprachanalytische Philosophie* (1976), Frankfurt/M.
- Tugendhat, E./Wolf, U: *Logisch-semantische Propädeutik* (1983), Stuttgart.
- Turing, A.: ‚Computing Machinery and Intelligence‘ (1950), in: *Mind* 59.
- Walsh, D. M.: ‚The Scope of Selection: Sober an Neander on what Natural Selection Explains‘ (1998), in: *Australasian Journal of Philosophy* Vol. 76, No 2.
- Weizenbaum, J.: *Die Macht der Computer und die Ohnmacht der Vernunft* (1976/1978), Frankfurt/M.

- Williams, B.: *Ethics and the Limits of Philosophy* (1985), London.
- Wilson, N.: ‚Substances without Substrata‘ (1959), in: *Review of Metaphysics* 12.
- Winch, P.: *Die Idee der Sozialwissenschaft und ihr Verhältnis zur Philosophie* (1958/1974), Frankfurt/M.
- Winograd, T./Flores, F: *Understanding Computers and Cognition* (1987), Norwood/NJ.
- Wittgenstein, L.: *Philosophische Grammatik* (Werkausg. Bd. 4), Frankfurt/M 1984.
- *Philosophische Untersuchungen* (Werkausg. Bd. 1), Frankfurt/M. 1984
- *Über Gewissheit* (Werkausg Bd. 8), Frankfurt/M. 1984.
- Worrall, J.: ‚Structural Realism: The Best of both Worlds?‘ (1989), in: *Dialectica* 43.
- von Wright, G. H.: *Norm and Action* (1963), London.
- *Erklären und Verstehen* (1971/1974), Königstein/Ts.
- Wright, L.: ‚Functions‘ (1973), in: *Philosophical Review* 82.