

# Das Bewusstsein erklären

Dissertation

zur Erlangung des Grades  
Doktor der Philosophie (Dr. phil.)

des Fachbereichs Philosophie  
der Universität Konstanz

vorgelegt von  
Michael Schmitz

Erstgutachter: Prof. Dr. Gottfried Seebaß  
Zweitgutachter: PD Dr. Bernhard Thöle  
Tag der mündlichen Prüfung: 21.12.2005

*Zusammenfassung.* In dieser Arbeit wird die These vertreten, dass das Leib-Seele-Problem vor dem Hintergrund der Elimination der manifesten physischen Welt entsteht und die einzig befriedigende Reaktion darauf nur sein kann, diese Elimination wieder rückgängig zu machen und das Problem damit aufzulösen. Materialistische und dualistische Reaktionen werden als unzulänglich erwiesen. Sie sind nur verschiedene Formen eines ontologischen Fundamentalismus – Physikfundamentalismus und Bewusstseinsfundamentalismus – der zu letztlich sinnlosen metaphysischen Konstruktionen führt. Dagegen ist nach dem hier vertretenen ontologischen Pluralismus die Tatsache, dass die gesamte raumzeitliche Welt eine physikalische Mikrostruktur hat, mit einem naiven Realismus sowohl in Bezug auf die manifeste physische Welt als auch in Bezug auf das Bewusstsein vereinbar. Wir müssen nur beiden eine physikalische Mikrostruktur zuschreiben ohne sie mit dieser zu identifizieren.

Nachdem in der Einleitung die Elimination der manifesten physischen Welt in der Neuzeit als der metaphysische Hintergrund des Leib-Seele-Problems identifiziert und die durch die typischen Reaktionen darauf entstehende dilemmatische Ausgangssituation beschrieben worden ist, klärt das erste Kapitel, was Bewusstsein ist. Im zweiten Kapitel werden dann einige gängige Formulierungen des so genannten „Bewusstseinsrätsels“ vorgestellt. Es wird gezeigt, dass diese sich nicht in eine wirklich beantwortbare Frage transformieren lassen. Die Vorstellung, dass sich andere Makrophänomene wesentlich besser erklären lassen als das Bewusstsein, oder gar in vollständig transparenter Weise, wird als Illusion, als Artefakt der stillschweigenden Elimination der manifesten Welt erwiesen. Im dritten Kapitel wird die These, dass das Bewusstsein räumliche Eigenschaften hat, gegen verschiedene Formen der Skepsis verteidigt. Im vierten Kapitel beginnt die Auseinandersetzung mit der Identitätstheorie mit einer Diskussion sprachphilosophischer Positionen zu Identitätsaussagen, die die These begründet, dass die Vorstellung informativer Identitätsaussage sinnlos ist. Auf dieser Grundlage wird im fünften Kapitel die Vorstellung einer empirisch gehaltvollen so genannten „wissenschaftlichen Identifikation“ von Bewusstseinsphänomenen mit ihren neuronalen Korrelaten kritisiert. Im sechsten Kapitel wird gezeigt, dass sich das Problem mentaler Verursachung auflösen lässt, wenn man Bewusstseinsphänomenen eine physiologische und letztlich auch physikalische Mikrostruktur zuschreibt. In einer Schlussbetrachtung wird erläutert, warum der Vorschlag in dieser Arbeit eine Auflösung statt einer Lösung des Leib-Seele-Problems darstellt.

*Abstract.* The dissertation defends the thesis that the mind-body problem arises against the background of the elimination of the manifest physical world, and that the only satisfactory response to it is to take back that elimination and thus to dissolve the problem. Various materialist and dualist responses are shown to be inadequate. They are only different forms of ontological fundamentalism – physics fundamentalism and consciousness fundamentalism – that lead to ultimately meaningless metaphysical constructions. By contrast, on the ontologically pluralist view outlined here, the fact that the entire spatiotemporal world has a physical microstructure is compatible with a naive realism with regard to both the manifest physical world and to consciousness. We only need to ascribe physical microstructures to both while identifying neither with these structures.

After an introduction that identifies the elimination of the manifest physical world in modernity as the metaphysical background of the mind-body problem and describes the dilemmatic situation created by the typical reactions to it, the first chapter attempts to clarify what consciousness is. The second chapter then discusses some of the standard formulations of the so-called “puzzle” or “mystery” of consciousness. It is shown that these cannot be transformed into a genuine, answerable question. The notion that other macrophenomena can be explained much better than consciousness, perhaps even in a completely transparent way, is revealed to be an illusion, an artefact of the tacit elimination of the manifest physical world. In the third chapter, the thesis that consciousness has spatial properties is defended against various forms of skepticism. The fourth chapter begins the debate with the mind-body identity theory by discussing various positions on identity statements in the philosophy of language. This discussion leads to the conclusion that the idea of informative identity statements is meaningless. On this basis, the notion of an empirically contentful so-called “scientific identification” of consciousness phenomena with their neuronal correlates is criticized in the fifth chapter. The sixth chapter argues that it is possible to dissolve the problem of mental causation if a physiological and ultimately also a microphysical structure is ascribed to consciousness. In a concluding reflection it is explained what it means that the present suggestion amounts to a dissolution rather than a solution of the traditional mind-body problem.

<b>0</b>	<b><i>Einleitung</i></b>	<b>1</b>
<b>1</b>	<b><i>Erklären, was Bewusstsein ist</i></b>	<b>19</b>
1.1	Einige grundlegende Merkmale des Bewusstseins	22
1.2	Bewusstsein als innergeistige Relation	24
1.3	Bewusstsein als funktionales Bewusstsein	25
1.4	Bewusstsein als phänomenales Bewusstsein	26
<b>2</b>	<b><i>Das Bewusstsein verräteln</i></b>	<b>29</b>
2.1	Diachrone vs. synchrone Emergenz	32
2.2	Die diachrone Emergenz des Bewusstseins erklären	33
2.3	Die Möglichkeit des Bewusstseins erklären?	37
2.4	Das Bewusstsein als begrifflich notwendig erweisen?	38
2.5	„Funktionales“ vs. „phänomenales“ Bewusstsein erklären	39
2.6	Sind physische Makrophänomene reduktiv erklärbar?	43
2.7	Die Abwesenheit von Bewusstsein und anderen Makroentitäten	48
2.8	Fazit	50
<b>3</b>	<b><i>Das Bewusstsein im Raum</i></b>	<b>52</b>
3.1	Zeigt das Bewusstsein sich als wesentlich nichträumlich?	53
3.2	McGinns Begründung der Nichträumlichkeitsthese	54
3.3	Der intentionale Gehalt der basalen Selbstlokalisierung	56
3.4	Basales und begriffliches Selbstbewusstsein	57
3.5	Kritik der Vorstellung einer bloß abgeleiteten Lokalisierung	58
3.6	Kritik des moderaten Skeptizismus	60
<b>4</b>	<b><i>Das Bewusstsein der Identität</i></b>	<b>64</b>
4.1	Das Paradox der Identität	64
4.2	Frege: Sinn vs. Bedeutung	67
4.3	Russell: Referenz vs. Denotation	72
4.4	Wittgenstein: Die Elimination des Identitätszeichens	78
4.5	Kripke: Kontingenz <i>a priori</i> und Notwendigkeit <i>a posteriori</i>	81
4.6	Kontingenz <i>a priori</i>	82
4.7	Notwendigkeit <i>a posteriori</i>	86
4.8	„Informative“ Identitätsaussagen als empirisch motivierte Redefinitionen	88
4.9	Empirisch motivierte Redefinitionen und das dynamische <i>a priori</i>	92
<b>5</b>	<b><i>Die Identität des Bewusstseins</i></b>	<b>105</b>
5.1	Wissenschaftliche Identifikationen als notwendig <i>a posteriori</i>	105
5.2	Die Illusion der Notwendigkeit	110
5.3	Kripkes Argument gegen die psychophysische Identitätstheorie	113

5.4	Kritik der psychophysischen Identitätstheorie	117
5.5	1. Verteidigungsstrategie: opake Bezugnahme durch Beschreibungen	122
5.6	2. Verteidigungsstrategie: direkte opake Bezugnahme	128
5.7	Komposition oder Mikrostruktur statt Identität	136
6	<i>Das Bewusstsein wirkt</i>	<i>138</i>
6.1	Kritik des Epiphanomenalismus	141
6.2	Kritik der Vorstellung kausaler Überdeterminiertheit	147
6.3	(Auflösung des Problems der mentalen Verursachung	152
7	<i>Schlussbetrachtung: Lösung oder Auflösung?</i>	<i>159</i>
8	<i>Literatur</i>	<i>163</i>

## 0 Einleitung

*How can technicolor phenomenology arise from soggy grey matter?*

– Colin McGinn, Can we solve the mind-body problem?

*But what sort of chemical process could lead to the springing of existence of something non-physical? No enzyme can catalyse the existence of a spook!*

– J.J.C. Smart, Materialism

*A materialist thesis is advanced. But the thesis encounters difficulties; the difficulties take different forms, but they are always manifestations of an underlying deeper difficulty, namely, the thesis in question denies obvious facts that we all know about our own minds. And this leads to ever more frenzied efforts to stick with the materialist thesis and try to defeat the arguments put forward by those who insist on preserving the facts. After some years of desperate maneuvers to account for the difficulties, some new development is put forward that allegedly solves the difficulties, but then we find that it encounters new difficulties, only the new difficulties are not so new – they are really the same old difficulties. If we were to think of the philosophy of mind over the past fifty years as a single individual, we would say of that person that he is a compulsive neurotic, and his neurosis takes the form of repeating the same pattern of behavior over and over again.*

– John Searle, The Rediscovery of The Mind

Warum sind bestimmte Prozesse im Gehirn mit bewusstem Erleben verbunden und warum ist dieses Erleben so wie es ist und nicht anders? Diese Frage ist in der Tradition des Nachdenkens über Bewusstsein schon verschiedentlich gestellt worden, aber immer wieder, wie etwa Emil Dubois-Reynolds berühmter Ausruf "Ignorabimus" (1916) dokumentiert, als unlösbar angesehen worden. Diese scheinbare Unerklärlichkeit des Bewusstseins, die Aura der Rätselhaftigkeit, die dieses uns doch andererseits so vertraute Phänomen umgibt, hat im Verein mit den mit dem Erklärungsproblem eng verbundenen ontologischen und epistemologischen Problemen – welchen Platz kann das Bewusstsein in einem von der Physik geprägten Weltbild haben und wie können objektive Daten über das Bewusstsein erhoben werden? – dazu ge-

führt, dass das Bewusstsein im letzten Jahrhundert vom Mainstream der Psychologie und der an den Wissenschaften orientierten Philosophie nicht als Gegenstand ernsthafter Forschung angesehen und mitunter gar seine Existenz offen oder verdeckt geleugnet wurde. Der Psychologe Bernard Baars hat für diese Epoche die treffende Bezeichnung *age of avoidance* geprägt.<sup>1</sup>

In den letzten Jahrzehnten ist das Bewusstsein aber in einer zuerst nur langsamen, sich dann aber immer mehr beschleunigenden Bewegung, deren Beginn man, bei allen Vorbehalten, die bei solchen Datierungen angebracht sind, vielleicht auf die Publikation von Thomas Nagels berühmtem Aufsatz „What is it like to be a bat?“ im Jahre 1974 datieren könnte, von einem Tabu- zu einem Modethema geworden. Das *age of avoidance* ist im Wesentlichen vorbei, die von John Searle (1992) seinerzeit noch mehr geforderte als konstatierte „Wiederentdeckung des Geistes“ in vollem Gang. Nicht nur die philosophische Beschäftigung mit dem Bewusstsein hat sich wieder intensiviert und zu einer wahren Flut von Büchern und Aufsätzen geführt. Auch die empirische Forschung hat das Thema wieder entdeckt. Tausende von Wissenschaftlern beschäftigen sich mit dem Bewusstsein, Psychologen, so genannte Kognitionswissenschaftler und Neurowissenschaftler, die nach seinen neuronalen Korrelaten suchen. Positionen dazu, wie diese Suche vor sich gehen sollte, werden in programmatischen Aufsätzen bezogen (z.B. Crick & Koch 2003).

Bedeutet dies nicht, dass wir nun dabei sind, das philosophische Stadium der Beschäftigung mit dem Bewusstsein hinter uns zu lassen? Und ist es vor diesem Hintergrund sinnvoll, der Flut von philosophischen Büchern zum Bewusstsein noch ein weiteres hinzuzufügen? Dass die Beschäftigung mit dem Bewusstsein eine neue Blüte erreicht hat, bedeutet aber nicht, dass die Probleme, die ursprünglich zum Zeitalter der Vermeidung geführt haben, schon als überwunden gelten können. Ein Blick in die Literatur lehrt, dass dies nicht der Fall ist. Die Leib-Seele-Debatte hat zwar einen neuen Grad von technischer Komplexität und Raffinesse erreicht und eine Fülle von neuen – mitunter allerdings auch scholastisch anmutenden – Begriffsbildungen und Argumentationen hervorgebracht. In wesentlichen Hinsichten verläuft sie aber immer noch in althergebrachten Bahnen. Sie wird immer noch durch den traditionellen Gegensatz zwischen monistischen und dualistischen Positionen strukturiert, und die meisten der klassischen Theorien wie Parallelismus, Epiphänomenalismus, die Zwei-Aspekte-Lehre oder die Identitätstheorie werden immer noch in der einen oder ande-

---

<sup>1</sup> "Note to Participants" eines Internet-Seminars der ASSC (Association for the Scientific Study of Consciousness): <http://www.phil.vt.edu/ASSC/baars/baars3.html>

ren Form vertreten. Sogar der Panpsychismus hat in letzter Zeit eine Art Revival erlebt (z.B. Chalmers 1996, Strawson 1994). Dies liegt auch nicht, wie ein Skeptiker vermuten könnte, allein am Starrsinn und Dogmatismus der Philosophen. Auch die empirischen Wissenschaftler zeigen sich in ihrer Methodik und Theoriebildung von traditionellen philosophischen Haltungen und Problemen beeinflusst, etwa vom Problem der mentalen Verursachung (z.B. Wegner 2002, Haggard 2005) oder philosophischen Problemen der Wahrnehmung (z.B. Crick 1994). Es besteht also weiterhin Klärungsbedarf.

Woran liegt es aber, dass trotz aller technischen Raffinesse sowenig substanzieller Fortschritt zu erkennen ist? Searle hat dies auf die Tendenz zurückgeführt, unkritisch das traditionelle, im weitesten Sinne cartesianische Verständnis der Kategorien des Geistigen und des Physischen als einander wechselseitig ausschließende, fundamental verschiedene ontologische Bereiche zu übernehmen. Die Originalität dieser Diagnose besteht darin, dass Searle diese Tendenz auch und gerade bei den materialistischen Monisten erkennt, die sich als die schärfsten Kritiker des Cartesianismus verstehen. Auch sie bleiben den traditionellen Denkmustern verhaftet, indem sie die cartesianischen Prämissen akzeptieren, der cartesianischen Konklusion aber dennoch zu entgehen versuchen (Searle 1992: Kap. 1). Man kann diese Tendenz auch so beschreiben, dass ein Konditional des Inhalts akzeptiert wird, dass, wenn das Bewusstsein weder auf Dispositionen zu Verhalten oder funktionale Zustände reduzierbar, noch mit Gehirnzuständen identisch ist, irgendeine Form des Dualismus wahr sein muss. Die vorliegende Arbeit versucht, Searles Diagnose zu präzisieren und zu erweitern, die Kritik der monistischen und dualistischen Positionen fortzuführen und eine positive Theorie zu entwickeln, die dem Geist von Searles Konzeption entspricht, ohne an jedem ihrer Buchstaben zu kleben.

Warum ist es so schwer, das Denken in der Opposition zwischen dualistischen und monistischen Theorien zu überwinden? Eine der Hauptthesen dieser Arbeit lautet, dass das Leib-Seele-Problem in seiner herkömmlichen Form nur vor dem Hintergrund einer metaphysischen Haltung zu verstehen ist, die in der traditionellen Unterscheidung zwischen primären und sekundären Eigenschaften verkörpert ist, die die westliche Philosophie seit der Neuzeit wesentlich geprägt hat. Das Problem ist nur unzureichend als das beschrieben, wie die graue, weiche und feuchte Materie des Gehirns mit unserem reichhaltigen Bewusstseinsleben, unseren Farb-, Ton-, Geschmacks-, Geruchs- und Wärmeempfindungen, unserem Fühlen, Denken und Wol-



len verbunden sein kann; zu einem wirklichen Rätsel wird das Bewusstsein erst dadurch, dass der dominierenden Metaphysik zufolge das Gehirn in Wirklichkeit gar nicht grau, die Materie an sich nie farbig, heiß, scharf oder laut ist – jedenfalls nicht in dem Sinne, in dem der naive Alltagsverstand dies annimmt. Austen Clark charakterisiert diese Metaphysik wie folgt:

The qualities of size, figure (or shape), number, and motion are for Galileo the only real properties of objects. All other qualities revealed in sense perception – colours, tastes, odours, sounds, and so on – exist solely in the sensitive body, and do not qualify anything in the objects themselves. They are the effects of the primary qualities of things on the senses. Without the living animal sensing such things, these ‘secondary’ qualities (to use the term introduced by Locke) would not exist. For that reason they are subjective properties (see Burt 1954, p. 84).

Much of modern philosophy has devolved from this fateful distinction. While it was undoubtedly helpful to the physical sciences to make the mind into a sort of dustbin into which one could sweep the troublesome sensory qualities, this stratagem created difficulties for later attempts to arrive at some scientific understanding of the mind. In particular, the strategy cannot be reapplied when one goes on to explain sensation and perception. If physics cannot explain secondary qualities, then it seems that any science that can explain secondary qualities must appeal to explanatory principles distinct from those of physics. Thus are borne various dualisms. (Clark 1993: 6)

Nicht zufällig ist mit dieser Metaphysik neben den Namen von Galileo und Locke, von Boyle, Malebranche, Newton, Hume und anderen Philosophen und Naturforschern auch der von Descartes verbunden. Cartesianisch ist nicht nur die Opposition zwischen *res extensa* und *res cogitans* sowie eine gewisse subjektivistische Grundhaltung in der Erkenntnistheorie; cartesianisch ist auch die Metaphysik primärer und sekundärer Eigenschaften. Auch sie gehört zu dem Bündel eng miteinander verknüpfter philosophischer Grundhaltungen, die die mit der Neuzeit beginnende Periode bis in die Gegenwart hinein wesentlich bestimmen, und für die im heutigen Verständnis Descartes’ Name paradigmatisch steht. Und dieser zentrale Teil der neuzeitlichen Metaphysik wird auch von den ihrem Selbstverständnis nach dezidiertesten Anti-Cartesianern unter den zeitgenössischen materialistischen Monisten akzeptiert, ja normalerweise als selbstverständlich vorausgesetzt.<sup>2</sup>

---

<sup>2</sup> Damit soll nicht gesagt sein, dass der zeitgenössische Materialismus nur an diesem Punkt Cartesianischen Denkmustern verhaftet bleibt.

Aber warum genau ist diese Metaphysik für das Leib-Seele-Problem so ‚schicksalhaft‘, warum hat sie die Tendenz, das Bewusstsein als inhärent mysteriös und sich jeder Erklärung entziehend erscheinen zu lassen? Nun, wenn diese Metaphysik die richtige ist, wenn die Elimination der manifesten physischen Entitäten eine zwingende oder zumindest angemessene Konsequenz aus dem Erfolg der naturwissenschaftlichen, insbesondere der physikalischen Theorien ist, dann besteht das Problem darin, dass, falls die naturwissenschaftliche Erklärungen geistiger Phänomene ähnlich erfolgreich sein sollten, sich diese metaphysische Strategie, wie Clark sagt, nicht ein zweites Mal auf die geistigen Phänomene anwenden lässt. Wir schafften der Naturwissenschaft die manifesten Phänomene vom Hals, indem wir sie aus der Theorie der nichtgeistigen Realität eliminierten. Übrig blieben die entsprechenden Empfindungen und (vermeintlichen) Wahrnehmungen, die nunmehr als ein bloßer Schleier von Ideen rekonzeptualisiert wurden, der dem Bewusstsein den Zugang zur Welt, wie sie an sich ist, versperrt.<sup>3</sup>

Damit haben wir das Bewusstsein oder die Bewusstseine der Menschen und Tiere zu Inseln von Farben, Geräuschen und Gerüchen in einem Meer von farb-, geräuschs- und geruchsloser Materie gemacht. Wir haben den qualitativen Überschuss, den die manifesten Entitäten gegenüber den ihnen von der physikalischen Theorie zugeschriebenen wenigen abstrakten Eigenschaften wie Masse, Geschwindigkeit, Ladung usw. zu haben schienen, in gewissem Sinn, wie Clark sagt, als bloßen Schein im ‚Mülleimer‘ des Bewusstseins entsorgt. Wir können dieses Manöver nicht wiederholen, weil nach der Elimination des Bewusstseins noch nicht einmal ein Schein von Farbigkeit, von Tönen, von Wärme usw. übrig bliebe. Der qualitative Überschuss könnte nirgendwo mehr untergebracht werden. Ein solches Bild der Welt scheint letztlich unverständlich zu sein.

Wenn wir einen solchen durchgängigen eliminativen Physikalismus aber vernünftigerweise als Option ausschließen, auf der Elimination der manifesten physischen Entitäten jedoch beharren, wird das Bewusstsein in unserem Weltbild als der Ort dessen, was ich als den „qualitativen Überschuss“ bezeichnet habe, notwendigerweise eine ganz einzigartige Position bekommen. Es muss uns dann rätselhaft vorkommen, dass anscheinend bestimmte physische Phänomene, die neuronalen

---

<sup>3</sup> Hiermit ist auch der Zusammenhang zwischen der Metaphysik sekundärer Eigenschaften und der Cartesianischen subjektivistischen Epistemologie angedeutet. Es liegt nahe, den Konflikt zwischen der eliminativistischen Metaphysik und der Tatsache, dass man doch die manifesten Entitäten zu fühlen, zu sehen, zu hören und zu riechen scheint, dadurch aufzulösen, dass man nunmehr die Empfindungen zu den unmittelbaren Objekten der Wahrnehmung macht.

Korrelate des Bewusstseins, mit einem solchen qualitativen Überschuss verbunden sind. Wir konnten doch die physische Welt scheinbar restlos physikalisch erklären; warum will uns das beim Bewusstsein nicht gelingen?

Diese Frage stellt sich deshalb mit besonderer Schärfe, weil die Erwartung, dass sich das Bewusstsein prinzipiell genauso erklären lassen muss wie die unbewusste physische Welt, in bestimmter Hinsicht sehr berechtigt zu sein scheint. In der Geschichte der Physik sind sukzessive die Kräfte chemischer Phänomene und biologischer Organismen auf fundamentale physische Kräfte zurückgeführt worden.<sup>4</sup> Kein seriöser Wissenschaftler glaubt heutzutage noch an so etwas wie einen irreduziblen *élan vital*, und es gibt auch keine rein naturwissenschaftlichen Evidenzen für die Annahme, dass das Bewusstsein Kräfte hat, die sich nicht auf die fundamentalen physikalischen Kräfte zurückführen lassen. Die Neurowissenschaft unterstellt, dass sich das, ‚was das Bewusstsein tut‘ – das Heben von Armen, Formen von Sprachlauten usw., d.h. das ganze komplexe Verhalten, das wir normalerweise ganz selbstverständlich als bewusstseinsgesteuert verstehen – physiologisch und letztlich physikalisch erklären lässt.

Die Erfolgsgeschichte der Naturwissenschaft rechtfertigt also die Annahme, oder macht sie zumindest wahrscheinlich, dass sich Bewusstseinsphänomene in einer wesentlichen Hinsicht prinzipiell genauso erklären lassen wie physische.<sup>5</sup> Aber damit ist die Bühne für das Drama der Leib-Seele-Debatte bereitet, das sich nun abspielt. Der philosophische Materialist wird, gestützt auf diese plausible Annahme, versuchen, dem Bewusstsein auch die gleiche metaphysische Behandlung angedeihen zu lassen wie der manifesten physischen Welt. Damit ist er aber ständig der Ge-

---

<sup>4</sup> Eine übersichtliche Darstellung dieser Geschichte findet sich bei David Papineau (2002: 232ff).

<sup>5</sup> Natürlich bleibt bei einer Aussage dieser Art immer ein spekulatives Moment. Aber auch die Kritiker des Materialismus in der Leib-Seele-Debatte stellen sie normalerweise nicht in Frage, sondern versuchen, eine dualistische oder auf andere Weise bewusstseinsfundamentalistische Position zu formulieren, die mit ihr vereinbar ist. (Der Begriff „Bewusstseinsfundamentalismus“ wird später im Haupttext erläutert.) Ich werde in dieser Arbeit gleichfalls davon ausgehen, dass sie wahr ist. Wer dennoch gegenüber dieser Aussage skeptisch ist, könnte die vorgetragenen Überlegungen als eine Versuch dazu ansehen, wie mit einem solchen Faktum, wenn es denn eins wäre, philosophisch umgegangen werden könnte. Ferner ist noch anzumerken, dass sich natürlich verschiedene sehr berechtigte Fragen daran knüpfen, wie denn die Rede davon, dass Bewusstseinsphänomene – aber auch makroskopische physische – ‚prinzipiell‘ mikrophysikalisch erklärbar sind, zu verstehen ist. Ich werde hier diese Fragen ausklammern und davon ausgehen, dass diesen Redeweisen ein bestimmter, wenn auch vielleicht nur sehr begrenzter Sinn gegeben werden kann. Nichts folgt daraus aber unmittelbar für die weitere Frage, welche Erwartungen in Bezug auf die tatsächliche Reichweite etwa von physiologischen Erklärungen menschlichen Verhaltens realistisch sind. Schließlich sollten die Bemerkungen zur prinzipiellen Erklärbarkeit des Verhaltens nicht in Richtung der These einer Ersetzbarkeit des mentalistischen Vokabulars verstanden werden. Im Gegenteil, die These der nicht nur praktischen, sondern auch prinzipiellen, also ontologischen Unverzichtbarkeit dieses Vokabulars ist eine der zentralen Thesen dieser Arbeit.

fahr ausgesetzt, von der Bühne in den eben beschriebenen Graben der globalen physikalistischen Elimination zu stürzen, oder erliegt ihr gar. Diese Gefahr wird ihm aber normalerweise nicht bewusst sein, weil er dazu neigt, die Elimination der manifesten Welt für selbstverständlich zu halten. Dabei hilft ihm, dass er diese Elimination auch gar nicht als eine solche versteht. Er wird sich nicht als jemand beschreiben, der die Realität von Wärme, Licht oder Wasser usw. leugnet, sondern als ein wissenschaftlich aufgeklärter Philosoph, der glaubt, dass die Wissenschaft entdeckt hat, was diese Phänomene ‚in Wirklichkeit‘, ihrem eigentlichen Wesen oder ihrer eigentlichen Natur nach sind. Wärme etwa ist für ihn in Wirklichkeit mit der Bewegung von Molekülen identisch. Aufgrund seiner Voraussetzung, dass die Wissenschaft Bewusstseinsphänomene prinzipiell genauso erklären kann wie Wärmephänomene, scheint ihm nun, dass es möglich sein muss und auch geboten ist, Bewusstseinsphänomene ihrerseits mit neuronalen Phänomenen zu identifizieren. Dass seine Kritiker ihm bei seinem Unterfangen, diese Identifikation in begrifflich kohärenter Weise durchzuführen, immer wieder Aporien und Widersprüche nachweisen können, wird ihn kaum irre machen, da er die wissenschaftliche Vernunft auf seiner Seite glaubt – und in einer wichtigen Hinsicht hat er damit ja auch recht. Seine Opponenten, die ihrerseits zumindest darin Recht haben, dass sie seine vermeintliche Identifikation von Bewusstsein mit etwas, das wir doch gemeinhin davon unterscheiden, als den Versuch seiner Elimination erkennen, erscheinen ihm wie Menschen, die hinter dem wissenschaftlichen Fortschritt zurückgeblieben sind. Die Insistenz auf der Nichtidentität von Bewusstseins- und Gehirnphänomenen kommt ihm vor wie der Glaube an Phlogiston, den *élan vital*, oder gar an Hexen oder Geister. So kommt es zu dem von Searle (vgl. das Motto oben und Searle 1992: Kap. 1) metaphorisch als zwanghaft beschriebenen Verhalten: Er versucht auf immer neue Weise, das ihm als ein Dorn im Auge der wissenschaftlichen Vernunft erscheinende Bewusstsein loszuwerden. Darin wird er ferner von seiner Überzeugung bestärkt, dass in einer Welt, in der die Kausalität des Bewusstseins durchgängig physikalisch erklärt werden kann, unsere Alltagsüberzeugung, dass das Bewusstsein kausal wirksam ist, nur durch seine Identifikation mit physiologischen und letztlich mit physikalischen Entitäten verständlich gemacht und damit gerettet werden kann. Schließlich fühlt er sich noch dadurch in seiner Haltung bestätigt, dass sein typischer Opponent, der Dualist, sich ebenfalls dazu gedrängt fühlt, extreme und unplausible Positionen einzunehmen.

Dieser wird vor allem oft das Gefühl haben, dem Materialisten darin Recht geben zu müssen, dass die Kausalität des Bewusstseins nur um den Preis seiner Identifikation mit physikalischen Entitäten zu bewahren ist. Schon Descartes spürte bekanntlich den Stachel des Problems der mentalen Verursachung. Aber dieser Preis kann dem Dualisten durchaus mit einigem Recht als einer erscheinen, der zwar hoch ist, den zu zahlen aber letztlich nicht unvernünftig ist. Kein Preis, so kann der Dualist sich sagen, ist zu hoch, um die unverständliche Idee abzuwehren, dass das Bewusstsein mit etwas anderem identisch ist oder die absurde Behauptung, dass es gar nicht existiert. Der Dualismus ist die natürliche Position, gegeben die auch vom Dualisten normalerweise als selbstverständlich akzeptierte Elimination der manifesten physischen Welt. Durch diese Elimination und die Tatsache, dass das Bewusstsein nicht auch noch eliminiert werden kann, hat das Bewusstsein einen Sonderstatus in der Ontologie erhalten, mag der Materialist sich noch so sehr dagegen sträuben. Ferner scheint dem Dualisten die Idee, dass Bewusstseinsphänomene räumliche Eigenschaften haben könnten, als sinnlos. Und zur Stützung dieser Position kann er darauf verweisen, dass wir tatsächlich mit einer Aussage wie der, dass ein Bewusstseinszustand eine bestimmte Ausdehnung hat, keine anschauliche Vorstellung verbinden können.

Der Dualist hat einige Verwandte, die streng genommen keine Dualisten sind, aber doch zur selben Familie gehören, die ich als die Familie der Bewusstseinsfundamentalisten bezeichnen möchte. Nicht-dualistische Bewusstseinsfundamentalisten wollen Bewusstseinsentitäten oder so genannte Protobewusstseinsentitäten *als solche* zum Gegenstand physikalischer Theorien machen. D.h. diese Theorien sollen entsprechend erweitert werden. Da diese Theoretiker damit sowohl unzweideutige Bewusstseinsrealisten als auch Physikalisten sein wollen, ist ihre Theorie auch mit dem Physikalismus verwandt. Weil sie aber genau wie der Dualist und im Gegensatz zum Identitätstheoretiker glauben, dass die bloße Existenz des Bewusstseins zeigt, dass die Ressourcen der gegenwärtigen Physik nicht ausreichen, damit die Physik Ontologie sein kann – also eine ontologisch vollständige Beschreibung der Welt in rein physikalischer Begrifflichkeit gegeben werden kann –, rechne ich sie doch eher dem dualistischen Lager zu.<sup>6</sup> Das in der gegenwärtigen Debatte wohl wichtigste Beispiel für eine solche Theorie ist der so genannte Russellsche Monismus.<sup>7</sup> Russellsche Mo-

---

<sup>6</sup> Dies ist aber eine rein taxonomische Frage, an der nichts Substanzielles hängt.

<sup>7</sup> Dieser Terminus stammt von David Chalmers (2002) und ist durch Bertrand Russell (1927) inspiriert. Russells Position ist aber dem speziellen Kontext des neutralen Monismus verhaftet, den ich hier ausklammern möchte. Zu den Vertretern des Russellschen Monismus gehören neben Chalmers selbst

nisten glauben, dass die gegenwärtige Physik die Welt nur in einer rein strukturellen, relationalen Begrifflichkeit beschreibt. Die Bewusstseinsentitäten sollen sozusagen die damit scheinbar innerhalb der Physik klaffenden ontologischen Leerstellen mit qualitativem Gehalt füllen.

Auch in diesem Fall kann man die Motivation für diese Auffassung vor dem Hintergrund der Elimination der manifesten physischen Welt durchaus nachvollziehen. *In abstracto* klingt die Idee, den qualitativen Überschuss direkt auf der physikalischen Ebene zu verorten und damit die Zurückweisung der Identitätstheorie mit einem durchgängigen Physikalismus zu vereinbaren, sehr verlockend. Sie ist allerdings äußerst spekulativ, da sie nur durch rein metaphysische Überlegungen motiviert wird. Es sprechen meines Wissens keinerlei physikalische Überlegungen dafür, Bewusstseinsphänomene in die Physik zu inkorporieren. Und die Physik sollte den Physikern überlassen werden, die sich nach Maßgabe der für die Physik charakteristischen Erkenntnisinteressen für bestimmte Theorien entscheiden. Der Russellsche Monismus scheint somit ein Artefakt der Tendenz zu sein, die Physik mit dem Anspruch zu belasten, Ontologie zu sein. Ferner muss jede Theorie dieser Art schnurstracks zum Panpsychismus führen. Die Physik, insbesondere die mikrophysikalischen Grundlagentheorien wie die Quantenmechanik und die Stringtheorie, ist schließlich dadurch ausgezeichnet, dass ihr theoretischer Apparat auf buchstäbliche alle Phänomene der raumzeitlichen Welt anwendbar ist. Nur deshalb hat die Vorstellung, sie könne Ontologie sein, ja überhaupt eine gewisse Scheinbarkeit. Bewusstsein ist aber nach unserem gewöhnlichen Verständnis ein sowohl zeitlich als auch räumlich äußerst begrenztes Phänomen. In den die menschliche Vorstellungskraft übersteigenden Weiten des Universums findet sich nur auf der winzigen Erde mit Sicherheit Bewusstsein, und dies nach kosmischen Maßstäben auch erst seit einem Augenblick. Der Russellsche Bewusstseinsfundamentalist mutet uns also den allem Anschein nach unüberprüfaren Gedanken zu, dass Bewusstsein in der physischen Welt allgegenwärtig ist und löst damit den Bewusstseinsbegriff von unserer ihm allein Sinn gebenden gewöhnlichen Praxis der Zuschreibung von Bewusstseinszuständen völlig ab.<sup>8</sup> Dies ist der Preis, der für diese Form des Fundamentalismus zu zahlen ist.

---

(siehe etwa auch Chalmers 1996) auch z.B. Michael Lockwood (1989) und Galen Strawson (1994, 2002).

<sup>8</sup> Die Behauptung, dass die Präsenz von Bewusstsein doch ohnehin unüberprüfbar sei, wäre meiner Ansicht nach an dieser Stelle keine angemessene Replik. Denn nach den Maßstäben unserer gewöhnlichen Praxis trifft dies ja nicht zu, und es scheint mir nicht angemessen zu sein, eine extreme metaphysische Position durch eine extreme Form des Skeptizismus zu rechtfertigen. Allerdings zeigt diese

Ich hoffe, dass der Leser nach diesem einleitenden Überblick über einige grundlegenden Positionen, Probleme und argumentativen Züge in der Leib-Seele-Debatte die Motivation nachvollziehen kann, einen Standpunkt jenseits dieser Standardauffassungen zu suchen, die uns in je eigener Weise mehr oder minder unplausibel bis unverständliche Aussagen zumuten, und dass die Diagnose, dass sich diese Probleme lösen oder auflösen lassen, wenn wir den fatalen ersten Schritt der Elimination der manifesten physischen Welt rückgängig machen, zumindest eine gewisse Anfangsplausibilität bekommen hat. Die Rücknahme dieses Schritts kann man auch als die Überwindung einer gewissen Form von ontologischem Fundamentalismus beschreiben.

Alle Standardpositionen sind verschiedene Formen eines ontologischen Fundamentalismus. Ihr Streit ist ein Streit darüber, welche Entitäten als fundamental anzusehen sind. Der Physikalist ist Physikfundamentalist, für ihn sind nur die gewöhnlichen physikalischen Entitäten fundamental; der Bewusstseinsfundamentalist meint, die Bewusstseinsphänomene auf der gleichen Ebene der Fundamentalität anzusiedeln zu müssen wie die physikalischen. Wiederum ist an beiden Positionen etwas Richtiges. Physikalische Entitäten sind in dem schon angedeuteten Sinn fundamental, dass alle Objekte der raumzeitlichen Welt (mikro)physikalische Strukturen und Eigenschaften aufweisen. Das Bewusstsein ist dagegen fundamental für die menschliche Existenz und Intentionalität; für Wahrnehmen und Denken, Handeln und Wollen, Werte und Bedeutung.<sup>9</sup> Es sollte aber offensichtlich sein, dass diese Aussagen einander nicht widersprechen. Hier ist von Fundamentalität in verschiedenen Bedeutungen dieses Wortes die Rede. Nur wenn man diese Bedeutungen nicht auseinander hält, kann man meinen, dass hier ein Konflikt vorliegt. Man muss weder die Allgegenwärtigkeit physikalischer Entitäten leugnen, um die Realität des Bewusstseins

---

Replik wieder, dass der Bewusstseinsfundamentalismus und der epistemische Subjektivismus sich sehr gut ergänzen und eine natürliche Allianz bilden. Nicht umsonst ist Descartes der Urvater von beiden! Natürlich ist es auch nicht möglich, diese beiden Positionen argumentativ zu widerlegen, wenn gar kein *common ground* zugestanden wird. Ich erhebe hier auch gar nicht den Anspruch, dies zu tun, sondern wäre schon zufrieden, sollte es mir gelingen, eine kohärente alternative Position zu formulieren, die die Extreme vermeidet und sowohl den alltäglichen als auch den wissenschaftlichen *common sense* respektiert. Wenn dies gelingt, ließe sich dann allerdings auch argumentieren, dass die extremen Positionen weder für unsere alltägliche, noch für unsere wissenschaftliche Praxis eine Rolle spielen und letztlich nur ein leeres Spiel mit Begriffen treiben, die durch diese Praxen ihren Sinn bekommen. Die damit angesprochenen grundsätzlichen Fragen zu den Zielen und Methoden der Philosophie sollen aber an dieser Stelle nicht weiter explizit verfolgt werden, obwohl diese aus einer bestimmten Haltung zu diesen Fragen heraus geschrieben ist und diese in ihr auch erkennbar sein sollte. Vgl. aber die Schlussbetrachtung.

<sup>9</sup> Zum Verhältnis von Bewusstsein und Werten vgl. Siewert (1998: Kap. 9), zu dem von Intentionalität und Bewusstsein Searle (1992), Strawson (1994) und Schmitz (2004).

und seine fundamentale Bedeutung für die menschliche (und tierische) Existenz zu bewahren, noch umgekehrt diese Bedeutung herunterspielen<sup>10</sup> oder gar implizit oder explizit die Realität des Bewusstseins leugnen, um dem umfassenden Anspruch der Fundamentalphysik gerecht zu werden.

Der Schlüssel zu einer Lösung oder Auflösung des Leib-Seele-Problems besteht somit meiner Ansicht nach darin, sowohl die verschiedenen Bedeutungen von Fundamentalität zu unterscheiden als auch zu akzeptieren, dass Fundamentalität weder in der einen, noch in der anderen Bedeutung und schon gar nicht in einer der vielen weiteren Bedeutungen, die es hat, mit *Realität* gleichzusetzen ist. (Beispiele für Fundamentalität in anderen Lesarten: „die fundamentale Bedeutung der Trennung von Kirche und Staat für die moderne französische Gesellschaft“ oder „die fundamentalen Gegenstände, Denkweisen und Methoden der Informatik.“) Das Bewusstsein ist zweifellos real, obwohl es nicht fundamental im Sinne der Fundamentalität der physikalischen Entitäten ist und auch – was sowohl im Hinblick auf seine Realität als auch seine Fundamentalität auf dasselbe hinausläuft – nicht mit ihnen oder ihren Aggregaten identisch ist. Es sollte damit auch klar geworden sein, wie eine emphatische Insistenz auf der Verschiedenheit von Bewusstseinsentitäten von anderen, physikalischen und manifesten physischen Entitäten mit der gleichzeitigen Zurückweisung des Dualismus zu vereinbaren ist.

Wie soll aber nun die alternative Auffassung jenseits von Monismus und Dualismus konkret aussehen? Der entscheidende Schritt besteht, wie schon deutlich gemacht wurde, darin, die Realität der manifesten physischen Welt anzuerkennen. Man kann dies so beschreiben, dass damit der ontologische Monismus oder Dualismus durch einen ontologischen Pluralismus oder Polyismus (Searle 1992) ersetzt wird. Um nur, neben Wärme, Farben usw., einige beliebig herausgegriffene Entitäten zu nennen: Auch Wasser, Blumen, Stühle, Mittagessen, Vorstandssitzungen und Bücher sind uneingeschränkt real, ohne doch mit (mikro)physikalischen Entitäten identisch zu sein. (Diese Behauptung mag trivial klingen und ohne den letzten Teilsatz wäre sie es auch; aber man muss sehen, dass viele Philosophen implizit oder explizit die Realität gängiger makroskopischer Entitäten geleugnet haben; auf andere Weise als die physikalistische hat dies natürlich auch die idealistische Tradition getan, auf die einzugehen den Rahmen dieser Arbeit sprengen würde.) Die Beziehung zwischen Makro- und Mikroentitäten ist stattdessen *Komposition*. Gewöhnliches Wasser

---

<sup>10</sup> Owen Flanagan (1992) bezeichnet die Tendenz, die Bedeutung des Bewusstseins herunterzuspielen als *consciousness inessentialism* und kritisiert sie in verschiedenen ihrer Erscheinungsformen.



z.B. besteht bekanntlich hauptsächlich aus  $\text{H}_2\text{O}$ - und einem geringen Anteil von  $\text{D}_2\text{O}$ -Molekülen – neben den so genannten ‚Verunreinigungen‘. Äquivalent kann man auch sagen, dass Wasser eine bestimmte molekulare – und natürlich auch subatomare – *Mikrostruktur* hat. Was an der physikalistischen These richtig ist, kann man entsprechend ausdrücken, indem man sagt, dass die gesamte raumzeitliche Welt aus physikalischen Entitäten besteht oder alle Entitäten eine (mikro)physikalische Struktur haben.

Während es natürlich ist zu sagen, dass Wasser aus  $\text{H}_2\text{O}$  besteht, wird man sich, selbst wenn man den Fall des Bewusstseins erst einmal ausklammert, fragen, woraus denn etwa ein Mittagessen besteht. Die Antwort darauf ist, dass ein Mittag ein – wenn auch vage bestimmter – Zeitraum ist, der als solcher aus kleineren Zeiteinheiten, aus Stunden, Minuten, Sekunden usw. besteht. Essen ist ein Vorgang der Nahrungsaufnahme, der aus Teilvorgängen wie Kauen, Schlucken usw. besteht, welche wiederum aus biologischen, chemischen und letztlich physikalischen Vorgängen bestehen. Diese Vorgänge oder Prozesse involvieren entsprechende Entitäten wie Mund, Zähne usw., die ihrerseits aus Mikroentitäten bestehen. Der Kompositionsbegriff lässt sich also nicht nur auf Dinge anwenden, sondern auch auf Zeiträume und Vorgänge. Dies entspricht durchaus dem *common sense*. Wir sagen auch im Alltag, dass ein Mittagessen aus vier Gängen besteht oder ein Drama aus fünf Akten.

Warum ist es wichtig, dass die Beziehung zwischen Makro- und Mikroentitäten Komposition ist und nicht Identität? Jeder wird zugestehen müssen, dass es einen Sinn von „meinen“ gibt, in dem derjenige, der sagt „Dies ist Wasser“ nicht dasselbe meint wie jemand, der sagt „Dies ist  $\text{H}_2\text{O}$ “ – und zwar auch dann, wenn wir von der schon erwähnten Tatsache absehen, dass Wasser gewöhnlich auch  $\text{D}_2\text{O}$ , also so genanntes ‚schweres Wasser‘ oder Deuteriumoxid, und die besagten ‚Verunreinigungen‘ enthält. Schließlich haben die Menschen auch schon etwas mit Ausdrücken wie „Wasser“ gemeint, bevor Lavoisier und andere das  $\text{H}_2\text{O}$ -Molekül entdeckten. Ich werde später ausführlich dafür argumentieren, dass der Sinn von „meinen“ oder auch „Bezug nehmen“, für den gilt, dass mit „Wasser“ und „ $\text{H}_2\text{O}$ “ Verschiedenes gemeint wird, auch der primäre Sinn von „meinen“ oder „Bezug nehmen“ ist und in gewissem Sinn sogar der einzige. Die grundlegende Frage ist folgende: Gelingt es den Menschen auch schon vor der Entdeckung des  $\text{H}_2\text{O}$ -Moleküls, sich erfolgreich auf Wasser zu beziehen? Da Begriffe manifester physischer Entitäten wie „Wasser“ sehr eng mit der Fähigkeit verknüpft sind, diese wahrzunehmen, ist diese Frage wiederum

sehr eng mit der Frage verbunden, ob uns die Wahrnehmung *als solche* einen Zugang zur Realität eröffnet. Wenn wir, wie hier vorgeschlagen wird, aufhören, die Realität der manifesten Welt zu leugnen, ist klar, dass beide Fragen zu bejahen sind. Die Menschen hatten auch schon zu vorwissenschaftlichen Zeiten einen intentionalen Zugang zur Realität und haben immer noch einen solchen, von der Wissenschaft unabhängigen Zugang. Darauf zu insistieren ist nicht ‚unwissenschaftlich‘ und auch mit einem vernünftigen wissenschaftlichen Realismus problemlos vereinbar; es bedeutet allerdings, sich gegen die Überhöhung der Wissenschaft zur Ontologie zu wehren. Wenn die Wahrnehmung aber einen Zugang zur Realität, z.B. zu Wasser eröffnet, und der Wahrnehmende mit ‚Wasser‘ primär das meint, was sich ihm in der Wahrnehmung eröffnet, kann Wasser nicht mit  $H_2O$  identisch sein, da der Wahrnehmende in der Wahrnehmung offensichtlich keinen intentionalen Zugang zu den  $H_2O$ -Molekülen hat, aus denen Wasser unter anderem besteht. Diese Elemente der Realität eröffnen sich ihm in der Wahrnehmung nicht. Zugang zu ihnen kann er nur über die wissenschaftliche Theoriebildung bekommen. Deshalb ist der Realismus in Bezug auf die manifeste physische Welt unvereinbar mit Identitätsbehauptungen wie der Behauptung der Identität von Wasser und  $H_2O$ , und deshalb ist es im Argumentationskontext dieser Arbeit wesentlich, darauf zu insistieren, dass diese Beziehung in Wirklichkeit die Kompositionsbeziehung ist. Eine genaue parallele Überlegung lässt sich natürlich auch in Bezug auf den Versuch der Identifikation von Bewusstseins- und Gehirnphänomenen anstellen: In der Introspektion und im Hineinversetzen in das Bewusstsein anderer eröffnet sich uns ein Bereich der Realität, aber diese Erkenntnisvermögen eröffnen uns keinen Zugang zu Gehirnphänomenen. Die hiermit eingeführte Argumentform wird in den Kapiteln 4 und 5 noch ausführlich diskutiert und gegen Einwände verteidigt werden.

Nach der hier vertretenen Auffassung verbindet die Kompositionsbeziehung auch Bewusstseins- mit Gehirn- und damit letztlich auch mit chemischen und physikalischen Entitäten. Bewusstseinsphänomene haben physiologische, chemische und physikalische Mikrostrukturen. Dieser Gedanke ist ungewohnt und mag zuerst kontraintuitiv erscheinen, aber ich glaube, dass er doch eine ernsthafte Prüfung verdient, weil er verspricht, einen Ausweg aus den eingangs umrissenen Dilemmata der Leib-Seele-Debatte zu eröffnen. Wenn wir die Beziehung zwischen Makro- und Mikroebenen durchgängig als Kompositionsbeziehung konzeptualisieren, können wir sowohl den empirischen Evidenzen gerecht werden, die dafür sprechen, dass die Kau-

salität des Bewusstseins prinzipiell genauso naturwissenschaftlich erklärt werden kann wie die von Wasser, Wärme oder Licht, als auch die metaphysischen Aporien der traditionellen Leib-Seele-Debatte hinter uns lassen.

Wenn man dagegen einwenden will, dass es *a priori*, aus rein begrifflichen Gründen, ausgeschlossen sei, dass Bewusstseinsphänomene eine physische Struktur haben und damit unter anderem auch räumliche Eigenschaften, so möchte ich darauf erwidern, dass meiner Ansicht nach die Autorität einer Begriffsanalyse allein nicht hinreichend sein kann, um den skizzierten Vorschlag zu verwerfen. Das Cartesianische Verständnis von Körper und Geist als *res extensa* und *res cogitans* ist in einer bestimmten historischen Situation entstanden und reflektiert bestimmte, empirische und metaphysische Annahmen und Grundhaltungen, von denen einige eingangs grob skizziert wurden. Es kann keinen Anspruch auf ewige Gültigkeit erheben, sondern muss gegebenenfalls den Fortschritten der empirischen Wissenschaften und der Philosophie angepasst werden. Wenn plausibel gemacht werden kann, dass nur eine Revision der überkommenen Kategorien des Geistigen und des Physischen es möglich macht, zu einer sowohl empirisch als auch metaphysisch adäquaten Auffassung zu kommen, ist es vernünftig, diese Revision vorzunehmen. Dies soll in dieser Arbeit versucht werden. Ferner wird im 3. Kapitel, das eigens der Frage der Räumlichkeit des Bewusstseins gewidmet ist, für die These argumentiert, dass auch wichtige Intuitionen zugunsten der Räumlichkeitsthese mobilisiert werden können.

Das Bewusstsein erscheint nicht nur in explanatorischer und metaphysischer Hinsicht als problematisch und rätselhaft, sondern auch in epistemologischer und bedeutungs- und intentionalitätstheoretischer sowie – im Kontext des Problems von Freiheit und Verantwortlichkeit – in moralischer. Der eigentliche Gegenstand dieser Arbeit ist die Erklärung und Ontologie des Bewusstseins. Dabei ist eine wesentliche These die, dass die explanatorischen und ontologischen Fragen auch sehr eng miteinander zusammenhängen. Zwar ist die Vorstellung eines ‚Bewusstseinsrätsels‘ oder einer ‚Erklärungslücke‘ in einigen Fällen von Autoren in die Debatte eingeführt und propagiert worden, die, wie z.B. Joseph Levine (1983, 1987) und Colin McGinn (1991, 2002), gleichzeitig an einem ontologischen Physikalismus festhalten wollen. Aber letztlich drückt sie meiner Ansicht nach vor allem ein Unbehagen an der physikalistischen Ontologie aus. Im dritten Kapitel werde ich dafür argumentieren, dass wir keine klare Vorstellung von einer Erklärung haben, die die Antwort auf die Rätselfrage liefern könnte. Sie kann nur die Funktion haben, uns klarzumachen, dass wir

das Bewusstsein als ein eigenständiges und begrifflich irreduzibles Phänomen in die Ontologie aufnehmen müssen.

Man könnte nun nicht ohne Plausibilität auch entsprechend für einen engen Zusammenhang zwischen den grundlegenden epistemologischen und den explanatorischen und damit letztlich auch den ontologischen Fragen argumentieren. Es wurde schon angedeutet, dass die Cartesianische Metaphysik und Epistemologie eng miteinander zusammenhängen und sich in natürlicher Weise ergänzen. Gleiches gilt für die materialistische Metaphysik und eine Epistemologie, die die einzig zulässige epistemische Grundlage wissenschaftlicher Theoriebildung in der Beobachtung äußeren, ‚öffentlichen‘ Verhaltens sieht. Schließlich hängt auch die Argumentation dieser Arbeit eng mit der epistemologischen Haltung eines direkten Realismus in der Wahrnehmungstheorie zusammen oder setzt diese sogar voraus. Es ist aber nötig, den Gegenstandsbereich der Untersuchung, die sonst leicht uferlos würde, einzuschränken, und ich werde deshalb diese Frage hier ausklammern und den direkten Realismus als Voraussetzung behandeln.

Diese Entscheidung wird dadurch erleichtert, dass der direkte Realismus in jüngerer Zeit eine Renaissance erlebt (z.B. Searle 1983, McDowell 1994) und von einer wachsenden Anzahl von Philosophen verteidigt wird. Ferner wird diese Position auch gestärkt, wenn gezeigt werden kann, dass unter ihrer Voraussetzung eine befriedigende Antwort auf die Leib-Seele-Problematik gegeben werden kann. Bis auf wenige Bemerkungen ausgeklammert werden muss auch die extensive Literatur zur Unterscheidung zwischen sekundären und primären Eigenschaften selbst, die selbst wiederum eng mit der Frage des direkten Realismus zusammenhängt.<sup>11</sup> Man kann letztlich wohl kaum direkter Realist in Bezug auf die Wahrnehmung sein, wenn Wärme, Töne, Farben usw. als externe wahrnehmbare Entitäten gar nicht existieren. Auch hier kann man aber sagen, dass es ein starkes Argument für den Realismus bezüglich der manifesten Welt ist, wenn er eine Lösung oder Auflösung des Leib-Seele-Problems ermöglicht. Da die zentralen Fragen der theoretischen Philosophie allesamt sehr eng miteinander zusammenhängen, aber hier unmöglich allesamt behandelt werden können, bleibt nur, an einem Punkt anzusetzen und zu zeigen zu versuchen, dass von diesem Punkt aus ein befriedigendes Gesamtbild zumindest entworfen werden könnte. Dagegen scheint es mir wegen dieses engen Zusammenhangs der

---

<sup>11</sup> Auch hier gibt es Anzeichen eines Wandels zugunsten eines Realismus bezüglich des Manifesten, der z.B. von Barry Stroud in seiner umfassenden Untersuchung *The Quest for Reality: Subjectivism and the Metaphysics of Colour* (2000) verteidigt wird. Siehe auch Campbell 1997.

zentralen Fragen nicht möglich zu sein, eine Position zu einer dieser Fragen einzunehmen ohne sich implizit oder explizit auch auf Antworten auf andere festzulegen oder diese zumindest nahe zu legen.<sup>12</sup>

Als den archimedischen Punkt der hier vorgetragenen Überlegungen könnte man die Auseinandersetzung mit der Identitätstheorie bezeichnen. Wenn die so genannten theoretischen Identifikationen von makroskopischen mit mikroskopischen Entitäten sich als solche verständlich machen ließen und sich nicht, wie hier behauptet wird, tatsächlich als verkappte Eliminationen der ersteren zugunsten der letzteren herausstellen würden, wäre die Hauptargumentation dieser Arbeit, die von dieser Voraussetzung ausgeht, gegenstandslos. An dieser zentralen Stelle ist es deshalb nicht möglich, die mit der Identitätstheorie wesentlich verknüpften sprachphilosophischen Fragen auszuklammern. Deshalb soll im 4. Kapitel die Auseinandersetzung mit der Identitätstheorie durch eine Analyse der grundlegenden sprachphilosophischen Probleme des Konzepts empirisch informativer Identitätsaussagen vorbereitet werden. Auf der Grundlage einiger klassischen Überlegungen von Gottlob Frege, Bertrand Russell und Ludwig Wittgenstein, und in kritischer Auseinandersetzung mit der neueren Position von Saul Kripke soll gezeigt werden, dass dieses Konzept aus ganz allgemeinen begrifflichen Gründen scheitert, die nichts Wesentliches mit der speziellen Frage einer Identifikation von Bewusstseinsphänomenen zu tun haben. Im 5. Kapitel wird diese Auseinandersetzung dann mit einer ausführlichen Analyse der Debatte um theoretische Identifikationen im Allgemeinen und die psychophysische Identitätstheorie im Besonderen abgeschlossen.

In dieser Einleitung wurden die Grundpositionen in der Leib-Seele-Debatte idealtypisch dargestellt, um einige wesentlichen Muster in dieser Debatte erkennbar werden zu lassen. Dabei konnten natürlich nicht alle Varianten dieser Positionen berücksichtigt werden, und auch der folgende Text hat nicht den Anspruch, einen umfassenden Überblick über die Vielzahl tatsächlicher oder gar möglicher Positionen zu geben.<sup>13</sup> Im Verlauf der Untersuchung werden aber auch schwächere Versionen des Physikalismus diskutiert, die weder der Identitätstheorie noch dem Eliminativismus zuzurechnen sind, z.B. im 3. Kapitel die Auffassung, dass die mikrophysikali-

---

<sup>12</sup> Barry Stroud spricht mir deshalb aus dem Herzen, wenn er schreibt: „Another conviction out of which I write is that philosophy is one subject and that progress in one place depends on the resolution of issues that lie elsewhere. One is led eventually into almost all other areas and questions. This is certainly true of the work of the great philosophers of the past. Against that high standard, the current professional fixation on distinct “fields” or areas of academic “specialization” looks like no more than a bad joke.” (2000: ix).

<sup>13</sup> Einen taxonomischen Überblick gibt Chalmers (2002).

schen Tatsachen alle Tatsachen determinieren. Auch diese schwächeren Versionen des Physikalismus werden zurückgewiesen. Aus der hier vertretenen Perspektive ist die wesentliche asymmetrische Relation zwischen Mikro- und Makroentitäten nicht Determination oder Verursachung, sondern die Kompositionsbeziehung. Im 6. Kapitel wird gezeigt, dass sich aus dieser pluralistischen Sicht auch das Problem der mentalen Verursachung lösen lässt, das allgemein als das stärkste Argument für die Identitätstheorie angesehen wird. Die Vorstellung, dass nur die physikalische Mikroebene Kausalkräfte besitzt, wird als verfehlt erwiesen. Dass die Makroentitäten keine Kausalkräfte haben, die über die der Mikrophenomene hinausgehen, bedeutet nicht, dass sie gar keine haben. Die Mikrostruktur der Makroentitäten *erklärt* ihre Kausalität, statt sie ihnen sozusagen wegzunehmen. Auch das Problem der mentalen Verursachung reflektiert den ontologischen Fundamentalismus. Es beruht auf der Vorstellung, dass nur fundamentale Kausalkräfte überhaupt Kausalkräfte sind.

Ich möchte die Hauptargumentation noch einmal kurz zusammenfassend im Gang durch die folgenden Kapitel Revue passieren lassen. Das 1. Kapitel versucht durch Zusammentragen einiger mehr oder minder alltäglicher Beobachtungen unseren Sinn für das Bewusstsein als ein eminent reales und konkretes Phänomen zu schärfen, das weder auf seine kausale Rolle noch auf innergeistige Relationen zu reduzieren ist. Im 2. Kapitel wird das Bewusstseinsrätsel eingeführt. Seine Hauptthese ist, dass die Vorstellung des Bewusstseinsrätsels nicht mehr leistet, als uns daran zu erinnern, dass wir das Bewusstsein als ein begrifflich irreduzibles Phänomen in unsere Ontologie aufnehmen müssen. Die Illusion, es zeige uns die Notwendigkeit einer weiteren fundamentalen, etwa über mögliche neurophysiologische Erklärungen des Bewusstseinsgeschehens hinausgehenden *Erklärung*, beruht nach den Überlegungen dieses Kapitels auf der Elimination der manifesten physischen Welt, die wir so missverstehen, als sei eine vollkommen transparente Erklärung dieser Welt gegeben worden. Das 3. Kapitel setzt die Auseinandersetzung mit dem Bewusstseinsrätsel anhand der Schwierigkeit fort, sich das Bewusstsein als räumlich zu denken, die als sein intuitiver Kern gelten kann. Im 4. Kapitel werden Identitätsaussagen allgemein, im 5. Kapitel theoretische Identifikationen diskutiert, mit dem Ziel zu zeigen, dass die Vorstellung einer empirisch informativen Identitätsaussage keinen Sinn hat und es sich bei den Identifikationen tatsächlich um verkappte Eliminationen handelt. Nach dem 6. Kapitel zur mentalen Verursachung wirft eine kurze Schlussbetrachtung

schließlich die Frage auf, ob wir es bei der vorgeschlagenen Reaktion auf das Leib-Seele-Problem mit einer Lösung oder Auflösung desselben zu tun haben.

# 1 Erklären, was Bewusstsein ist

*Was man bewusst heißen soll, braucht man nicht zu erörtern, es ist jedem Zweifel entzogen.*

– Sigmund Freud, Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse

*This consciousness that is myself of selves, that is everything, and yet nothing at all – what is it?*

– Julian Jaynes, The Origin of Consciousness in the Breakdown of the Bicameral Mind

*...das Spiel mit diesen Worten, ihre Verwendung im sprachlichen Verkehr, dessen Mittel sie sind, ist verwickelter – die Rolle dieser Wörter in unserer Sprache eine andere, als wir versucht sind, zu glauben.*

*(Diese Rolle ist es, die wir verstehen müssen, um philosophische Paradoxe aufzulösen. Und darum genügt dazu gewöhnliche nicht eine Definition; und schon erst recht nicht die Feststellung, ein Wort sei 'undefinierbar'.)*

– Ludwig Wittgenstein, Philosophische Untersuchungen

Ist es nicht offensichtlich, was Bewusstsein ist? Manche glauben wie Freud unmittelbare Klarheit darüber zu besitzen. Sie haben das Gefühl, dass dies auch jedem klar sein muss und halten ausführliche Erörterungen der Definition von Bewusstsein für überflüssig. Anderen erscheint diese vorgebliche Gewissheit als ihrerseits dubios und das Fehlen einer allgemein anerkannten, expliziten Definition als äußerst verdächtig. Es ist ihnen ein Zeichen dafür, dass mit dem Bewusstseinsbegriff etwas grundlegend nicht stimmt. Vielleicht bezeichnet er gar kein kohärentes Phänomen und muss zumindest aus dem ontologisch ernstzunehmenden wissenschaftlichen Diskurs verbannt werden. Vielleicht liegt hier nur ein begriffliche Mischmasch verschiedener Vorstellungen vor, oder ein sogenannter ‚Kofferbegriff‘ (*suitcase term*), womit gemeint ist, dass die unter ihn fallenden Entitäten etwa den gleichen Grad an Ähnlichkeit und innerem Zusammenhang haben wie die Socken, die Zahnpasta und das Buch, die der Reisende mit einer Vielzahl weiterer Dinge in seinen Koffer packt. Diese extremen Zweifel erscheinen aber ihrerseits kaum glaubwürdig. Meist sind sie erkennbar durch die eingangs skizzierte materialistische Metaphysik und/oder einen epistemologischen Skeptizismus in Bezug auf die Erkennbarkeit von Bewusstsein



motiviert. Die Zweifel am Bewusstseinsbegriff sind dann nur ein billiges Mittel, ein suspektes, philosophisch irritierendes Phänomen einfach loszuwerden.<sup>14</sup>

Die These der Ambiguität des Bewusstseinsbegriff ist ein Leitmotiv der Bewusstseinsliteratur, und es herrscht kein Mangel an Vorschlägen zur Unterscheidung verschiedener Bedeutungen des Bewusstseinsbegriffs: transitives/intransitives Bewusstsein, phänomenales/funktionales Bewusstsein, kognitives Bewusstsein, sensorisches Bewusstsein, intentionales Bewusstsein, introspektives Bewusstsein, Bewusstsein 1./2./x.ter Stufe, Kontrollbewusstsein, Monitorbewusstsein, Zugriffsbewusstsein, Zustandsbewusstsein, Speziesbewusstsein (*creature consciousness*). Einige Begriffe in dieser Aufzählung können zwar als synonym oder nahezu synonym gelten, aber die Vielfalt der Unterscheidungen ist doch erstaunlich, obwohl diese Liste keinen Anspruch auf Vollständigkeit erheben kann. Methodisch kranken die Erörterungen der Ambiguitätsthese fast ausnahmslos daran, dass keine Rechenschaft darüber abgelegt wird, in welchem Sinn von „Bedeutung“ „Bewusstsein“ denn mehrere Bedeutungen haben soll. Dies ist fatal, denn wie Bedeutungen zu zählen sind, ist nicht nur eine sehr schwierige theoretische Frage für die Sprachphilosophie und Linguistik. Wie jeder, der einmal die Einträge zu einem Begriff in verschiedenen Wörterbüchern verglichen hat, bestätigen kann, kommt es darüber auch in der Praxis zu keiner Einigung. Und soll es sich bei der Ambiguität um Homonymie handeln oder um Polysemie? Der Ton vieler Diskussionen legt sicherlich ersteres nahe, womit wir wieder bei der These wären, dass mit dem Wort „Bewusstsein“ nur ein begriffliches Sammelsurium verbunden ist. Ich werde im Folgenden für die These argumentieren, dass man zwar natürlich verschiedene *Arten* von Bewusstsein unterscheiden kann, dass hier aber doch ein einheitliches Phänomen vorliegt. Es gibt Konzepte sprachlicher Bedeutungen (z.B. Lakoff 1987), nach denen diese These durchaus damit vereinbar ist, dass „Bewusstsein“ polysem ist. Wie man sich hier entscheidet, spielt aber für die Zwecke dieser Untersuchung keine Rolle. Wesentlich ist nur, ob tatsächlich ein einheitliches Phänomen vorliegt.

Auf der anderen Seite versuchen manche Autoren, Bewusstsein mit ausgefeilteren theoretischen Konstruktionen oder bestimmten Metaphern näher zu charakterisieren. In diesem Zusammenhange werden dann etwa Begriffe wie „Subjektivität“ oder „Qualia“ ins Spiel gebracht oder die berühmte Nagelsche Formel des „wie es ist“ ein bewusstes Wesen, ein Mensch oder Tier zu sein (*what it's like to be that*

---

<sup>14</sup> Ein Beispiel hierfür ist Wilkes 1988.

*being*; vgl. Nagel 1974). Anderen wiederum erscheinen diese Begriffe und Metaphern als mindestens so mysteriös wie der Bewusstseinsbegriff, den sie erläutern sollen. Auch ich glaube nicht, dass es an dieser Stelle besonders hilfreich ist, philosophische Kunstausdrücke einzuführen und einen großen begrifflichen Aufwand zu betreiben, um verständlich zu machen, was mit „Bewusstsein“ gemeint sein soll.

Den Bewusstseinsbegriff zu definieren gilt gemeinhin als äußerst schwieriges, wenn nicht gar hoffnungsloses Unterfangen. Dies mag tatsächlich zutreffen, wenn, wie dies in philosophischen Erörterungen oft geschieht, stillschweigend oder ausdrücklich ein äußerst anspruchsvolles Konzept von Definition zugrunde gelegt wird. Demnach besteht das Definiens in einem komplexen Ausdruck, der mit dem Definiendum in allen Kontexten austauschbar, gar synonym sein soll. Dennoch soll die Definition informativ sein, sie soll uns etwas mitteilen und das Verständnis des Ausdrucks verbessern. Es ist aber schwer zu verstehen, wie diese Bedingungen allesamt erfüllt sein könnten. Macht die Vorstellung wirklich Sinn, dass ich dadurch etwas erfahre, dass ich zwei synonyme Ausdrücke einander zuordne?<sup>15</sup> (Natürlich könnte ich dadurch etwas über bestimmte Zeichen lernen, nämlich dass sie die gleiche Bedeutung haben; aber diese Art von Information ist normalerweise nicht gemeint, wenn in philosophischen Debatten gefordert wird, eine Definition solle informativ sein.) Und ist die Annahme plausibel oder auch nur verständlich, dass ein einfacher und ein komplexer Ausdruck synonym sind, dass also ersterer nur so etwas wie eine Abkürzung für den zweiten ist und zwar so, dass dies eine von uns zu erkennende Tatsache über die gewöhnliche Bedeutung von Ausdrücken wie z.B. „Bewusstsein“ ist und nicht nur eine bloße Stipulation? Mir scheint äußerst zweifelhaft, dass diese Fragen positiv und befriedigend beantwortet werden können. Wenn daher, wie weit hin anerkannt wird, Bewusstsein *in diesem Sinn* von Definition nicht definierbar ist, so bedeutet dies nicht zwangsläufig, dass der Bewusstseinsbegriff einen Sonderstatus hat, etwa deshalb, weil er zu einer exklusiven Liste primitiver Grundbegriffe gehört. Die Erklärung dafür könnte vielmehr sein, dass überhaupt kein Begriff in diesem Sinn definierbar ist. Dieses ganze Bild – primitive Grundbegriffe und abgeleitete Begriffe, die durch diese Grundbegriffe definiert, auf sie vollständig zurückgeführt werden können, sie letztlich nur abkürzen – könnte überhaupt verfehlt und nicht anwendbar sein. Und tatsächlich dürften nur wenige Philosophen heutzutage noch die-

---

<sup>15</sup> Diese Frage oder ihr doch sehr nah verwandte sind in der Tradition der analytischen Philosophie oft als das so genannte Paradox der Analyse diskutiert worden.

ses Bild verteidigen.<sup>16</sup> Aber manche Denkgewohnheiten überleben den Untergang der Theorien, von denen sie einst ihre Berechtigung abgeleitet haben. Bestimmte Erwartungen daran, was eine Definition ist und was sie leisten muss, scheinen mir in diese Kategorie zu gehören.

Wenn aber ein Wort zu definieren einfach nur heißt: erklären, was damit gemeint ist, die Grenzen seines Bezugsobjekts abstecken, kann „Bewusstsein“ durchaus definiert werden. Wenn man sich von der Vorstellung freimacht, eine Definition müsse zu einer Elimination des Definiendums führen, und es gebe eine scharfe Grenze zwischen dem Definieren eines Wortes und gehaltvollen Aussagen, die wir über seinen Gegenstand machen, gibt es keinen vernünftigen Grund, warum man nicht das Sammeln einiger zentraler Aussagen über das Bewusstsein und die Auflistung einiger seiner grundlegenden Merkmale als einen Beitrag zu seiner Definition auffassen sollte. Ich möchte dies deshalb im Folgenden kurz tun. Später komme ich dann noch näher auf einige in der philosophischen Debatte zentralen Bewusstseinskonzeptionen zu sprechen.

## 1.1 Einige grundlegende Merkmale des Bewusstseins

Zur ersten Orientierung kann die folgende Definition des Psychologen George Turnbull Ladd dienen:

What we are when we are awake, as contrasted with what we are when we sink into a profound and perfectly dreamless sleep or receive an overpowering blow upon the head – *that* is to be conscious. What we are less and less, as we sink gradually down into dreamless sleep, or as we swoon slowly away: and what we are more, as the noise of the crowd outside tardily arouses us from our after-dinner nap, or as we come out of the midnight darkness of the typhoid fever crisis – *that* is to become conscious.

(zit. nach Güzeldere 1997: 6)

Ladd charakterisiert Bewusstsein also unter anderem durch einen Kontrast mit verschiedenen Formen der Bewusstlosigkeit: traumloser Schlaf und Ohnmacht. Durch die Betonung der Traumlosigkeit des Schlafes wird angedeutet, dass das Bewusstsein

---

<sup>16</sup> Zeitgenössische Verteidiger der Vorstellung begrifflicher Wahrheiten wie z.B. David Chalmers (1996) und Frank Jackson (1998) gehen von einer schwächeren Konzeption aus, nach der begriffliche Wahrheiten keine Synonymie erfordern oder mit sich bringen. Auf diese Konzeption komme ich im nächsten Kapitel zu sprechen, wenn es um die These geht, dass manche physikalischen Erklärungen manifester physischer Phänomene den Status begrifflicher Wahrheiten haben.

eng mit dem Zyklus des Schlafens, Wachens und Träumens verbunden ist. Statt der negativen Bestimmung von Bewusstsein durch die Zustände seiner Abwesenheit, könnte man auch sagen, dass Bewusstsein das ist, was dem Wach- und dem Traumzustand – als seinen zwei grundlegendsten Formen – gemeinsam ist. Traum- und Wachbewusstsein sind globale Bewusstseinszustände. Dass sie globale Bewusstseinszustände sind, bedeutet, dass sie die Vielzahl anderer Bewusstseinszustände und Prozesse umfassen. In Bezug auf jeden anderen Bewusstseinszustand können wir fragen, ob er eine Modifikation des Wach- oder des Traumbewusstseins ist.

Das zweite zentrale Charakteristikum von Bewusstsein, auf das Ladd hinweist, ist, dass es Grade hat. Die Intensität des Bewusstseins ist variabel, sie kann allmählich abnehmen oder steigen. In dieser Hinsicht kann man Bewusstsein mit einem Rheostaten wie z.B. einem Lichtschalter mit Dimmer vergleichen (Searle 1992). Das Bewusstsein ist entweder „an“ oder „aus“, aber innerhalb des „An“-Zustandes gibt es verschiedene Intensitätsstufen. Die Intensität des Bewusstseins ist eine Dimension, auf der sowohl globale als auch spezifischere Bewusstseinszustände unabhängig von ihrem Inhalt und davon, um welchen Typ von Bewusstseinszustand es sich handelt, eingeordnet werden können. Grade von Wachheit sind eine globale Spezies von Graden der Bewusstseinsintensität. Die verschiedenen Grade von Intensität, die ein Schmerz haben kann, sind ein Beispiel für eine lokalere Unterscheidung von Bewusstseinsintensität.

Es erscheint mir nicht übermäßig spekulativ, anzunehmen, dass die Bewusstseinsintensität als eine globale Dimension des Bewusstseins leichter der quantifizierenden naturwissenschaftlichen Theoriebildung zugänglich ist als etwa die unübersichtliche Vielfalt der verschiedenen Inhalte des Bewusstseins. So lässt sich etwa erwarten, dass dieser Dimension auf einer Skala Zahlenwerte zugeordnet werden könnten. Gleichzeitig sind wir mit dem Begriff der Bewusstseinsintensität dem Bewusstsein ‚als solchem‘ begrifflich sehr nahe, weil es sich um eine sehr allgemeine Bestimmung handelt, die auf jeden Bewusstseinsinhalt angewendet werden kann.

Wenn wir unsere Aufmerksamkeit darauf richten, dass unser Bewusstseinsleben durch den Zyklus des Schlafens, Träumens und Wachens strukturiert wird und Grade der Intensität hat, schärfen wir unseren Sinn dafür, dass das Bewusstsein ein konkretes biologisches Phänomen ist, ein Zustand, der wesentlich mit einer mehr oder minder großen Aktivierung eines Organismus verbunden ist und diese Aktivie-

rung selbst zum Teil ausmacht. Umso bemerkenswerter ist es, dass diese Aspekte des Bewusstseins in der philosophischen Literatur kaum beachtet werden.

Dies gilt natürlich nicht für die Intentionalität oder Repräsentationalität des Bewusstseins, die ich als sein drittes zentrales Merkmal aufführen möchte. Dabei kann es hier an dieser Stelle natürlich nicht um eine Auseinandersetzung mit den philosophischen Problemen der Intentionalität gehen. Ich möchte hier nur versuchen, einen wesentlichen Aspekt unseres intuitiven, vorphilosophischen Verständnisses von Bewusstsein zu formulieren. Am anschaulichsten lässt sich das in Bezug auf die Wahrnehmung tun. „Es gibt dort etwas in mir“, so könnte man versuchen, dieses intuitive Verständnis auszudrücken, „das mich meine Umgebung sehen lässt. Ich bin in einem bestimmten Zustand, der macht, dass ich die Dinge um mich herum erkennen kann. Wenn ich die Augen schließe, sind sie immer noch da, aber ich habe keinen Zugang mehr zu ihnen. Wenn ich müde bin, scheinen die Dinge vor meinen Augen zu verschwimmen, aber ich weiß, dass nicht sie sich geändert haben, sondern ich. Es ist etwas in mir, mein Zustand, das mir diesen Zugang eröffnet.“ Auch diese Überlegung stößt uns darauf, dass das Bewusstsein ein konkretes, reales Phänomen ist, nicht ein bloßes theoretisches Konstrukt, das etwa eingeführt wird, um abstrakte Muster im Verhalten zu erklären oder zu erfahren. Dass es dies sein könnte, ist natürlich eine philosophische Vorstellung und zwar eine von drei für die philosophische Bewusstseinsdebatte zentralen Bewusstseinskonzeptionen, denen ich mich nun zuwende.

## **1.2 Bewusstsein als innergeistige Relation**

Sowohl in der Tradition der Sinnesdatentheorie als auch im Funktionalismus findet man Versuche, das Bewusstsein als eine innergeistige Relation zu verstehen.<sup>17</sup> Die Sinnesdatentheoretiker neigten dazu, die Sinnesdaten in einer unauflöslchen Relation mit einem nicht näher spezifizierten Subjekt der Erfahrung zu konzipieren, das sich der Sinnesdaten bewusst ist. Funktionalisten haben vorgeschlagen, die Bewusstheit eines intentionalen Zustands durch eine Relation zu einem intentionalen Zustand 2. Stufe zu explizieren, der den Inhalt hat, dass man sich in dem Zustand 1. Stufe befindet (z.B. Rosenthal 1988, Lycan 1996). Verwandt sind Versuche, diese Relation

---

<sup>17</sup> In Bezug auf den Funktionalismus stellt diese Diagnose auch Alvin Goldmann (1997).

zu modalisieren und die Bewusstheit des Zustandes als seine Berichtbarkeit in irgendeiner Form zu verstehen (vgl. Goldman 1997).

Mir scheint, dass diese Konzeption unserem natürlichen Verständnis von Bewusstsein nicht gerecht werden kann. Wenn ich z.B. den eben erwähnten Gedanken fasse, dass meine visuelle Wahrnehmung verschwommen ist und ich deshalb wohl müde sein muss, sind sowohl dieser Gedanke als auch mein visuelles Erleben Bewusstseinszustände. Beide sind Modifikationen meines Wachzustandes. Mein visuelles Bewusstsein ist nicht bewusst, weil ich diesen Gedanken habe, sondern weil es eine *Form* von Bewusstsein ist. Der Gedanke hat ihm Bewusstsein nicht *verliehen*, sondern konstatiert nur, dass Bewusstsein vorliegt. Die Bewusstheit ist sowohl dem visuellen Erleben als auch dem Gedanken inhärent. Weder die Vorstellung, dass einem an sich unbewussten Zustand Bewusstsein verliehen wird, noch die, dass die Bewusstheit in der Relation selbst besteht, scheint einen klaren Sinn zu haben. Eine Bewusstseinsrelation besteht nur insofern, als zumindest eines ihrer Relata inhärent bewusst ist. Wenn ich mir bewusst bin, dass es regnet, besteht diese Relation deshalb, weil ich bei Bewusstsein bin und mein Bewusstsein sich in einem entsprechenden Zustand befindet. Und die Tatsache, dass es regnet, wird natürlich nicht dadurch bewusst, dass ich Bewusstsein von ihr habe. Entsprechende Bemerkungen gelten auch für die Relation zwischen einem Erkenntnissubjekt und seinen Sinnesdaten. Auch ein Erkenntnissubjekt kann nicht anderen Entitäten Bewusstsein verleihen.

### **1.3 Bewusstsein als funktionales Bewusstsein**

Der Begriff des funktionalen Bewusstseins reduziert Bewusstsein auf eine bestimmte kausale Rolle bei der Vermittlung zwischen Input und behavioralem Output. Dem gewöhnlichen Verständnis von Bewusstseinsentitäten entspricht diese Konzeption sicherlich nicht. Die natürliche Auffassung ist die, dass ein Schmerz eine kausale Rolle, also Ursachen und Wirkungen, *hat*, nicht dass er mit dieser identisch ist. Noch meinen wir, dass ein Schmerz irgendetwas ist, was immer es auch an sich sei, das diese Rolle spielt. Wir meinen mit „Schmerz“ eine spezifische Entität, die wir aus eigenem Erleben kennen, nicht etwas uns als solches Unbekanntes, dessen Identität wir erst noch bestimmen müssen.

Für den Begriff des funktionalen Bewusstseins scheint wenig mehr als der Wunsch zu sprechen, das Bewusstsein als ein metaphysisch und/oder epistemolo-

gisch anstößiges Phänomen einfach wegzudefinieren. Deshalb wird er normalerweise nur von Eliminativisten implizit oder explizit vertreten.<sup>18</sup> Eine Ausnahme ist David Chalmers (1996), der eine durchgängige Ambiguität aller mentalen Begriffe zwischen einer funktionalen und einer phänomenalen Lesart behauptet, wobei letztere im Wesentlichen dem gewöhnlichen Bewusstseinsbegriff entspricht. Bei Identität der kausalen Rolle soll – so Chalmers’ Gedankenexperiment – auch einem Wesen, dem jedes Bewusstsein im gewöhnlichen Sinne abgeht, ‚funktionales Bewusstsein‘ zugesprochen werden. In einer Hinsicht scheint es sich hierbei um eine rein terminologische Festlegung ohne substanziellen Gehalt zu handeln, um die Entscheidung, eine bestimmte kausale Struktur als Bewusstsein zu bezeichnen, die der Aussage, dass diese Struktur vorliegt, nichts hinzufügt.<sup>19</sup> Gleichzeitig ist sie aber auch ein Symptom einer Verwirrung, die zu der Vorstellung führt, es gebe zwei verschiedene Erklärungsprojekte in Bezug auf das Bewusstsein, die bloße Erklärung seiner kausalen Rolle, das so genannte ‚leichte Problem‘ und die Erklärung des phänomenalen Bewusstseins, das ‚schwere Problem‘. Diese Vorstellung wird im nächsten Kapitel noch eingehender zu diskutieren sein.

#### **1.4 Bewusstsein als phänomenales Bewusstsein**

Zum Begriff der Qualia oder des qualitativen oder phänomenalen Bewusstseins habe ich ein gespaltenes Verhältnis. Einerseits bin ich auf der Seite derjenigen, die mit diesen Begriffen auf etwas verweisen wollen, das uns sehr wichtig ist, und das z.B. von den behavioristischen oder funktionalistischen Theoretikern entweder gar nicht thematisiert oder gar geleugnet wird. Andererseits bin ich der Meinung, dass dieses Etwas das Bewusstsein selbst ist, während diejenigen, die diesen Terminus gebrauchen, damit etwas Spezielleres zu meinen scheinen.

Paradigmatisch für qualitative Zustände sind z.B. Farb-, Geruchs- oder Geräuschempfindungen, kurzum alle Arten sensorischer Zustände. Auch die vielfältigen Varianten emotionalen Empfindens wird man wohl allgemein unter den Begriff des

---

<sup>18</sup> Der Funktionalismus und andere Formen des Neobehaviorismus sind in jüngerer Zeit so oft überzeugend widerlegt worden, dass ich es an dieser Stelle mit diesen knappen Bemerkungen bewenden lassen möchte. Vgl. z.B. Searle 1992, Strawson 1998 und für volitionale Bewusstseinsphänomene Seebaß 1993.

<sup>19</sup> Wobei hier nur *arguendo* zugestanden werden soll, dass es dem funktionalistischen Konzept überhaupt gelingt, eine kausale Struktur zu identifizieren. Zum entsprechenden Vorwurf der universellen Realisierbarkeit siehe z.B. Putnam 1991, Searle 1992 und Goldstrasz 2002.

qualitativen oder phänomenalen Bewusstseins fassen. Wer dagegen auch bewusstes Denken, Wünschen, Hoffen etc. als qualitative Zustände verstehen will, wird auf Widerstand stoßen. Nun ist sicherlich z.B. das Denken ein Zustand von anderer Art als sensorische und emotionale Zustände, und deshalb kann es leicht scheinen, dass derjenige, der auch das Denken als einen qualitativen Zustand bezeichnen möchte, behaupten will, es gebe für jeden möglichen Inhalt des Denkens eine charakteristische Empfindung in dem Sinn, in dem es eine charakteristische Empfindung des Rotsehens gibt. Dies ist sicherlich nicht der Fall. Daraus folgt aber nicht, dass das Denken, Verstehen oder Wünschen nicht erlebt wird, dass es nicht ‚irgendwie ist‘, in diesen Zuständen zu sein, dass sie kein Aspekt des bewussten Erlebens sind (Strawson 1994: 5ff).

Die grundlegende Frage, ob etwas ein Bewusstseinszustand oder -ereignis ist, lässt sich im Anschluss an vorige Überlegungen wie folgt formulieren: würde die An- oder Abwesenheit dieses Zustands eine Modifikation des Wach- oder Traumzustandes bedeuten? Wäre mein Bewusstseinsleben anders, wenn ich die Sprachlaute, die ich gerade gehört habe, nicht verstanden hätte, oder wenn ich nicht denken würde, dass die Person gerade etwas sehr Kluges gesagt hat? Und da scheint die Antwort doch ein eindeutiges „Ja“ zu sein. Diese Antwort lässt aber noch zwei verschiedene Interpretationen zu. Sie begründet nur, was man als die These der Supervenienz der höherstufigen Zustände des (bewussten) Meinens, Denkens, Verstehens etc. gegenüber den Empfindungszuständen, die paradigmatisch für den Begriff des qualitativen oder phänomenalen Bewusstseins stehen, bezeichnen könnte. Nach dieser Supervenienzthese kann es keinen Unterschied in Bezug auf das (bewusste) Denken ohne entsprechende Unterschiede im phänomenalen Bewusstsein geben. Diese These ist aber noch mit der Auffassung kompatibel, dass die Unterschiede im phänomenalen Bewusstsein nur darin bestehen, welche Empfindungen man hat und wie diese Empfindungen miteinander verbunden sind. Demnach wäre etwa ein Gedanke so etwas wie ein Konglomerat<sup>20</sup> verschiedener basaler Empfindungen, er bestünde in der Struktur und dem Zusammenhang dieser Elemente. Der Unterschied etwa zwischen dem Hören von Sprachlauten bevor und nachdem man die Sprache erlernt hat, läge also allein darin, dass für den, der die Sprache versteht, die auditorischen Empfindungen nunmehr z.B. mit Empfindungen der Familiariät verknüpft sind und mit sensorischen Empfindungen in anderen Modalitäten. Jetzt, wo man weiß, was „Pasta“

---

<sup>20</sup> Dieser Ausdruck wird in diesem Zusammenhang von Wittgenstein verwendet, der diese Vorstellung im in Bezug auf die Furcht zurückweist (*Zettel* §502).



bedeutet, löst das Hören dieses Wortes mitunter Geruchs- und Geschmacksempfindungen aus, man spürt, wie das Wasser im Munde zusammenläuft, stellt sich einen Teller voll Nudeln vor, freut sich oder empfindet einen Widerwillen, weil man sich gerade schon den Magen voll geschlagen hat oder Nudeln nicht mag, usw.. Aber es gäbe keine darüber hinausgehende spezifische Erfahrung des Verstehens und auch keine des Denkens, Hoffens oder Wünschens.

Aber sind wir uns wirklich sicher, dass wir unser bewusstes Denken und Wollen auf solch elementare Gegebenheiten reduzieren können? Mir scheint, dass wir dem Gedanken, dies müsse doch möglich sein, mit Misstrauen begegnen sollten. Er verrät den gleichen reduktionistischen und fundamentalistischen Impuls, der uns auch in der Leib-Seele-Debatte in so große Schwierigkeiten bringt. Ich möchte auf diesem Punkt an dieser Stelle aber nicht insistieren, da die Existenz phänomenalen Bewusstseins im weithin akzeptierten, nur von manchen radikalen Neobehavioristen oder Funktionalisten in Frage gestellten, Sinn auf jeden Fall ausreicht, um das Bewusstseinsrätsel aufzuwerfen, dem ich mich nun gleich zuwenden möchte. Wie schon bemerkt, liegt aber ein Problem in der Opposition zwischen phänomenalem und funktionalem Bewusstsein, und es ist mir deshalb wichtig, den Bewusstseinsbegriff ohne die Qualifikation „phänomenal“ zu verwenden, die mir bestenfalls pleonastisch, schlimmstenfalls aber ein Ausdruck der angesprochenen Verwirrung zu sein scheint, um die es im folgenden Kapitel unter anderem gehen wird.

## 2 Das Bewusstsein verräteln

*Im unendlichen Raum zahllose leuchtende Kugeln, um jede von welchen etwan ein Dutzend kleinerer, beleuchteter sich wälzt, die inwendig heiß, mit erstarrter, kalter Rinde überzogen sind, auf der ein Schimmelüberzug lebende und erkennende Wesen erzeugt hat; – dies ist die empirische Wahrheit, das Reale, die Welt.*

– Arthur Schopenhauer, Die Welt als Wille und Vorstellung II

Wie kann unser reichhaltiges Bewusstseinsleben, wie kann seine „Technicolor-Phänomenologie“ (McGinn 1989) aus grauer – eigentlich sogar farbloser (Metzinger 2000) – Gehirnmasse entstehen, wie kann das Gehirn sie „produzieren“ oder „erzeugen“? Betrachten wir die Geschichte des Universums oder auch, da wir ja nicht sicher wissen, ob überhaupt und wenn ja, wann, an anderer Stelle im Universum Bewusstsein vorgekommen ist, vorkommt, oder noch vorkommen wird, nur die unseres Planeten. Nach den allgemein akzeptierten wissenschaftlichen Theorien unserer Zeit ordnete sich die irdische Materie über Milliarden von Jahren zu ständig wechselnden, zuerst anorganischen, dann organischen Konfigurationen an, bis auf einmal mit einer bestimmten, bis dato nie vorgekommenen organischen Konfiguration – und damit auf einer Stufe der Evolution des Lebens, die wir bis dato nicht genau bestimmen können, da uns weder unsere vorthoretischen Intuitionen noch wissenschaftliche Erkenntnisse erlauben, welche Lebensformen Träger von Bewusstseinszuständen sind – eine Entität gänzlich neuer Art entstand, das 'Licht des Bewusstseins' aufflackerte, wenn auch wohl in einer eher trüben und matten Form. *Warum* geschah dies? Ist dieses Ereignis nicht mindestens genauso rätselhaft wie das Erscheinen des Dschinn, wenn Aladin an seiner Lampe reibt (Huxley 1866: 193)? Wie sollen wir erklären, dass Bewusstsein ausgerechnet mit dieser Konfiguration und nicht mit einer der zahllosen anderen verbunden ist? Und warum ausgerechnet dieses Bewusstsein? Könnte es nicht genauso gut ein Bewusstsein gänzlich anderer Art sein? Wäre ein Bewusstsein gänzlich anderer Art nicht genauso gut – und das heißt an dieser Stelle natürlich: genauso schlecht – aus der mit ihm – wie? – verbundenen Materiekonfiguration zu erklären?

Oder stellen wir uns ein auf die Dimensionen einer Mühle vergrößertes Modell des Gehirns vor. In seinen Windungen könnten wir endlos umherklettern, aber wie sollten wir je Bewusstseinszustände dort finden (Leibniz 1720, Bieri 1996)?

Auch bei einem echten, lebendigen und wachen Gehirn kann die Suche leicht als von vornherein vergeblich erscheinen. Die – vermeintlich – dort zu verortenden Bewusstseinszustände verweigern sich unseren gewohnten Zugangswegen zur Realität, bleiben seltsam ungreifbar. Stellen wir uns vor, ein Gehirn in einem geöffnetem Schädel sei unserer Beobachtung zugänglich, während sein Besitzer das ganze Spektrum gängiger Bewusstseinszustände durchlebt, vor allem Erfahrungen in den verschiedenen Modalitäten der sinnlichen Wahrnehmung. Würde uns dies helfen? Kaum! Das Gehirn erscheint visuell grau, auch wenn sein Besitzer Farbempfindungen hat; es schmeckt nicht nach Schokolade, auch wenn sein Besitzer Schokolade isst und schmeckt, selbst dann, wenn man an den Stellen schleckt, die die physiologische Basis der Geschmacksempfindungen sind. Es würde sich weich und feucht anfühlen (wie man hört, ich habe noch keines angefasst!), egal ob sein Besitzer einen Gegenstand anfasst, den er als weich und feucht oder einen, den er als trocken und hart empfindet (Nagel 2004). Und auch wenn wir unser Ohr auf das Areal legen würden, in dem sich die materielle Basis auditiver Empfindungen lokalisieren lässt, während der Besitzer dieses Gehirns Musik hört: wir würden diese dadurch nicht hören. Wie können wir da verstehen, dass das Gehirn die materielle Basis, ja sogar der Ort von Bewusstseinszuständen sein soll? Ist es nicht ein Rätsel, wie das möglich sein sollen, stehen wir nicht vor einer „Grenze der Naturerkenntnis“ (Dubois-Reymond 1916), müssen „Ignorabimus“ rufen und damit gestehen, dass die Auflösung dieses Rätsel zumindest uns Menschen aufgrund unserer kognitiven Begrenztheit unmöglich ist und auch bleiben wird (McGinn 1991)?

Dies sind einige der gängigen Formulierungen, Illustrationen und Dramatisierungen des Bewusstseinsrätsels. Wer diesen Einflüsterungen gegenüber völlig taub bleibt, wer gar kein Sensorium hat für die das Bewusstsein umgebende Atmosphäre der Rätselhaftigkeit, ist wahrscheinlich entweder kein voll sozialisiertes Mitglied unserer wissenschaftlich-philosophischen Kultur – oder vielleicht durch endlose Beschäftigung mit diesem Thema abgestumpft. Durch die Wiederentdeckung des Geistes – sprich: des Bewusstseins – in den letzten Jahrzehnten ist auch das Bewusstseinsrätsel wieder in den Mittelpunkt des intellektuellen Interesses gerückt und hat naturgemäß eine Vielzahl einander widerstreitender Haltungen und Diagnosen ausgelöst, die das ganze Spektrum typischer Positionen in der gegenwärtigen Leib-Seele-Debatte repräsentieren. Wie alle großen Fragen der Philosophie, bündelt auch das Bewusstseinsrätsel eine Vielzahl verschiedener Probleme so, dass diese uns in

einer Art von Gestaltwahrnehmung als ein einziges Rätsel erscheinen. Um Fortschritte zu erzielen, müssen die Fäden entwirrt werden, ohne dass dabei der große Zusammenhang aus dem Blick gerät. Diese Aufgabe ist für ein einzelnes Kapitel zu groß. Deshalb ist, obwohl nur dieses zweite Kapitel sich speziell dem Bewusstseinsrätsel widmet, die darin entwickelte Position nur im Gesamtzusammenhang dieser Untersuchung verständlich.

Schon auf den ersten Blick lässt sich aber erkennen, dass ein guter Teil der das Bewusstsein umgebenden Atmosphäre der Rätselhaftigkeit daher rührt, dass es die Erwartung frustriert, es müsse, wie viele, wenn auch nicht alle, raumzeitlichen Entitäten *wahrnehmbar* sein. So führen die Proponenten des Bewusstseinsrätsels uns vor, dass man das Bewusstsein nicht sieht, anfasst oder gar schmeckt, wenn man das Gehirn sieht, anfasst, schmeckt usw., obwohl doch das Gehirn vielen Theorien zufolge Träger von Bewusstseins-eigenschaften ist, der Ort des Bewusstseins und der Identitätstheorie zufolge Bewusstseinsphänomene gar mit Gehirnentitäten identisch sind. Die intuitiven Schwierigkeiten, die sich daraus ergeben, dass das Bewusstsein nicht wahrnehmbar ist, hängen eng mit dem Problem der Räumlichkeit des Bewusstseins zusammen. Man könnte sogar behaupten, dass sie dieses allein ausmachen. Deshalb werden sie erst im nächsten Kapitel ausführlich behandelt, das sich ausschließlich mit der Frage der Räumlichkeit des Bewusstseins befasst.

In diesem Kapitel soll es erst einmal um verschiedene Versuche gehen, die durch die vorgestellten Veranschaulichungen evozierte Atmosphäre der Rätselhaftigkeit in eine spezifische substanzielle, philosophisch oder empirisch wirklich beantwortbare Frage zu transformieren. Seine Hauptthese ist, dass dies nicht gelingt, oder jedenfalls zumindest bis jetzt nicht gelungen ist. Das Bewusstseinsrätsel als solches leistet nicht mehr, als Bewusstseinsphänomene als Explananda wieder in Erinnerung zu rufen. Historisch gesehen hat es diese Aufgabe wahrscheinlich bereits weitgehend erfüllt.

Die Rätselfrage könnte man vielleicht so paraphrasieren: „Warum überhaupt Bewusstsein, wie ist das möglich?“ Es ist aber nicht ohne weiteres klar, was mit dieser Frage gemeint sein soll. Ein spezifischer Sinn muss ihr erst *gegeben* werden. Zweifelsfrei legitime Explananda sind spezifische *Tatsachen* über das Bewusstsein, z.B. „Warum verlieren wir das Bewusstsein durch Anästhesie?“, „Warum sind bewusste Intentionen leichter zu ändern als unbewusste?“, „Warum gibt es einen blinden Fleck im visuellen Bewusstsein und warum bemerken wir diesen normalerweise

nicht?“ und unzählige mehr. Aus der Perspektive des Rätseltheoretikers, wie ich den Proponenten des Bewusstseinsrätsels im Folgenden nennen will, berühren Antworten auf diese Frage die wesentliche Frage aber gar nicht. Zu zeigen, dass der Sinn der Rätselfrage nicht *selbstverständlich* ist, ist natürlich methodisch schwierig. Im Folgenden kann ich nur nacheinander verschiedene Versuche sie zu präzisieren durchgehen und mich bemühen zu zeigen, dass diese Präzisierungen entweder auf die Gehirn-Bewusstsein-Beziehung gar nicht anwendbar sind, oder, dass sie zwar zu einer beantwortbaren Frage führen, die möglichen Antworten auf diese Frage aber den Rätseltheoretiker nicht zufrieden stellen werden. Mit letzter Sicherheit ausschließen, dass es doch irgendeine Präzisierung der Frage geben könnte, die den Intentionen des Rätseltheoretikers näher kommt, kann man natürlich nicht. Es wäre ja auch gar nicht sinnvoll, dieses Ziel anzustreben. Wenn es eine solche Präzisierung gibt: umso besser. Es wäre schon ein Fortschritt, wenn gezeigt werden kann, dass die Frage in ihren bisherigen Formulierungen keine sinnvolle Antwort erlaubt und deshalb eine Präzisierung unumgänglich notwendig ist.

## 2.1 Diachrone vs. synchrone Emergenz

Beginnen wir mit der Beobachtung, dass in den Veranschaulichungen der Rätselfrage die Tatsache, dass das Bewusstsein einen bestimmten *qualitativen Überschuss* mit sich bringt, als Grund für seine Rätselhaftigkeit präsentiert wird. Das Gehirn ist grau, gegeben die Elimination der manifesten Welt sogar farblos, das Bewusstsein umfasst die ganze Reichhaltigkeit unserer Farberfahrung; unsere bewusste Geschmackserfahrung umfasst viel mehr als die begrenzten Geschmackserfahrungen, die wir beim Schlecken am Gehirn (oder beim Verzehr desselben) machen können, etc.. Wenn man aber von einem solchen qualitativen Überschuss des Bewusstseins spricht, muss man auch sagen, im Vergleich wozu er besteht. Hier können im Wesentlichen zwei Bezugspunkte verwendet werden. Erstens der naturgemäß vor dem ersten Auftreten des Bewusstseins bewusstlose Zustand des Universums oder ‚der Materie‘. Zweitens der Gehirnzustand, der die Basis des Bewusstseinszustandes ist. Die erste Betrachtung ist eine historische, evolutionäre. Das historisch erste Auftreten des Bewusstseins bringt neue, noch nicht da gewesene Qualitäten in die Welt. Entsprechend spricht man von der diachronen Emergenz des Bewusstseins. Dagegen ist das Verhältnis zwischen dem Bewusstsein und seiner materiellen Basis im Gehirn ein syn-

chrones. Der qualitative Überschuss besteht dann im Vergleich zu den relevanten, zur gleichen Zeit instanziierten Gehirneigenschaften. Oft wird hier analog zur diachronen Emergenz von synchroner Emergenz gesprochen und davon, dass auch diese eine Beziehung zwischen etwas Bewusstlosem oder Unbewussten – dem Gehirnzustand, der die materielle Basis des Bewusstseins bildet – und dem Bewusstsein ist. Entsprechend lässt sich unsere Rätselfrage entweder als die Frage nach einer Erklärung der diachronen oder nach einer Erklärung der synchronen Emergenz des Bewusstseins präzisieren. „Warum ist überhaupt dieses Phänomen namens ‚Bewusstsein‘ an einem bestimmten Punkt der historischen Entwicklung aufgetaucht?“ und „Warum ist Bewusstsein – und gerade diese spezifische Art von Bewusstsein – mit diesen Gehirnzuständen verbunden?“

Diese Fragen sind aber zweifelsohne eng miteinander verknüpft. Nehmen wir an, jemand würde versuchen, die diachrone Emergenz wie folgt zu erklären: „Bewusstsein emergierte, weil zum ersten Mal in der Geschichte eine bestimmte (wahrscheinlich neuronale) Materiekonfiguration auftrat, die gesetzesmäßig mit Bewusstsein verknüpft ist. Und das Auftreten dieser Materiekonfiguration wiederum lässt sich aus den Naturgesetzen ableiten und damit erklären.“ Der Rätseltheoretiker wird aber zweifelsohne das Gefühl haben, dass die entscheidende Frage nicht beantwortet wurde, sondern bestenfalls auf das Verhältnis zwischen Materiekonfiguration und Bewusstsein und damit auf das Problem der synchronen Emergenz hin verschoben worden ist. Dieses ist also vermutlich das zentrale Problem. Betrachten wir aber dennoch zunächst die diachrone Emergenz des Bewusstseins noch etwas genauer.

## **2.2 Die diachrone Emergenz des Bewusstseins erklären**

Ist eine Erklärung der diachronen Emergenz möglich, die mehr leistet als zu sagen, dass das Bewusstsein synchron mit einer bestimmten Konfiguration der Materie gesetzesmäßig verknüpft ist, und das Auftreten dieser Konfiguration wiederum aus den Naturgesetzen ableitet? Es ist schwer zu sehen, wie eine solche Erklärung aussehen sollte. Die einzige mir denkbare erscheinende Erklärung wäre eine, die das Auftreten des Bewusstseins nicht aus seiner materiellen Basis ableitet, sondern aus einem Gesetz, das das Bewusstsein mit bestimmten ihm zeitlich vorausgehenden Bedingungen verknüpfen würde. Ein solches Gesetz könnte man sich in Analogie zu vielleicht existierenden historischen Gesetzen denken, die gesellschaftliche Bedingungen be-

schreiben, auf die unweigerlich revolutionäre Veränderungen folgen müssen, oder zu einem Gesetz in der Meteorologie, dass das Auftreten von Stürmen aus bestimmten Antezedentien abzuleiten erlaubt. Dieses Gesetz wäre also ein diachrones Verlaufsgesetz. Es scheint aber wiederum klar, dass dies nicht die Art von Erklärung ist, nach der der Rätseltheoretiker sucht. Warum gilt denn dieses Gesetz, könnte er fragen. So ist die Rätselfrage wohl nicht zu beantworten. Kehren wir also zur Frage der Erklärung der synchronen Emergenz zurück.

Die naheliegendste Interpretation dieser Frage, die sicherlich auch die meisten, wenn nicht alle Autoren, die sie stellen, im Sinn haben, ist die, dass hier nach einem fehlenden Bindeglied, einem "missing link" (McGinn 1991) zwischen Gehirn und Bewusstsein gefragt wird, das die so genannte "Erklärungslücke" (Levine 1983, 1997) zwischen denselben schließt oder überbrückt. Und eine attraktive Möglichkeit, diese Frage wiederum zu präzisieren, scheint zu sein, sie als die Frage danach zu verstehen, wie denn das Gehirn Bewusstsein erzeugt oder produziert, oder wie es aus den relevanten Zuständen entsteht. Aber obwohl sich Formulierungen dieser Art in zahllosen philosophischen, wissenschaftlichen und populären Texten finden, kann man sich leicht klarmachen, dass sie in ihrem buchstäblichen Verständnis gar nicht auf die Relation der synchronen Emergenz angewandt werden können, mithin sinnlos sind. Denn die Frage danach, wie etwas erzeugt wird oder entsteht, verlangt als Antwort die Beschreibung eines Prozesses, der zu dem relevanten Resultat führt. Die Fragen danach, wie man Spaghetti macht, wie ein Unterdruck erzeugt wird, wie eine Stadt oder eine Galaxie entsteht und zahllose weitere Fragen dieses Typs werden durch Beschreibungen bestimmter Prozesse beantwortet. Eine Antwort dieses Typs kann es in unserem Fall aber nicht geben, weil ein Prozess zeitlich ausgedehnt ist, das Verhältnis zwischen Bewusstsein und materieller Basis aber ja ein synchrones ist. Wenn die Frage aber entgegen allem Anschein doch als Frage nach einem diachronen Verhältnis gemeint wäre, so ist nicht zu sehen, warum sie nicht recht leicht zu beantworten ist. Denn wie man Bewusstsein überhaupt oder einen bestimmten Bewusstseinszustand erzeugt, lässt sich durchaus sagen. Man könnte etwa in irgendeiner Weise dafür sorgen, dass ein bewusstes Wesen das Licht der Welt erblickt, man könnte jemanden aus einer Ohnmacht oder aus einem traumlosen Schlaf erwecken oder einen bestimmten Bewusstseinszustand in einer Person dadurch erzeugen, dass man etwas zu ihr sagt oder in anderer Weise auf ihr Bewusstsein einwirkt. Und die Frage, wie Bewusstsein entsteht, lässt sich durch eine Beschreibung der Prozesse

beantworten, die zur Evolution bewusstseinsfähiger Wesen wie z.B. Menschen führen. All dies sind verschiedene Konkretisierungen des oben angegebenen Schemas zur Beantwortung der diachronen Erklärungsfrage.

Zweifelsohne wird der Rätseltheoretiker mit diesen Antworten aber wiederum nicht zufrieden sein. So war die Frage auch nicht gemeint! Aber wie? Eine alternative Formulierung wäre, dass danach gefragt wird, warum die materielle Basis die Eigenschaft der Bewusstheit hat, oder – falls man etwas dagegen hat, dass hier die materielle Basis als Träger dieser Eigenschaft konzeptualisiert wird – warum die materielle Basis mit dieser Eigenschaft verknüpft ist. Die Fragen, warum etwas eine bestimmte Eigenschaft hat, kann im Wesentlichen auf zwei verschiedene Arten und Weisen beantwortet werden. Erstens durch Verweis auf einen Vorgang, durch den der Gegenstand die Eigenschaft bekommen hat: Der Tisch ist rot, weil ich ihn so gestrichen habe und verwittert, weil er schon seit langem ungeschützt der Witterung ausgesetzt ist. Zweitens durch Ableitung des entsprechenden Sachverhalts aus einem allgemeinen Gesetz, das besagt, dass alle Gegenstände eines bestimmten Typs die fragliche Eigenschaft haben: Dieser Mensch hat ein Herz, weil alle Menschen ein Herz haben. (Wenn alle Prozess- oder Kausalerklärungen letztlich auf Gesetzen beruhen, es also keine irreduzibel singulären Prozess- oder Kausalerklärungen gibt, können wir den ersten Fall als Spezialfall des zweiten ansehen; diese Frage braucht uns hier aber nicht zu interessieren.) Der erste Fall wurde im letzten Abschnitt schon erörtert. Die Forderung nach einer Erklärung dieses Typs ist entweder sinnlos, oder viel leichter zu beantworten, als es den Intentionen des Rätseltheoretikers entspricht.

Letzteres gilt auch für den zweiten Fall. Denn die Annahme, dass es Gesetze gibt, die Gehirnzustände mit Bewusstseins-eigenschaften verbinden, wird ja von den meisten Teilnehmern der Leib-Seele-Debatte geteilt.<sup>21</sup> Und aus einem solchen Ge-

---

<sup>21</sup> Dies wird sowohl von den Materialisten vorausgesetzt als auch von Dualisten oder Bewusstseinsfundamentalisten wie Chalmers und, vor seiner Konversion zum Materialismus, Jackson. Als Ausnahme sind neben Eliminativisten oder Fiktionalisten wie Dennett (1991) auch interpretationistische Positionen zu nennen, für die hier stellvertretend die von Donald Davidson (1980) stehen kann. Auch hier mögen beide Lager über einen Teil der Wahrheit verfügen. Die Existenz von Gesetzen scheint plausibel in Bezug auf die Eigenschaft der Bewusstheit selbst, die Intensität des Bewusstseins, sowie globale Bewusstseinszustände wie das Wach- und Traumbewusstsein (vgl. Kapitel 1). Hierzu gibt es ja auch bereits konkrete Vorschläge (z.B. Crick & Koch 1997, Edelman 2000). Weitere Gesetze könnten die verschiedenen Modalitäten der Wahrnehmung mit neuronalen Parametern verbinden. Auf der anderen Seite wird man vernünftigerweise kaum erwarten, Gesetze zu finden, die etwa die Überzeugung, dass Schopenhauer ein großer Philosoph war, den Wunsch, Deutschland möge 2006 Fußballweltmeister werden, oder andere propositionale Einstellungen mit bestimmten Typen von Gehirnzuständen verbinden. Schließlich ist noch anzumerken, dass die Frage der Existenz von psychophysischen Gesetzen des beschriebenen Typs in zweierlei Hinsicht vom Problem von Freiheit und Determinismus zu trennen ist. Erstens betrifft dieses nach herkömmlicher Deutung propositionale prakti-



setz lässt sich dann natürlich – in Übereinstimmung mit dem bekannten Hempel-Oppenheim-Schema wissenschaftlicher Erklärungen – die Tatsache, dass ein bestimmter Gehirnzustand eine Bewusstseins-eigenschaft hat oder mit ihr verknüpft ist, ableiten und somit erklären. Die Ableitung einer spezifischen Instanz der Korrelationsbeziehung aus dem allgemeinen Gesetz beantwortet seine Frage nicht.

Umgekehrt hilft dieses Faktum aber auch wenig bei dem Versuch zu verstehen, auf welche Art von Antwort der Rästeltheoretiker denn abzielt. Die Art von Antwort, die gewöhnlich auf eine Frage des Typs "Warum hat x F?" gegeben wird, befriedigt ihn nicht. Was aber hat er im Sinn? Offenkundig muss es ihm um die Erklärung des Gesetzeszusammenhanges selbst gehen. Warum besteht überhaupt diese Beziehung zwischen Typen von Gehirn- und Bewusstseinsentitäten? Das einzige mir bekannte Modell für die Erklärung eines Gesetzes ist die Subsumption unter ein noch allgemeineres Gesetz. Im Standardmodell der Theorienreduktion (Nagel 1961) wird etwa angenommen, dass sich die Gesetze der reduzierten Theorie mit Hilfe von Korrelationsgesetzen aus den Gesetzen der reduzierenden Theorie ableiten lassen, und damit gelten erstere Gesetze auch als erklärt. Prinzipiell kann man sich vorstellen, dass auch Gesetze, die Gehirn- und Bewusstseinszustände verknüpfen, auf diese Weise erklärt werden könnten. Es wäre ein Gesetz denkbar, das abstraktere physikalische Merkmale, die nicht nur terrestrische Gehirne aufweisen, sondern auch andere, von ihnen ansonsten in mancherlei Hinsichten verschiedene, oder gar Entitäten, die wir gar nicht als Gehirne bezeichnen würden, mit Bewusstsein überhaupt in all seinen verschiedenen, terrestrischen und nicht-terrestrischen Formen, verbinden würde. Mit Hilfe dieses Gesetzes ließe sich dann die Tatsache, dass bestimmte menschliche oder terrestrische Gehirnzustände mit Bewusstsein verbunden sind, dadurch erklären, dass diese Gehirnzustände bestimmte abstraktere physikalische Merkmale haben. Wiederum scheint aber klar, dass auch diese Erklärung den Rästeltheoretiker nicht zufrieden stellen würde. Warum, würde er fragen, sind diese abstrakten physikalischen Parameter mit Bewusstsein verknüpft und nicht andere?

---

sche Einstellungen wie Wünsen und Beabsichtigen, für die die Existenz von Gesetzen nicht zu erwarten ist. Zweitens ist meiner Ansicht nach die Existenz von psychophysischen Korrelationsgesetzen damit vereinbar, dass die fundamentalen Gesetze probabilistisch sind, also auch mit der Falschheit des Determinismus. Für diese These kann hier nicht im Einzelnen argumentiert werden. Es sei nur darauf verwiesen, dass es trotz des probabilistischen Charakters der Quantenmechanik ja auch Korrelationsgesetze gibt, die die quantenmechanische Ebene mit der der Chemie verbinden.

### 2.3 Die Möglichkeit des Bewusstseins erklären?

Wir wissen immer noch nicht, auf welche Art von Antwort die Rätselfrage abzielt. Vielleicht hilft es, wenn wir die Frage so umformulieren: "Wie kann es sein, wie ist es überhaupt möglich, dass physische Vorgänge wie Gehirnzustände mit Bewusstseinszuständen verbunden sind?" Aber was unterscheidet die Frage nach einer Erklärung dafür ist, dass etwas möglich ist, von der Frage nach einer Erklärung dafür, dass es tatsächlich der Fall ist?

Nun ist in der philosophischen Tradition mitunter die Ansicht vertreten worden, dass Fragen der Form "Wie ist es überhaupt möglich, dass p?" oder "Wie ist x überhaupt möglich?" einen ganz eigenen – transzendentalen – Typ von Fragen darstellen. Die Kantische Frage nach den Bedingungen der Möglichkeit von Erfahrung überhaupt kann als Paradigma für eine transzendente Fragestellung gelten. Aber selbst abgesehen davon, dass es sehr umstritten ist, ob es tatsächlich eine eigenständige und legitime spezifisch transzendente Art des Fragens gibt, scheint klar, dass, auch wenn man dies bejaht, diese Art des Fragens dem hier zur Debatte stehenden Gegenstand nicht angemessen ist. Die paradigmatische Kantische Frage soll ja durch den Aufweis von Elementen, die notwendigerweise Teil einer jeden Erfahrung sein müssen, beantwortet werden. Es ist nicht zu sehen, wie dieses Schema auf das Verhältnis zwischen Gehirn- und Bewusstseinszuständen angewendet werden soll. Die transzendente Fragestellung scheint hier nicht relevant zu sein.

Im Alltag drücken wir mit Ausrufen wie „Ja, ist das denn die Möglichkeit?!“ unser Erstaunen über einen Sachverhalt aus, der unseren Erwartungen zuwiderläuft. Der Verdacht, dass auch die Atmosphäre der Rätselhaftigkeit, die die Beziehung zwischen dem Bewusstsein und seiner materiellen Basis umgibt, etwas damit zu tun haben könnte, dass die Art dieser Beziehung bestimmten Erwartungen widerspricht, lässt sich nur schwer von der Hand weisen. Die Annahme einer naturgesetzlichen Verbindung des Bewusstseins mit seiner physischen Basis konfligiert mit der dualistischen Konzeption des Mentalen und des Physischen als zweier fundamental verschiedener Seinsbereiche. Und auch für den physikalistischen Monisten, der meint, dass eine rein physikalische Beschreibung der Welt ontologisch vollständig sein kann, ist sie problematisch. Anstößig ist dabei nicht das Fehlen einer Erklärung, sondern dass die Existenz dieser Verbindung eine *Tatsache* zu sein scheint, die in seiner physikalistischen Ontologie nicht vorkommt. Dies legt den Verdacht nahe, dass der Anschein der Rätselhaftigkeit der Verbindung zwischen Bewusstsein und Gehirn

zumindest zum Teil daher rührt, dass ihre bloße Existenz dualistischen und monistischen Erwartungen widerspricht.

Leider hilft der – an sich auch wenig überraschende – Verdacht, dass das Bewusstsein vor allem vor dem Hintergrund bestimmter metaphysischer Annahmen problematisch oder gar rätselhaft erscheint, uns auch nicht dabei, die Frage zu beantworten, welche Art von Erklärung der Rätseltheoretiker eigentlich fordert. Es gibt keine charakteristische Art der Erklärung von Sachverhalten, die Erwartungen zuwiderlaufen. Sie werden erklärt wie andere Sachverhalte auch. Wenn wir gute Gründe haben, ihre Existenz anzuerkennen, zwingen sie uns allerdings, unsere Erwartungshaltung zu verändern und uns an die überraschenden Tatsachen zu gewöhnen. Dies scheint nach unseren bisherigen Überlegungen auch hier der Fall zu sein.

## **2.4 Das Bewusstsein als begrifflich notwendig erweisen?**

Kommen wir zu einem weiteren Versuch der Präzisierung der Rätselfrage. Unter einer Erklärung wird ja in der Philosophie oft auch eine Explikation oder Analyse eines Begriffs verstanden, von der angenommen wird, dass sie durch eine begrifflich notwendige Verbindung zwischen Erklärtem und Erklärung charakterisiert wird. Nach einer traditionellen Konzeption von Philosophie ist das Fragen nach Erklärungen in diesem Sinn sogar das Spezifische philosophischen Fragens. Könnte es sein, dass die Rätselfrage auch aus diesem philosophischen Impetus heraus entsteht und eine Erklärung fordert, die die Emergenz des Bewusstseins als in irgendeinem Sinne begrifflich – und nicht bloß naturgesetzlich – notwendig, gegeben seine materielle Basis, erweist? Dies würde erklären, warum aus der Perspektive des Rätseltheoretikers alle bisher besprochenen Erklärungen als unzureichend erscheinen müssen.

Nun haben zwar in der Geschichte der Leib-Seele-Debatte logische Behavioristen und analytische Funktionalisten die Existenz begrifflich notwendiger Verbindung zwischen der Bewusstseinsbegrifflichkeit bzw. der Begrifflichkeit mentaler Zustände und einer nicht-mentalistischen Begrifflichkeit behauptet. Aber dabei handelte es sich erstens nicht um Begriffe von Gehirnzuständen, sondern um Begriffe äußeren Verhaltens; und zweitens scheinen diese Auffassungen, jedenfalls insofern der Anspruch erhoben wird, die Bewusstseinsbegrifflichkeit vollständig ‚wegzuanalysieren‘, Versuche zu sein, das Bewusstsein zu eliminieren. Wer dagegen das Bewusstsein als ein konkretes, nicht auf Verhalten oder Verhaltensvermittlung redu-

zierbares Phänomen erkennt, wird kaum auf die Idee kommen, eine begrifflich notwendige Verbindung zwischen dem Bewusstsein und seinen neuronalen Korrelaten zu behaupten oder nach einer solchen zu suchen. Auch dieser Versuch der Explikation der Rätselfrage scheint also zu scheitern, bevor er auch nur richtig in Gang gekommen ist.

Die Vorstellung einer begrifflich notwendigen Verbindung kann aber noch auf andere Weise ins Spiel kommen, nämlich wenn die Überzeugung, dass es sie tatsächlich *nicht* gibt, mit dem Gedanken verbunden wird, es *müsse* sie eigentlich doch geben. Auf diese Weise kommt diese Vorstellung in David Chalmers' Entgegensetzung des „leichten Problems“ (*easy problem*) der Erklärung des ‚funktionalen‘ Bewusstseins und des „schweren Problems“ (*hard problem*) der Erklärung des (phänomenalen) Bewusstseins ins Spiel. Dieser in jüngster Zeit vielleicht einflussreichste Versuch zur Explikation des Bewusstseinsrätsels soll nun diskutiert werden.

## **2.5 ‚Funktionales‘ vs. ‚phänomenales‘ Bewusstsein erklären**

Chalmers (1996) stellt die These auf, die Gegebenheiten auf den verschiedenen physischen Makroebenen und der Ebene des ‚funktionalen Bewusstseins‘ ließen sich aus den mikrophysikalischen Tatsachen mit Hilfe von begrifflich notwendigen Verbindungen, so genannten „Supervenienzkonditionalen“, ableiten, für das (phänomenale) Bewusstsein gelte dies aber nicht. Die Verbindung zwischen den Tatsachen des (phänomenalen) Bewusstseins und den mikrophysikalischen Tatsachen sei nur naturgesetzlich notwendig. Damit soll der Materialismus widerlegt sein. Dies setzt voraus, dass der Materialismus als die These verstanden wird, dass alle Tatsachen begrifflich auf den mikrophysikalischen Tatsachen supervenieren, oder dass die mikrophysikalischen Tatsachen alle Tatsachen mit begrifflicher Notwendigkeit determinieren. Die letzte Formulierung scheint nur eine andere Beschreibung desselben Sachverhalts zu sein.

Wir haben im letzten Abschnitt schon bemerkt, dass die so verstandene materialistische These die Elimination des (phänomenalen) Bewusstseins beinhaltet oder voraussetzt: es muss ‚weganalysiert‘ werden. Demnach würde die bloße Existenz des (phänomenalen) Bewusstseins den Materialismus widerlegen. Das (phänomenale) Bewusstsein frustriert also die Erwartung des Materialisten, dass die mikrophysikalischen Tatsachen alle Tatsachen determinieren.

Chalmers stützt seine These vor allem durch Gedankenexperimente, die zeigen sollen, dass man sich, gegeben die mikrophysikalischen Tatsachen, vorstellen kann, dass die Tatsachen des (phänomenalen) Bewusstseins anders sein könnten, als sie tatsächlich sind. Für die physischen Makrotatsachen und das ‚funktionale Bewusstsein‘ soll dies nicht gelten. Demzufolge kann man sich etwa eine Welt vorstellen, die mikrophysikalisch mit der unsrigen identisch ist, in der aber die Bewusstseinstatsachen ganz andere sind, oder es gar kein Bewusstsein gibt. Aber man kann sich nicht eine mit der unsrigen mikrophysikalisch typenidentische Welt vorstellen, in der die Wälder andere Ausdehnungen haben oder es überhaupt keine Bäume gibt. Und man kann sich auch nicht eine mit der unsrigen mikrophysikalisch typenidentische Welt vorstellen, in der die Menschen kein ‚funktionales Bewusstsein‘ haben. Wohl könnte ihnen aber in dieser Welt (phänomenales) Bewusstsein völlig fehlen. Dies wäre dann eine ‚Zombie‘-Welt.

Nach Chalmers Konzeption liegt also die Rätselhaftigkeit des (phänomenalen) Bewusstseins, oder das schwere Problem, vor das es uns stellt, darin, dass es nicht begrifflich reduktiv durch die Mikroebene erklärt werden kann. Im Fall des Bewusstseins können keine begrifflich notwendigen Verbindungen zur Mikroebene (bzw. der entsprechenden Begrifflichkeit) hergestellt werden. Chalmers möchte es aber nicht dabei bewenden lassen, die Rätselhaftigkeit des Bewusstseins zu konstatieren. Das schwere Problem soll sehr schwer, aber nicht unlösbar sein. Eine positive, konstruktive Reaktion auf das Bewusstseinsrätsel soll möglich sein. Das Bewusstsein kann zwar unter Bezug auf die neurobiologische Ebene allein nicht befriedigend erklärt werden. Auf dieser Ebene ließe sich nur das leichte Problem der Erklärung des ‚funktionalen Bewusstseins‘ lösen. Eine neue Grundlagentheorie könnte aber die Lösung bringen. Sie könnte z.B. in einer neuen, naturalistischen Form des Dualismus bestehen oder im eingangs schon erwähnten Russellschen Monismus. (Chalmers diskutiert mehrere Vorschläge, ohne sich auf einen festzulegen.) Diese Grundlagentheorie würde die uns vertrauten Bewusstseinsphänomene wie Tastempfindungen, Schmerzen oder Freude unter Bezug auf fundamentale Bewusstseinsentitäten erklären, für die Chalmers den Begriff „protophänomenaler“ Eigenschaften verwendet.

Wir hatten schon eingangs herausgearbeitet, dass die Postulierung solch fundamentaler Bewusstseinsentitäten die Tendenz hat, schnurstracks zum Panpsychismus zu führen. Chalmers fühlt sich deshalb auch dahin gedrängt, den Panpsychismus ernsthaft zu erwägen. Wir sollen uns also eine Vorstellung von protophänomenalen

Eigenschaften machen, die auch z.B. Tischen oder Felsen zugeschrieben werden könnten. Diese Vorstellung mit konkretem Inhalt zu füllen, bleibt notgedrungen dem Leser überlassen. Ich möchte diese Bedenken aber bis auf weiteres zurückstellen und mich zuerst wieder dem mittlerweile vertrauten Problem widmen, ob es Chalmers denn gelungen ist, das Bewusstseinsrätsel in eine beantwortbare Frage zu transformieren.

Fragen wir uns zuerst, wie denn die explanatorische Beziehung zwischen den uns allen vertrauten Bewusstseinsphänomenen und den protophänomenalen Eigenschaften aussehen soll. Ich glaube, dass Chalmers an dieser Stelle vor einem Dilemma steht. Entweder diese Beziehung ist begrifflich notwendig, oder sie ist es nicht. Wenn sie es nicht ist, ist schwer zu verstehen, was das ganze Manöver in Bezug auf das Erklärungsproblem eingebracht haben soll. Es ging ja von der Prämisse aus, die Erklärung des Bewusstseins müsse mit begrifflicher Notwendigkeit erfolgen. Es sollte ein Problem darstellen, dass wir über eine solche Erklärung nicht verfügen. Wenn wir aber nun die Beziehung zwischen protophänomenalen Eigenschaften und den gewöhnlichen Bewusstseinsphänomenen als ein *factum brutum* akzeptieren sollen, fragt man sich, warum wir dann nicht auch die Beziehung zwischen dem Bewusstsein und seinen neuronalen Korrelaten als ein solches akzeptieren können. Wenn dieses Horn des Dilemmas ergriffen wird, haben wir nur ein *factum brutum* durch ein anderes ersetzt und uns dabei wahrscheinlich den Panpsychismus, auf jeden Fall aber mit den protophänomenalen Eigenschaften Entitäten eingehandelt, von denen wir keinen klaren Begriff haben. Das andere Horn des Dilemmas besteht darin, dass die Vorstellung, es gebe begrifflich notwendige Verbindungen zwischen unseren Bewusstseinsbegriffen und den Begriffen protophänomenaler Eigenschaften mit jeder geläufigen Vorstellung solch begrifflich notwendiger Verbindungen unvereinbar zu sein scheint. Es kann wohl kaum behauptet werden, dass die Vorstellung protophänomenaler Eigenschaften zu unserem Bewusstseinsbegriff gehört. Vielmehr haben wir ja gar keinen klaren Begriff von protophänomenalen Eigenschaften. Und selbst wenn wir diesem Begriff einen klareren Sinn *geben* könnten, so würde dies nur zeigen, dass dieser klarere Begriff protophänomenaler Eigenschaften nicht zu unserem gegenwärtigen Begriff des Bewusstseins gehört und deshalb mit ihm auch nicht durch eine begrifflich notwendige Verbindung verknüpft sein kann. Ferner ist die Vorstellung, die Verbindung sei begrifflich notwendig, nach landläufigem Verständnis auch damit unvereinbar, dass die neue Grundlagentheorie eine empirisch stützba-

re naturalistische Theorie sein soll. Demnach müssten die fundamentalen Eigenschaften *entdeckt* werden, so wie die Physiker einst die fundamentalen Eigenschaften der Materie entdeckt haben, nicht durch Begriffsanalyse aus unserem gewöhnlichen Bewusstseinsbegriff herausgewickelt werden.

Wenn es hier allein um die Widerlegung der Chalmersschen Theorie ginge, könnten wir die Diskussion hiermit abschließen. Im Kontext der in der Einleitung umrissenen Hauptargumentation der vorliegenden Untersuchung interessiert uns seine Konzeption aber nicht zuletzt als ein Beispiel dafür, zu welchen Extrempositionen die stillschweigende Akzeptanz der Elimination der manifesten physischen Welt diejenigen führen kann, die sie mit einem Realismus bezüglich des Bewusstseins verbinden. Denn ich möchte im Folgenden zeigen, dass seine Überlegungen tatsächlich nur vor dem Hintergrund dieser Elimination so plausibel, ja zwingend erscheinen können und es in gewisser Hinsicht auch sind. Man könnte nämlich sagen, dass Chalmers die in der Standardmetaphysik sekundärer Eigenschaften angenommene Sonderstellung des Bewusstseins konsequent bis an ihr logisches Ende denkt – oder, genauer, an das eine ihrer Enden, den Bewusstseinsfundamentalismus –, allerdings ohne zu bemerken, dass er damit eine *reductio ad absurdum* dieser Metaphysik liefert.

Kehren wir also zur Wurzel der Chalmersschen Überlegungen zurück. Diese besteht ja in der These einer fundamentalen Asymmetrie im Verhältnis zwischen mikrophysikalischen Gegebenheiten und makrophysischen Tatsachen auf der einen Seite und mikrophysikalischen und Bewusstseinstatsachen auf der anderen. Diese Asymmetrie lässt das Bewusstsein als etwas erscheinen, das einer ganz besonderen Erklärung bedarf. Wir haben gesehen, dass die Vorstellung einer solchen Erklärung in begriffliche Aporien führt. Ich möchte nun Einklang mit der in dieser Arbeit verfolgten Argumentationsstrategie zeigen, dass diese Asymmetrie in Wirklichkeit nicht besteht, oder dass sie jedenfalls nur dann besteht, wenn wir die manifeste physische Welt tatsächlich eliminieren. Wenn wir unbefangen an die Sache herangehen, ist die These, dass makrophysische Gegebenheiten auf mikrophysikalischen begrifflich supervenieren, nicht plausibel.

## 2.6 Sind physische Makrophänomene reduktiv erklärbar?

Um diese These angemessen diskutieren zu können, müssen wir uns zuerst über das Konzept begrifflicher Wahrheiten größere Klarheit verschaffen. Ich hatte dazu oben bei der Diskussion der Erklärung des Bewusstseins durch protophänomenale Eigenschaften bereits einige Bemerkungen gemacht, die aber noch präzisiert werden sollten. Dies gilt besonders angesichts der Tatsache, dass Chalmers und auch Jackson das Konzept in sehr eigener Weise interpretieren.

Das Konzept begrifflicher oder analytischer Wahrheiten *a priori* hat eine lange und verwickelte Geschichte, besonders natürlich in der Tradition der analytischen Philosophie, von der man ja meinen könnte, dass dieser Begriff für ihr Selbstverständnis zentral ist, wenn dies auch spätestens seit Quines (1953) berühmter Kritik nur noch sehr eingeschränkt zutrifft. Es ist hier nicht möglich, diese Geschichte oder den derzeitigen Stand der Debatten auch nur ansatzweise nachzuzeichnen. Ich möchte nur ein wesentliches Merkmal der traditionellen Konzeption begrifflicher Wahrheiten herausarbeiten, nämlich dass zwischen den in einer solchen Wahrheit miteinander verbundenen Begriffen ein- oder beidseitige Abhängigkeitsverhältnisse bestehen. Dabei kann man eine stärkere und eine schwächere Version des Konzeptes unterscheiden.

Stärkere Version: Eine begriffliche Wahrheit ist eine Analyse eines Begriffs, so dass der Begriff durch das Analysans ohne Bedeutungsänderung ersetzt werden kann; Analysans und Analysandum sind synonym.

Schwächere Version: Eine begriffliche Wahrheit ist eine Verbindung zwischen zwei Begriffen, so dass diese Verbindung für die Bedeutung zumindest eines dieser Begriffe wesentlich ist; d.h. wenn diese Verbindung aufgelöst würde, hätte sich die Bedeutung dieses Begriffs geändert, er wäre nicht mehr derselbe Begriff.

Die stärkere Version entspricht z.B. der von Quine diskutierten Konzeption begrifflicher Wahrheiten. Die traditionelle Charakterisierung einer Begriffsanalyse als eines Vorganges, bei dem etwas aus einem Begriff hervorgeholt wird, was in ihm bereits enthalten ist, wenngleich möglicherweise nur „verworren“ (Kant, KdrV: B 10), ist prinzipiell mit beiden Versionen vereinbar. Der herausanalyisierte Begriff könnte im



Verein mit anderen einen komplexen Ausdruck bilden, der mit dem analysierten Begriff synonym ist, oder die Verbindung zu dem enthaltenen Begriff könnte für die Identität des ihn enthaltenden Begriffes wesentlich sein, ohne dass es doch möglich wäre, ihn vollständig in anderen Begriffen zu analysieren. Als ein Beispiel für die letztere Art der Verbindung könnte man etwa die Beziehung zwischen den Begriffen der Handlung und der Absicht auffassen. Dass zumindest manche, wenn nicht gar die meisten, Handlungen von entsprechenden Absichten verursacht werden, kann wohl als ein plausibler Kandidat für eine begrifflich notwendige Verbindung gelten. (Da manches Tun/Handeln unabsichtlich ist, kann nicht jede Handlung von einer *entsprechenden* Absicht verursacht worden sein. Man kann aber argumentieren, dass auch in solchen Fällen das Tun von einer Absicht angeleitet worden sein muss – wenn auch nicht von der entsprechenden – damit es sich überhaupt um eine Handlung handelt.) In diesem Fall ist auch die Verbindung in umgekehrter Richtung ein Kandidat für eine begrifflich notwendige Beziehung: eine Absicht ist stets eine Absicht, etwas zu tun. Aber weder sind die beiden Begriffe Synonyme, noch muss man unbedingt annehmen, dass es möglich ist, mit Hilfe anderer Ausdrücke komplexe Synonyme dieser Ausdrücke zu bilden. Es könnte sein, dass die genannten (und möglicherweise weitere) Verbindungen dieser Ausdrücke miteinander für ihre Bedeutung wesentlich und damit begriffliche Wahrheiten sind, ohne dass doch die Vorstellung einer vollständigen, zu einem synonymen Ausdruck führenden Bedeutungsanalyse einen klaren Sinn hätte.

Sowohl in der stärkeren als auch in der schwächeren Version gilt aber mindestens in einer Richtung, dass nur der über den einen Begriff verfügen kann, der auch über den anderen verfügt. Im gerade diskutierten Beispiel gilt möglicherweise beides: man kann nur über den Begriff der Absicht verfügen, wenn man auch den der Handlung hat und umgekehrt. Hier liegt eine wechselseitige begriffliche Abhängigkeit vor. Beispiele für eine bloß einseitige begriffliche Abhängigkeit könnten die Beziehungen zwischen solchen Begriffen wie „Mensch“, „Mann“, „Frau“ auf der einen und Begriffen von Institutionen wie „Ehe“, „Familie“, aber auch „Regierung“, „Parlament“ und unzähligen anderen sein. Hier könnte man argumentieren, dass die letzteren Begriffe zwar die ersteren voraussetzen, dass eine umgekehrte Abhängigkeit aber nicht besteht.

Vorausgesetzt, dass dies als grundlegende Bedingung für einen adäquaten Begriff der Bedeutungswahrheit akzeptiert wird, lässt sich nun relativ leicht zeigen,

dass die so genannten Supervenienzkonditionale dieser Bedingung nicht genügen. Supervenienzkonditionale sollen ja Aussagen der Form „Wenn die Mikrotatsachen bezüglich einer Situation X sind, sind die Makrotatsachen Y“ sein (cf. Chalmers 1996: 52). Nehmen wir z.B. an, wir verfügten über eine vollständige Beschreibung der mikrophysikalischen Verhältnisse und Geschehnisse während einer Zeit von 10 Minuten in einem bestimmten Waldstück (die relevante Situation). Nach der zur Debatte stehenden Auffassung muss es dann entsprechende Supervenienzkonditionale geben, die in der mikrophysikalischen Begrifflichkeit beschriebene Tatsachen in Verbindung z.B. zu in der Makrobegrifflichkeit der Biologie beschriebenen Tatsachen setzen würden und den Status begrifflicher Wahrheiten hätten. Demnach wäre es dann, gegeben eine Beschreibung eines bestimmten Raumzeitinhalts in mikrophysikalischer Begrifflichkeit, z.B. begrifflich notwendig, dass es in diesem Raumzeitinhalt auch bestimmte Pflanzen und Tiere gibt, dass bestimmte Dinge mit diesen geschehen etc. Diese Makrotatsachen können aus den Mikrotatsachen abgeleitet oder ‚abgelesen‘ werden.

Nun wird aber wohl niemand behaupten wollen, dass mikrophysikalische Ausdrücke mit biologischen synonym sind, und auch wenn man die schwächere Version des Konzeptes begrifflicher Wahrheiten zu Grunde legt, ist die These, dass es sich hier um solche handelt, nicht haltbar. Denn sicherlich ist es weder notwendig, dass jemand, der über die mikrophysikalische Begrifflichkeit verfügt, auch die relevanten biologischen Begriffe hat, noch dass die umgekehrte Beziehung besteht. Und auch die Behauptung, die biologischen ließen sich aus den mikrophysikalischen Tatsachen ‚einfach ablesen‘, ist kaum verteidigbar. Nehmen wir an, ein Physiker bekäme eine vollständige mikrophysikalische Beschreibung der Gegebenheiten in dem besagten Waldstück über einen Zeitraum von 10 Minuten. (Diese Vorstellung ist schon sehr utopisch, und es ist auch fraglich, ob die Rede von einer ‚vollständigen mikrophysikalischen Beschreibung‘ einen sehr klaren Sinn hat, aber akzeptieren wir diese Voraussetzungen einmal um des Argumentes willen.) Wiederum wird er dadurch weder notwendigerweise *ipso facto* über die relevanten Speziesbegriffe etc. verfügen, noch lassen sich diese aus der mikrophysikalischen Begrifflichkeit irgendwie ableiten oder konstruieren. Und auch wenn unser Physiker über die relevante biologische Begrifflichkeit verfügt, ist ihm dadurch die Verbindung zwischen Mikrophysik und Biologie noch keineswegs *a priori* gegeben. Er könnte zwar prinzipiell lernen, anhand der Verteilung mikrophysikalischer Eigenschaften in einer bestimm-

ten Raumzeit z.B. die Präsenz eines Tieres einer bestimmten Art zu erkennen. Es soll hier natürlich keineswegs bestritten werden, dass Tiere oder makroskopische Entitäten überhaupt prinzipiell durch ihre mikrophysikalische Struktur identifizierbar sind.<sup>22</sup> Eine solche Identifikation beruht aber auf spezifischem empirischem Wissen, das sich *a priori* weder aus der makroskopischen Begrifflichkeit ergibt, noch aus der mikrophysikalischen, noch aus deren bloßer Addition.

Der grundlegende Punkt lässt sich auch an einem verwandten, aber viel realistischerem Beispiel illustrieren. Betrachten wir einen Medizinstudenten, der sowohl etwas von der Physik der Röntgenstrahlen versteht, so dass er weiß, welche physikalischen Eigenschaften eine Röntgenaufnahme reflektiert, als auch von den verschiedenen Erkrankungen der Lunge. Dennoch muss er erst noch lernen, diese Krankheiten auf dem Röntgenbild zu erkennen. Er muss seinen visuellen Zugang zur Realität trainieren, gewisse Fähigkeiten der Mustererkennung erwerben, und diese mit seinem medizinischen Wissen verknüpfen. Die Lungentatsachen sind ihm keineswegs ohne weiteres durch sein physikalisches Wissen gegeben.

Eine andere wichtige Überlegung betrifft die Projizierbarkeit der jeweiligen Begrifflichkeiten. Bisher haben wir Fälle betrachtet, wo die Makrogegebenheiten auf der Basis von Wissen in Mikrobegrifflichkeit bestimmt werden konnten und nur darauf bestanden, dass dies empirisch etablierte Verbindungen zwischen den Ebenen voraussetzt. Fragen wir uns nun, was geschieht, wenn die Makrobegrifflichkeit auf neue, strittige Fälle angewandt werden soll. Ich denke, es ist einleuchtend, dass auch die Entscheidungen, die wir dort treffen müssen, sich aus der Mikrobegrifflichkeit nicht ableiten oder -lesen lassen. Nehmen wir wieder einmal zumindest *arguendo* an, dass wir über eine vollständige mikrophysikalische Beschreibung von etwas verfügen, diesmal der Situation auf der Erde in, sagen wir, einer Million Jahren, und nehmen wir weiterhin an, dass es dann auf der Erde noch Wesen gibt, die direkt von den heutigen Menschen abstammen. Der Punkt ist nun, dass durch die mikrophysikalischen oder selbst durch die molekularbiologischen Tatsachen nicht festgelegt ist, ob diese Wesen zur Spezies *Homo sapiens sapiens* gehören oder auch, ob sie überhaupt Menschen sind. Wissen über Mikrotatsachen kann zwar Entscheidungen bezüglich solcher Fragen beeinflussen, aber die Antworten lassen sich nicht einfach von den Mikrotatsachen ablesen. Dies lässt sich wieder mittels eines verwandten realen Beispiels veranschaulichen. Inspiriert durch neue Ergebnisse der Molekulargenetik, de-

---

<sup>22</sup> Was aber keineswegs bedeutet, dass sie *mit* ihrer mikrophysikalischen Struktur identifiziert werden können.

nen zufolge es eine über 98%ige Übereinstimmung zwischen dem Genom des heutigen Menschen und dem der Bonobos gibt, haben einige Wissenschaftler vorgeschlagen, die Bonobos zur Gattung Homo zu zählen. Dies illustriert, dass Mikrodaten für Makrotaxonomien relevant und motivierend sein können, aber es zeigt auch, dass es keinen automatischen Übergang von ersteren zu letzteren gibt. Die wissenschaftliche Debatte wird durch den Hinweis auf das Ausmaß der genetischen Übereinstimmung eingeleitet, nicht beendet. Die Mikrobefunde müssen zuerst eingeordnet, ihre Relevanz für die Makrotheorie und -begrifflichkeit erst bestimmt werden. Dazu muss die Makrobegrifflichkeit weiterentwickelt werden, und dies geschieht primär aufgrund der ihr inhärenten Logik und vor allem im Hinblick auf eine angemessene Einordnung der in der Makrobegrifflichkeit beschriebenen relevanten Tatsachen, in unserem Fall etwa den anatomischen Tatsachen, an denen die Taxonomie von Arten und Gattungen immer noch hauptsächlich orientiert ist.

All diese Beobachtungen deuten auf eine *relative begriffliche Autonomie* von Mikro- und Makrobegrifflichkeiten, die mit der Vorstellung, die Makrotatsachen ließen sich aus den relevanten Mikrobeschreibungen einfach ableiten, unvereinbar ist. Zweifelsohne sind in der modernen Wissenschaft die Grenzen zwischen den Disziplinen durchlässiger und flexibler geworden, als dies früher der Fall war. Eine charakteristische Entwicklung ist ja etwa die Entstehung neuer Fächer, die traditionelle Disziplinen unter einem neuen Dach verknüpfen, wie z.B. die Biochemie, Biophysik, Molekulargenetik, Geophysik, Neuropsychologie etc. Und biologische oder geologische Forschung, die völlig ohne die Hilfe von Methoden auskommt, die auf chemischen oder physikalischen Erkenntnissen beruhen, ist heutzutage kaum noch vorstellbar. Dies bedeutet aber keineswegs, dass sich die zentralen Begrifflichkeiten der relevanten Disziplinen gar nicht voneinander unterscheiden lassen, oder dass z.B. gar die Unterscheidung zwischen dem charakteristischen Vokabular der Biologie und dem der Physik hinfällig würde. Es ist immer noch etwa möglich, manche Kernbegriffe der Biologie zu erwerben, ohne über (wissenschaftliche) physikalische Begriffe und Kenntnisse zu verfügen. Und nur innerhalb des Bereichs der Kernbegrifflichkeit einer Disziplin oder Theorie lassen sich wirklich plausible Kandidaten für den Status einer begrifflichen Wahrheit finden – vorausgesetzt, man ist der Meinung, dass sich eine sinnvolle Grenze zwischen begrifflichen und nicht-begrifflichen Wahrheiten überhaupt ziehen lässt. Dass Mitglieder derselben Spezies sich (unter normalen Umständen) miteinander fortpflanzen können, wäre z.B. ein Kandidat für eine begriffli-

che Wahrheit in der Biologie. Aussagen über die molekulare Struktur des Genoms sind es dagegen nicht.

Kommen wir nun zur behaupteten Asymmetrie in Bezug auf die Vorstellbarkeit der Abwesenheit von (phänomenalen) Bewusstseins- und physischen Makrotatsachen.

## **2.7 Die Abwesenheit von Bewusstsein und anderen Makroentitäten**

Wie schon erwähnt, wird die These einer prinzipiellen Disanalogie zwischen dem Verhältnis zwischen Gehirn und Bewusstsein und anderen Fällen von Mikro-Makrobeziehungen auch oft durch Behauptungen über die Vorstellbarkeit bestimmter Szenarien illustriert und zu plausibilisieren versucht. Die beliebtesten Szenarien sind dabei diejenigen mit so genannten philosophischen Zombies, die sich im Unterschied zu den aus Filmen bekannten Kreaturen wie normale Menschen verhalten, denen aber jedes (phänomenale) Bewusstsein fehlt. So wird dann etwa behauptet, ein Zombiedoppelgänger eines normalen Menschen, der mit diesem behavioral und auch bis zum letzten Molekül oder gar Quark oder Myon identisch ist, sei vorstellbar oder denkbar, ein mikrotypenidentischer Zwilling eines Baumes z.B., der kein Baum ist, aber nicht. Ich werde dafür argumentieren, dass hier kein prinzipieller Unterschied vorliegt.

Insofern es hier nur um die Frage der Denkbarkeit der relevanten Szenarien geht, könnte man meinen, dass der entscheidende Punkt im Wesentlichen schon gemacht worden ist. Denn wenn erstens die Schlussfolgerung des letzten Abschnitts richtig ist, dass es keine begrifflichen Wahrheiten gibt, die Mikro- und Makro Begrifflichkeiten verbinden, und zweitens die ja wohl allgemein akzeptierte Annahme, dass alles denkbar ist, wenn es nicht die Negation einer begrifflichen Wahrheit bedeutet, dann muss auch z.B. denkbar sein, dass etwas die Mikrostruktur eines Baumes besitzt, aber kein Baum ist. Dies wird aber vielen sehr kontraintuitiv erscheinen, so dass hier noch Klärungsbedarf besteht. Auch ist noch nichts über Vorstellbarkeit im engeren Sinn gesagt worden, d.h. die Rolle, die anschauliche Vorstellungen in diesem Zusammenhang spielen. Und da anschauliche Vorstellungen auch den Inhalt von Begriffen mitbestimmen, wird eine Klärung ihrer Rolle auch Licht auf die Problematik der Denkbarkeit werfen.

Fragen wir uns, was es eigentlich bedeutet, dass man sich einen Zombie vorstellen oder denken kann. Wie macht man das überhaupt? Dazu ist es wesentlich sich klarzumachen, dass es hier um einen negativen Sachverhalt geht, die Abwesenheit von etwas, und noch dazu um die Abwesenheit von etwas, das zu einer anderen begrifflichen Ebene gehört bzw. sozusagen um die Abwesenheit der ganzen Ebene. Es ist nicht so, wie wenn man sich einmal ein grünes und dann, um sich die Abwesenheit der Eigenschaft, grün zu sein, zu veranschaulichen, ein blaues Osterei vorstellt – wobei auch in diesem Fall die wesentlich begrifflich-abstrakte, nicht-anschauliche Operation der Negation eine wesentliche Rolle spielt. Es ist nicht einmal so, wie wenn man sich den gleichen Raum einmal voller Möbel vorstellt und einmal leer. Noch würde es helfen, sich den Schädel sehr voll vorzustellen – so dass sozusagen gar kein Bewusstsein mehr hineinpasst. Der Abwesenheit des Bewusstseins entspricht keine anschauliche Vorstellung. Sie kann nur gedacht werden durch die abstrakte Operation der Negation. Ich könnte mir also ein Gehirn vorstellen, oder auch, soweit mir das gelingt, das neuronale Korrelat eines Bewusstseinszustandes oder gar die ganze zelluläre Mikrostruktur eines menschlichen Körpers. Danach könnte ich mir z.B. vorstellen, wie die Welt diesem Organismus visuell oder auditorisch erscheint, diese Vorstellung aber mit einem Kopfschütteln oder anderen Zeichen der Verneinung begleiten. Schließlich würde ich dabei denken, dies sei mein Vorstellen/Denken von etwas, das auch der Fall sein könnte.

Nach dem gleichen Muster können wir uns vorstellen, dass ein Mikrozwilling eines Baums auch kein Baum hätte sein können. Um uns die Mikrostruktur des Baumes vorzustellen, könnten wir die Verteilung der Mikroentitäten in der vom Baum eingenommenen Raumzeit z.B. durch Zahlenwerte angeben, die Positionen in einem Koordinatensystem bezeichnen, oder wir könnten uns verschiedener Methoden der Visualisierung solcher Mikrobeschreibungen bedienen. Den Baum würden wir visuell und vielleicht noch auditorisch, durch das Rauschen der Blätter im Wind, imaginieren. Demnach würde man zuerst die Mikrostruktur in einer oder mehrerer der beschriebenen Weisen imaginieren, dann den Baum und dann die Operation der Negation ausführt indem man z.B. den Kopf schüttelt zum Zeichen, dass man sich vorstellt bzw. denkt, dass die Mikrostruktur auch nicht die eines Baumes hätte sein können.

Der wesentliche Punkt ist, dass es für die Frage, ob die Vorstellung von  $x$  bei gleichzeitiger Abwesenheit einer Eigenschaft  $Y$  von  $x$ , eine mögliche, sinnvolle, ko-

härente Vorstellung ist, entscheidend ist, ob es möglich ist, sich *x vorzustellen, ohne sich gleichzeitig Y vorzustellen*. Wenn dies der Fall ist, ist es leicht, sich die Abwesenheit von *x* zu *denken*, indem man sich *x* vorstellt und die Operation der Negation ausführt. Mir scheint, dass man sich nach diesem Schema die Abwesenheit von Bäumen bei gleichen Mikrogegebenheiten genauso vorstellen kann wie die von Bewusstsein.

## 2.8 Fazit

Wir haben eine Vielzahl von Beobachtungen zusammengetragen, die belegen, dass die Verbindungen zwischen physikalischen Mikro- und physischen Makrotatsachen nicht begrifflich notwendig sind. Und dabei haben wir den Realismus in Bezug auf die manifesten Eigenschaften wie Farbe, Wärme, Töne usw. noch nicht einmal unmittelbar ins Spiel gebracht. Wenn wir dies tun, ist natürlich sofort klar, dass die entsprechenden Mikro-Makro-Verbindungen nicht begrifflich notwendig sind. Chalmers Konzeption des schweren Problems setzt also die Elimination der manifesten Welt voraus. Die Asymmetriethese kann nur plausibel erscheinen, weil die physischen Makroentitäten stillschweigend auf bloße Aggregate von Mikroentitäten reduziert werden. Wenn man sich darüber keine Rechenschaft ablegt, entsteht die Illusion, es sei eine völlig transparente Erklärung gegeben worden, die die Existenz der physischen Makrowelt als notwendig, gegeben die mikrophysikalische Welt, erweist. Das Fehlen einer solchen Erklärung für das Bewusstsein erscheint als eine Anomalie, die den Sonderstatus des Bewusstseins zeigt. Dieser fordert die Suche nach einer Fundamentaltheorie des Bewusstseins. Aber die Vorstellung einer völlig transparenten Erklärung hat keinen klaren Sinn. Dies zeigte sich bei der Diskussion von Chalmers' Konzeption der Erklärung von Bewusstsein durch protophänomenale Eigenschaften. Wenn diese begriffliche Notwendigkeit mit sich führen sollte, müsste der Begriff phänomenaler Eigenschaften schon mit dem Bewusstseinsbegriff verknüpft sein. Das ist klarerweise nicht der Fall. Wenn sie aber keine begriffliche Notwendigkeit mit sich führt und sich auf ein *factum brutum* bezieht, gibt es keinen Grund, warum man nicht auch eine Erklärung akzeptieren sollte, die das Verhältnis zwischen dem Bewusstseinszustand und seinen neuronalen Korrelaten als ein *factum brutum* akzeptiert. Dass ein gewisser begrifflicher Abstand zwischen Erklärung und Erklärtem besteht und damit in gewissem Sinne auch ein *factum brutum*, ist tatsächlich der

Normalfall der Erklärung. Die diachronen und synchronen Erklärungen von Bewusstseinsphänomenen, die wir eingangs Revue passieren ließen, waren tatsächlich völlig adäquate Erklärungen, die allen Standards wissenschaftlicher Erklärung entsprechen. Das Bewusstseinsrätsel kann also bestenfalls dazu dienen, uns daran zu erinnern, dass wir Bewusstseinsphänomene als begrifflich irreduzible Addition zur Ontologie akzeptieren müssen. Es kommt aber alles darauf an, wie dies geschieht. Der Identitätstheoretiker meint, dies könne in Form einer empirisch informativen Identitätsaussage geschehen. Dieser Auffassung werden wir uns im übernächsten Kapitel zuwenden. Vorerst aber zur Frage der Räumlichkeit des Bewusstseins, die den rein intuitiven Kern des Bewusstseinsrätsels ausmacht.



### 3 Das Bewusstsein im Raum

*Alles, was ist, ist irgendwann und irgendwo.*

– Crusius zugeschrieben

Im Folgenden geht es um drei Formen der Skepsis bezüglich der Räumlichkeit des Bewusstseins. Die erste, radikale, bestreitet, dass überhaupt räumliche Begriffe auf das Bewusstsein angewandt werden können. Die zweite gesteht zwar zu, dass Bewusstseine mit ihrem jeweiligen Subjekt, also der Person oder dem Tier, lokalisiert werden können, hält diese Lokalisierung aber für sekundär, für abgeleitet. Das Bewusstsein werde so nicht „an sich“ lokalisiert (McGinn 1997). Die dritte, moderates-te Form des Skeptizismus schließlich akzeptiert eine grobe Lokalisierung des Bewusstseins im Körper der Person oder im Gehirn zwar uneingeschränkt, glaubt aber nicht, dass sie in Form einer Lokalisierung spezifischerer Bewusstseinszustände und -prozesse in bestimmten Hirnarealen präzisiert werden kann.

Wie kann die These der Nichträumlichkeit des Bewusstseins plausibilisiert werden? Der Skeptiker wird auf seine Intuition verweisen, dass Aussagen der Art, dass sein Schmerz die Form und Größe eines Champignons hat und sich ungefähr in der Mitte seines Schädels befindet, einfach absurd und sinnlos sind. Eine Verteidigung dagegen kann der Verweis darauf sein, dass das Beispiel etwas krude und wahrscheinlich unrealistisch ist. Vielleicht werden alle Bewusstseinszustände diffus und sogar diskontinuierlich lokalisiert sein. Aber selbst für die Beispielsbehauptung bedeutet das nicht, dass sie sinnlos ist. Sie mag sich als falsch oder unpräzise herausstellen, aber falsch oder unpräzise können nur sinnvolle Aussagen sein. Sie werden korrigiert durch andere Aussagen, die aber im Wesentlichen die gleiche Begrifflichkeit verwenden.

Der Verteidiger zeigt sich also unbeeindruckt. Er teilt die Intuitionen des Skeptikers eben einfach nicht. Ihm ist nicht evident, dass die gemachten Aussagen sinnlos sind. Der Skeptiker versucht nun, die Sensibilität des Optimisten dafür zu wecken, dass den Aussagen eine bestimmte Art von Hintergrund fehlt, der bei vielen anderen Aussagen eine notwendige Bedingung für ihre Sinnhaftigkeit zu sein scheint:

„Stell Dir mein Gehirn in meinem Kopf vor. Es ist grau und furchig und hat eine bestimmte charakteristische Form. Aber nun versuche einmal, Dir meine visuellen Erfahrungen, die ich gerade von meiner Umgebung mache, in meinem Kopf vorzustellen. Du behauptest, sie könnten sich in einem bestimmten Areal des Gehirnes befinden, oder über mehrere verteilt sein, vielleicht sogar diskontinuierlich. Nehmen wir einmal, um nur irgendeine beliebige räumliche Vorstellung heranzuziehen, an, sie hätten die Form eines Blumenstraußes, seien aber, da ein gewöhnlicher Blumenstrauß natürlich nicht in meinen Kopf hineinpasst, um ein Vielfaches kleiner. Wie viel kleiner brauchen wir hier nicht festzulegen! Nehmen wir weiter an, dass ich mir gleichzeitig einen Blumenstrauß anschau. Wie willst Du Dir denn nun vorstellen, dass Deine visuellen Erfahrungen die Form eines Blumenstraußes haben und in Deinem Gehirn lokalisiert sind? Wenn ich mir ein kleines Bild des Blumenstraußes vorstelle, wer soll sich dieses Bild ansehen? Vielleicht ein kleiner Mann in meinem Gehirn? Und wie soll ich die Größe des gesehenen Blumenstraußes mit der angeblichen Größe der visuellen Erfahrung in meinem Kopf in eine vernünftige Beziehung setzen? Siehst Du nicht, dass man sich hier gar nichts Richtiges vorstellen kann, dass die Idee, die Erfahrung befinde sich in meinem Kopf, gar keinen Sinn macht?“

Hiermit sind wir beim intuitiven Kern unserer Schwierigkeiten mit der Räumlichkeit des Bewusstseins. Unsere Konzeptionen der Räumlichkeit gewöhnlicher Gegenstände basieren auf entsprechenden Wahrnehmungen der Gestalt und räumlichen Beziehungen dieser Gegenstände. Aber das Bewusstsein scheinen wir weder als räumliches Objekt wahrnehmen noch uns als solches vorstellen zu können. Dies ist auch der Grund, warum die entsprechenden Aussagen leicht als suspekt erscheinen können. Ihnen fehlt der Hintergrund räumlicher Erfahrungen und Vorstellungen, der zum Sinn anderer Arten von Aussagen über räumliche Verhältnisse zumindest beiträgt. Diese Behauptung darf aber nicht mit einer anderen verwechselt werden, die oft von Vertretern der Nichträumlichkeitsthese aufgestellt wird.

### **3.1 Zeigt das Bewusstsein sich als wesentlich nichträumlich?**

So begründet etwa Colin McGinn seine Skepsis, indem er sagt, das Bewusstsein „zeige sich als wesentlich nichträumlich“ (McGinn 97: 97). Aber auf welche Weise zeigt das Bewusstsein sich so? Man könnte vermuten, dass McGinn also meint, wir hätten eine *Anschauung* vom Bewusstsein, in der sich uns dessen Nichträumlichkeit zeigt. Uns wären also etwa Farbempfindungen in einer speziellen Form von Anschauung als wesentlich nichträumliche Entitäten gegeben. Eine solche Anschauung

scheint es aber nicht zu geben. Einen Bewusstseinszustand kann man weder wahrnehmen noch sich eine anschauliche Vorstellung davon machen. Wenn man versucht, sich eine anschauliche Vorstellung z.B. von einer Gelbempfindung zu machen, landet man doch immer nur bei der Vorstellung eines gelben äußeren Gegenstands. Und eine solche Vorstellung zeigt ihr Objekt natürlich als ein wesentlich räumliches. McGinn hebt selbst diese Punkte hervor (97f). Aber wenn wir keine besondere Anschauung von Bewusstseinsentitäten haben, ja wenn wir anschauliche Vorstellungen vielleicht überhaupt *nur* von räumlich lokalisierbaren Objekten haben können, fragt sich, wie die Behauptung zu begründen ist, dass sich das Bewusstsein als wesentlich nichträumlich zeigt. Bisher haben wir nur das Resultat, dass das Bewusstsein sich gar nicht „präsentiert“ oder „zeigt“, weder als räumlich noch als nichträumlich – jedenfalls, wenn wir den Begriff des Zeigens oder Präsentierens so verstehen, dass er Anschaulichkeit und/oder Wahrnehmbarkeit impliziert. Es scheint hier also eine Gefahr zu bestehen, von der These, dass sich das Bewusstsein in der Wahrnehmung oder Anschauung *nicht als räumlich zeigt*, umstandslos zu der ganz anderen These überzugehen, dass das Bewusstsein sich als wesentlich nichträumlich *zeigt*.

### **3.2 McGinns Begründung der Nichträumlichkeitsthese**

Diese These ist mehr als fragwürdig, weil wir ja normalerweise ganz selbstverständlich eine grobe Lokalisierung des Bewusstseins vornehmen. Wir nehmen normalerweise an, dass das Bewusstsein einer Person da ist, wo sich auch diese Person selbst befindet. Es lohnt sich, genauer zu betrachten, wie McGinn mit diesem Sachverhalt umgeht und welche weiteren Überlegungen er zur Stützung der Nichträumlichkeitsthese anführt:

...we take each subject of consciousness to be somewhere in the vicinity of a distinguished body, and we locate conscious events in the approximate neighbourhood of the physical object we call the brain. ... How do we actually make the locational judgement that we do? Not, clearly, by perceiving that conscious events occupy particular places; rather, by trading upon certain causal considerations. Events in particular physical objects are directly causally involved in changes of mental state; and my states of consciousness are situated in the vicinity of that brain whose activity is most directly implicated in the causal relations controlling my mental life. For example, my visual states are in the whereabouts of the eyes and brains that produce them, and not somewhere in (say) the Grand Canyon (unless my eyes and brain hap-

pen to be there). But this kind of causally based location of the mental should be taken for what it is. First, it is parasitic on a prior location of physical objects; there is no independent route on to mental location, since that is based solely on bearing causal relations to things that can be nonderivatively located. ... Locating mental events as we do in the actual world is merely ‚theoretical‘ – a sort of courtesy location. Considered in themselves, intrinsically, we do not regard mental events as having a location. The imprecision of our locational judgement here is a mark of this. Second, to allow that consciousness can be roughly located is not to grant it the full panoply of spatial predications. We still do not get predications of shape, size, dimensionality etc. (98f).

Diese reichhaltige Passage enthält eine Vielzahl von Punkten, die im Folgenden diskutiert werden sollen. Zusammen mit der ersten These, dass das Bewusstsein sich als nichträumlich präsentiert, die ich der Vollständigkeit halber noch einmal aufführe, unterscheide ich insgesamt 6 Thesen:

- 1) Das Bewusstsein kann nicht wahrgenommen werden, ‚präsentiert‘ sich aber als nichträumlich.
- 2) Wir lokalisieren das Subjekt der Erfahrung in der ‚Nähe‘ eines ‚ausgezeichneten‘ Körpers und die Bewusstseinsereignisse ‚ungefähr‘ in der ‚Nachbarschaft‘ des Gehirns.
- 3) Die alltägliche Lokalisierung des Bewusstseins beruht auf rein kausalen Kriterien. Die Erlebnisse eines Subjekts sind in der Nähe des Gehirns zu lokalisieren, das am unmittelbarsten in die kausalen Beziehungen involviert ist, die mein Bewusstseinsleben kontrollieren.
- 4) Die Lokalisierung mentaler Entitäten ist parasitär, bloß abgeleitet von der vorgängigen Lokalisierung physischer Entitäten. An sich schreiben wir Bewusstseinsentitäten keinen Ort zu. Ihre Lokalisierung ist sozusagen bloß eine Lokalisierung ehrenhalber.
- 5) Die fehlende Präzision der Lokalisierung ist ein Indiz für ihren abgeleiteten Charakter.

- 6) Wir können nicht die ganze Palette räumlicher Begriffe auf Bewusstseinsentitäten anwenden.

Eine wesentliche Annahme dieser Argumentation ist die, dass in der Wahrnehmung nur der Körper lokalisiert wird. Zusammen mit der zusätzlichen Annahme, dass eine Lokalisierung des Bewusstseins an sich nicht auch auf andere Weise als durch Wahrnehmung erfolgen kann, ergibt sich daraus die Konklusion 4), dass die Lokalisierung des Bewusstseins rein derivativ ist und bloß ‚ehrenhalber‘ erfolgt. Die Argumente 5) und 6) können als Zusatzargumente für diese Konklusion verstanden werden. Dass eine Lokalisierung des Bewusstseins an sich ausschließlich auf der Grundlage einer Wahrnehmung von Bewusstseinsentitäten erfolgen könnte, versucht McGinn streng genommen gar nicht zu zeigen. Er behauptet nur, dass die in 2) charakterisierte Lokalisierung, die nach McGinn unserer gewöhnlichen Auffassung entspricht, *de facto* auf rein kausalen Kriterien beruht (3)), und dass diese nicht ausreichend sind, um eine Lokalisierung des Bewusstseins an sich zu begründen. Dass in der Wahrnehmung nur Körper lokalisiert werden, ist auch keine Behauptung, die McGinn explizit aufstellt. Sie wird ihm, gegeben dass das Bewusstsein selbst nicht wahrgenommen wird, als selbstverständlich erscheinen. Ich werde aber dafür argumentieren, dass sie falsch ist und nur dann als selbstverständlich erscheint, wenn man einer gewissen allgemeinen Tendenz des Denkens unterliegt, die als begrifflicher Dualismus bezeichnet worden ist (Searle 1992).

### **3.3 Der intentionale Gehalt der basalen Selbstlokalisierung**

Ein erster wesentlicher Schritt zur Lösung des Problems besteht in einem angemessenen Verständnis des intentionalen Gehalts der Wahrnehmung, insbesondere der visuellen Wahrnehmung. Die visuelle Wahrnehmung ist durch ihre perspektivische Struktur selbstlokalisierend. Wir sehen die Dinge als in räumlichen Beziehungen zu uns selbst stehend, z.B. als vor uns, rechts oder links von uns, nah oder fern etc.. Die basale Selbstrepräsentation in der Wahrnehmung<sup>23</sup> ist aber nicht die Repräsentation eines rein mentalen Subjekts, eines transzendentalen Egos oder dergleichen, noch werden in einer solchen Repräsentation in irgendeiner Form die Bewusstseinszustän-

---

<sup>23</sup> Zu basalen, nichtbegrifflichen Formen der Selbstrepräsentation vgl. die Studie von José Luis Bermúdez (1998, besonders Kap. 7).

de des Subjekts repräsentiert. Man kann sich in der visuellen Wahrnehmung verorten, ohne über den Begriff eines Bewusstseinszustandes auch nur zu verfügen. Dies ist die Situation, in der sich vermutlich viele Tiere und kleine Kinder befinden: Sie verfügen über eine basale Form von Selbstbewusstsein, in dem sie sich selbst als ein sich durch den Raum bewegendes Wesen verstehen; aber sie verfügen nicht über den Bewusstseinsbegriff und damit auch nicht über den Begriff eines rein mentalen Subjekts der Erfahrung.

Bezieht sich das basale Wahrnehmungsselbstbewusstsein deshalb auf den Körper des Subjekts der Wahrnehmung? Dies ist nur dann zwingend, wenn wir uns von dem schon angesprochenen begrifflichen Dualismus bestimmen lassen, wenn wir also davon ausgehen, alles müsse entweder geistig oder körperlich in einem Sinn sein, in dem das Geistige das Körperliche ausschließt und umgekehrt. Aber tatsächlich kann, wenn die angenommene Konzeption des Wahrnehmungsselbstbewusstseins korrekt ist, dieses seinen Gegenstand gar nicht als einen ‚bloßen‘, also nicht-geistigen Körper herausgreifen, denn als einen solchen Körper kann nur jemand etwas repräsentieren, der auch über eine Repräsentation des Geistigen verfügt, und eine solche Repräsentation gehört ja nicht zum intentionalen Gehalt des Wahrnehmungsselbstbewusstseins. Deshalb entzieht sich die Selbstrepräsentation im Wahrnehmungsbewusstsein dem üblichen Dualismus von Geist und Körper oder Leib und Seele, ist ihm gegenüber vorgängig und grundlegender.

### **3.4 Basales und begriffliches Selbstbewusstsein**

Was aber geschieht, wenn zu der basalen und nichtbegrifflichen Selbstrepräsentation in der Wahrnehmung anspruchsvollere, begriffliche Formen der Selbstrepräsentation hinzukommen und die Unterscheidung zwischen Geistigem und Körperlichen? Was bedeutet dies für die Lokalisierung des Bewusstseins? Halten wir zuerst fest, dass auch das Bezugsobjekt des Wortes „Ich“, jedenfalls wenn man sein gewöhnliches Verständnis zu Grunde legt, kaum als ein reiner Geist oder reiner Körper verstanden werden kann. „Ich hatte diesen Gedanken zuerst in Gelsenkirchen“ bedeutet weder „Mein Bewusstsein hatte diesen Gedanken zuerst in Gelsenkirchen“ noch „Mein Körper hatte diesen Gedanken zuerst in Gelsenkirchen“. Auch „Ich“ bezeichnet etwas, eine Person, einen Menschen, dem sowohl geistige als auch körperliche Attribute zugeschrieben werden, das sich aber an sich dem Dualismus von Leib und Seele

entzieht. Die Person ist weder ein bloßer Körper, noch ein bloßer Geist; am ehesten könnte man sie noch als einen beseelten, bewussten Körper charakterisieren.

Wenn man nun einen Begriff vom Bewusstsein hat und so die Frage nach seinem Ort überhaupt aufwerfen kann, ist es zweifellos natürlich, sie durch eine Lokalisierung des Bewusstseins innerhalb der räumlichen Grenzen der Person zu beantworten. Das Selbstbewusstsein als Bewusstseinssubjekt baut auf den primitiveren Formen des Selbstbewusstseins auf und ist nur vor ihrem Hintergrund verständlich. Tatsächlich wird man dazu neigen, eine noch spezifischere Lokalisierung vorzunehmen. Der Gedanke, dass das Bewusstsein sich im Kopf befindet, ist ein sehr natürlicher. Dies liegt in der Perspektivität der Wahrnehmung, der rechts/links- und unten/oben-Orientierung begründet. Man könnte sagen, dass die Welt uns so erscheint, als ob sie vom Kopf aus gesehen würde. Der Kopf ist auch der Ort der mimischen Äußerungen des Gefühlszustands und der Ursprungsort der Sprachlaute. Wenn wir mit dem Anderen sprechen oder ihn trösten, wenn wir wissen wollen, wie ihm zumute ist, orientieren wir uns vornehmlich auf seinen Kopf hin. Und wir sagen ständig solche Dinge wie: „Ich lass mir das durch den Kopf gehen!“

Aus der von McGinn verteidigten, im weitesten Sinn Cartesianischen Perspektive heraus müssten wir dagegen in dem Moment, in dem wir Geist und Körper einer Person unterscheiden können, räumliche Eigenschaften nur noch dem Körper zuschreiben – jedenfalls sofern es um Zuschreibung in einem nicht abgeleiteten Sinn geht. Dies entspricht unserer gewöhnlichen Praxis aber nicht. Dass McGinn das auch in gewisser Weise anerkennt, zeigt sich darin, dass er sich bemüht, unserer gewöhnlichen Praxis zumindest durch sein Konzept einer bloß abgeleiteten Lokalisierung Sinn abzugewinnen. Ich möchte aber nun zeigen, dass ihm das nicht gelingt, weil die Vorstellung einer bloß abgeleiteten Lokalisierung keinen klaren Sinn hat. McGinns Position kollabiert in die des radikalen Skeptikers.

### **3.5 Kritik der Vorstellung einer bloß abgeleiteten Lokalisierung**

Wie ist die Aussage zu verstehen, dass Bewusstsein sei an sich, intrinsisch, gar nicht räumlich und lokalisiert, könne aber durch seine Beziehungen zum Körper bzw. zum Gehirn „ehrenhalber“ (*courtesy location*) lokalisiert werden? Halten wir zuerst fest, dass an der Idee, ein Objekt könne „nur“ durch seine Beziehungen zu anderen Objekten räumlich verortet werden, eigentlich nur das „nur“ merkwürdig ist, da die Lokali-

sierung eines Objekts immer im Aufweis von Beziehungen zu anderen Objekten besteht. (Dies gilt unabhängig davon, ob der Raum selbst nur in räumlichen Beziehungen besteht oder noch mehr ist.) Tatsächlich merkwürdig ist aber die Vorstellung, ein an sich nichträumliches Objekt könne überhaupt durch seine kausalen Beziehungen zu anderen Objekten lokalisiert werden und sei es nur „ehrenhalber“. Dies gilt besonders angesichts einer Behauptung wie der in 2), dass die Bewusstseinsentitäten sich in der *Nähe* oder *Nachbarschaft* des Körpers oder des Gehirns befinden. Wie kann man von etwas, das an sich gar nicht lokalisiert ist, sagen, dass es sich in der Nähe von etwas befindet? Was sollte dann den Sachverhalt ausmachen, dass es einem anderen Gegenstand nah ist und nicht fern? Aussagen dieser Art scheinen letztlich sinnlos zu sein, und das gilt dann natürlich für jede Art von Ortsbestimmung. So ist der Versuch, mit dem Konzept der bloß abgeleiteten Lokalisierung ‚ehrenhalber‘ eine Art Mittelposition – die eingangs als die zweite Form des Räumlichkeitsskeptizismus bezeichnete Position – zwischen einer unzweideutigen Lokalisierung des Bewusstseins und der These seiner völligen Nichträumlichkeit zu beziehen, gescheitert. Wenn das Bewusstsein überhaupt lokalisiert werden kann, kann es auch ‚an sich‘ lokalisiert werden. Sonst wird nur das mit dem Bewusstsein verbundene Gehirn, oder der mit dem Bewusstsein verbundene Körper, lokalisiert. Dies ist wohl auch die Position, auf die McGinn sich letztlich zurückziehen würde. Demnach müssten wir aber statt „Ich habe diese Sache damals in Gaienhofen durchdacht“ streng genommen so etwas sagen wie „Ich habe diese Sache damals durchdacht, als mein Körper in Gaienhofen war.“ Dies ist aber, gelinde gesagt, merkwürdig. McGinn macht nicht von ungefähr den Versuch, unserer gewöhnlichen Lokalisierungspraxis Sinn abzugewinnen. Dass unser Denken an sich völlig ortslos sein soll, ist mit unserem gewöhnlichen Selbstverständnis nicht zu vereinbaren. Schließlich ist unter dieser Voraussetzung auch der Zusammenhang zwischen dem Bewusstsein und dem Körper eines Subjekts unverständlich. Was könnte den wesentlich lokalisierten Körper zum Körper eines (an sich) nirgendwo seienden Subjekts oder Bewusstseins machen? McGinn behauptet, dass dieser Zusammenhang auf kausalen Beziehungen beruht. Aber wie soll es kausale Beziehungen zwischen einer räumlichen und einer radikal-nichträumlichen Entität geben? Nach dem gewöhnlichen Verständnis von Kausalbeziehungen verbinden diese räumlich verschiedene Entitäten. McGinn kann also unserer gewöhnlichen Praxis der Lokalisierung von Bewusstseinszuständen mit ihrem Träger gar keinen Sinn abgewinnen. Aber dies scheint seine Position letztlich unver-



ständiglich zu machen, da diese Lokalisierung ein wesentlicher Aspekt unseres Selbstverständnisses ist.

### **3.6 Kritik des moderaten Skeptizismus**

Wenn man aber den radikalen Skeptizismus verabschiedet, gibt es auch keinen Grund auf halbem Wege innezuhalten und bei der dritten, moderateren Form des Skeptizismus stehen zu bleiben, der ich mich nun zuwenden möchte. Der moderate Skeptizismus läuft darauf hinaus, bestimmte basale Methoden der Ortsbestimmung von Bewusstseinszuständen – jene, die sich auf die bloße Lokalisierung der Person beziehen – zu akzeptieren, die Erweiterung der Methoden der Ortsbestimmung durch die Hirnforschung aber abzulehnen. Damit sind wir bei den letzten beiden von McGinns Einwänden – dass die mangelnde Präzision der Lokalisierung ein Indiz für ihren abgeleiteten Charakter ist und dass wir nicht die ganze Palette räumlicher Begriffe auf Bewusstseinsentitäten anwenden können, dass wir also etwa nicht ohne weiteres vom Umfang, der Länge etc. von Bewusstseinsentitäten reden können. Die These vom abgeleiteten Charakter der Lokalisierung von Bewusstseinszuständen hatte ich zurückgewiesen, weil die Intuition, dass Ortsbestimmungen an sich nur auf den bloßen Körper zu beziehen sind, nicht zwingend zu machen ist, und weil sich zeigte, dass es letztlich keinen Sinn macht zu sagen, eine Entität sei zwar lokalisiert, aber nicht an sich. Die Präzision der Ortsbestimmungen mentaler Phänomene wiederum lässt sich steigern, wenn die Instrumentarien der Hirnforschung zu diesem Zweck akzeptiert werden. Zweifellos gibt es ja schon eine Menge Fortschritte in dieser Richtung.

Ein Argument gegen den moderateren Skeptizismus könnte man wie folgt formulieren: „Auf das Bewusstsein können wir uns nur im begrifflichen Rahmen unserer Alltagspsychologie und deren Erweiterungen in der philosophischen und wissenschaftlichen Psychologie beziehen. Mit diesen begrifflichen Mitteln können wir eine genauere räumliche Eingrenzung des Bewusstseins aber nicht begründen. Genauere Bestimmungen ließen sich nur auf der physiologischen Ebene vornehmen, aber wenn wir auf dieser begrifflichen Ebene operieren, können wir auch nur Gliazellen, Neuronen, Synapsen usw. lokalisieren, aber eben nicht das Bewusstsein und seine Zustände. Um dieses lokalisieren zu können, müssten wir eine Übereinstimmung zwischen unabhängig gewonnenen Ortbestimmungen des Bewusstseins und den

Ortsbestimmungen auf der physiologischen Ebene feststellen können. Aber eine genauere Ortsbestimmung auf der begrifflichen Ebene des Bewusstseins ist ja gerade nicht möglich.“ Bezeichnen wir dies als das Argument von der Notwendigkeit einer unabhängigen Lokalisierung, oder allgemeiner, da die gleiche Art der Argumentation auch bezüglich anderer Charakteristika anwendbar ist, als das Argument von der Notwendigkeit einer unabhängigen Bestimmung. Ich möchte nun zeigen, dass diese Art von Argumentation zu unakzeptablen Konsequenzen führt. Wenn wir ihr folgten, müssten wir die nähere Bestimmung altbekannter Eigenschaften durch neue Kriterien, Messverfahren und Methoden auch in zahllosen anderen Fällen zurückweisen, in denen wir sie normalerweise als unproblematisch akzeptieren. Betrachten wir etwa den Fall, dass ein neues Verfahren zur Messung der räumlichen Dimensionen von Körpern entdeckt wird, das viel genauere Bestimmungen ermöglicht als die herkömmlichen. Nehmen wir nun an, jemand würde argumentieren, dass man, um das neue Vorgehen als ein Verfahren etwa zur Messung der Länge eines Stabs akzeptieren zu können, dessen Resultate auch auf unabhängige Weise, d.h. allein durch Anwendung der bisher verwendeten Verfahren zur Längenmessung, erzielen und damit bestätigen können müsste. Es ist jedoch offenkundig, dass dies niemals möglich ist, wenn die neuen Verfahren präziser sind, oder auch einen weiteren Anwendungsbereich haben als die konventionellen, also Messungen in Fällen erlauben, wo sie bisher noch gar nicht möglich waren. Das Resultat, dass es nie möglich wäre, Eigenschaften durch neue Methoden präziser und umfassender zu bestimmen, ist aber sicherlich inakzeptabel. Ich komme deshalb zu dem Schluss, dass das Argument der Notwendigkeit unabhängiger Bestimmung zurückzuweisen ist, da es zu einem inakzeptablen begrifflichen Konservatismus führt, der es uns unmöglich macht, neue Methoden und Messverfahren zur Bestimmung bereits bekannter Eigenschaften zu akzeptieren.

All dies soll aber nicht heißen, dass es nicht auch Grenzen möglicher Präzision bei der Ortsbestimmung von Bewusstseinsentitäten gibt. Solche Grenzen treten bekanntlich auch bei der Integration anderer Makroentitäten auf. Ob beispielsweise das eine oder andere Elektron wirklich zum Tisch gehört oder zu seiner Umgebung, ist letztlich unbestimmt und wäre nur durch eine Stipulation festzulegen. Die Kriterien, mit denen wir normalerweise bestimmen, was ein Tisch ist und wo seine räumlichen Grenzen verlaufen, reichen einfach nicht so tief auf die Mikroebene hinab, um eine solche Frage entscheiden zu können, und es ist auch nicht klar, wie sie entspre-

chend erweitert werden sollten. Es ist aber auch nicht nötig, hier eine Entscheidung herbeizuführen. Eine genauere Grenzziehung hat keinen praktischen Zweck, und es gibt auch keinen metaphysischen Zwang, warum sie möglich sein müsste. Eine gewisse Vagheit unserer Begriffe, auch im Hinblick auf räumliche Grenzen, können und müssen wir wohl akzeptieren. Dies ist kein vernünftiger Grund um etwa zu behaupten, es mache gar keinen Sinn zu sagen, dass der Tisch aus Mikroentitäten wie Elektronen besteht. Ich glaube, wir sollten uns die gleiche pragmatische Haltung auch in Bezug auf die Bestimmung der räumlichen Grenzen des Bewusstseins zu eigen machen. Man wird sicherlich an Punkte kommen, wo die räumlichen Grenzen des Bewusstseins sich nicht mehr genauer bestimmen lassen, wo die Begrifflichkeit, die Kriterien, die zur Bewusstseinssebene selbst gehören, oder noch eine nachvollziehbare Verbindung zu derselben haben, nicht mehr ausreichen, um entsprechende Entscheidungen zu begründen. Aber auch hier gilt, dass es auch keine Notwendigkeit gibt, warum solche Entscheidungen möglich sein sollten, und dass sie insbesondere nicht nötig sind, um der Aussage, Bewusstseinsentitäten hätten eine neuronale Mikrostruktur, Sinn zu verleihen.

Schließlich wird es durch die Erweiterung der Ortsbestimmungen durch neurowissenschaftliche Methoden auch möglich, die ganze Palette räumlicher Begriffe auf Bewusstseinsentitäten anzuwenden, also etwa von der Größe oder Gestalt von Bewusstseinsentitäten zu sprechen. Ohnehin scheint mir dies nicht so unintuitiv zu sein, wie es vielleicht auf den ersten Blick scheinen mag. Z.B. scheint es mir durchaus sehr plausibel zu sein, dass, wenn ich mich einem Gegenstand nähere, auch der den Gegenstand repräsentierende Teil meines visuellen Bewusstseins buchstäblich größer wird. Ich wäre sehr überrascht, wenn sich dies als falsch herausstellen würde. Allgemein ist es intuitiv plausibel, dass insbesondere das Wahrnehmungsbewusstsein topologische Formen der Repräsentation benutzt, und dies bedeutet ja, dass äußere Gestalten und räumliche Verhältnisse wiederum durch innere Gestalten und räumliche Verhältnisse repräsentiert werden. Für die Existenz topologischer Repräsentationen gibt es wiederum bereits Evidenzen auf der neuronalen Ebene. Man hat etwa das Gehirn einer Katze, die ein Kreuz wahrgenommen hatte, so präpariert, dass tatsächlich eine kreuzähnliche Struktur im visuellen Areal sichtbar wird (Damasio 1997: 149).

Der Gedanke, dass Bewusstseinsentitäten eine Gestalt und Ausdehnung haben, hat zugegebenermaßen zuerst etwas Irritierendes. Bei näherem Hinsehen spre-

chen aber nicht nur starke theoretische Argumente für ihn, sondern auch gewisse Intuitionen.

## 4 Das Bewusstsein der Identität

*Von zwei Dingen zu sagen, sie seien identisch, ist ein Unsinn, und von Einem zu sagen, es sei identisch mit sich selbst, sagt gar nichts.*

– Ludwig Wittgenstein, Tractatus-logico-philosophicus

*Though the notion of identity is so simple, confusion over it is not uncommon.*

– Willard Van Orman Quine, Word & Object

Die Auseinandersetzung mit der Identitätstheorie ist von zentraler Bedeutung für die Argumentation dieser Arbeit. Sie setzt voraus, dass das, was der ontologische Physikalist für empirisch informative Identitätsaussagen *a posteriori* hält, durch die z.B. das wirkliche Wesen von Farbe, Wärme oder anderen manifesten Phänomenen entdeckt wird, in Wirklichkeit verkappte Eliminationen sind. Das Argument gegen die psychophysische Identitätstheorie ist dabei ein allgemeines Argument gegen empirisch informative Identitätsaussagen überhaupt. Deshalb ist es sinnvoll, sich erst einmal unabhängig von der Debatte über psychophysische Identitätsaussagen und andere so genannte wissenschaftliche Identifikationen über die Logik von Identitätsaussagen klar zu werden. Dies gilt umso mehr, als das Paradox informativer Identitätsaussagen seit Gottlob Frege, nach dem es oft auch als Freges Problem benannt wird, es in die Debatte einführte, eines der wichtigsten und schwierigsten Probleme der Sprachphilosophie gewesen ist.

### 4.1 Das Paradox der Identität

Das Paradox der Identität ist dem Paradox der Existenz eng verwandt (Williams 1989). Das Paradox der Existenz ergibt sich aus der Annahme, dass Existenzaussagen Entitäten die Eigenschaft der Existenz zuschreiben. Da alle Entitäten existieren, müssten nach dieser Deutung alle positiven Existenzaussagen *a priori* und notwendig wahr sein. Negative Existenzaussagen dagegen müssten selbstwidersprüchlich und damit *a priori* und notwendig falsch sein: Man würde zuerst auf einen Gegenstand Bezug nehmen, um ihm dann die Existenz abzusprechen. Diese Folgen sind aber paradox, denn weder scheinen positive Existenzaussagen stets *a priori* wahr noch negative Existenzaussagen stets *a priori* falsch zu sein. Z.B. ist es nicht *a priori*

wahr, dass es bei mir um die Ecke einen Gemüseladen gibt und nicht *a priori* falsch, dass es dort keinen Juwelier gibt.

Das Paradox der Identität ergibt sich in ähnlicher Weise aus der Annahme, dass Identitätsaussagen das Bestehen einer Identitätsrelation behaupten, einer Beziehung der Selbstidentität, in der jeder Gegenstand notwendig mit sich selbst steht. Wenn es eine solche Beziehung gibt, so wissen wir sicherlich *a priori*, dass jeder Gegenstand notwendig in dieser Beziehung zu sich selbst steht und zwei Gegenstände niemals in dieser Beziehung zueinander stehen können. Identitätsaussagen wären also auch entweder *a priori* und notwendig wahr oder *a priori* und notwendig falsch. Aber dass Mark Twain Samuel Clemens ist oder der Morgenstern der Abendstern, scheinen wir nicht *a priori* zu wissen und es scheint auch nicht notwendig wahr zu sein. Und dass der SPD-Vorsitzende mit dem Bundeskanzler identisch ist, ist falsch, aber dies wissen wir nicht *a priori* und es scheint auch nicht notwendig so zu sein. Wir können uns sicherlich vorstellen, dass die Dinge sich so entwickelt hätten, dass es wahr geworden wäre. Wie sind also Identitätsaussagen *a posteriori* möglich, wenn Identitätsaussagen sich auf eine Relation der Selbstidentität beziehen sollen, von der wir doch *a priori* wissen, dass jeder Gegenstand notwendig in dieser Beziehung steht? Dies ist das Paradox der Identität.

Es ist wichtig zu sehen, dass sich dieses Problem nicht nur dadurch ergibt, dass wir *a priori* zwar wissen können, dass ein Identitätssatz der Form „ $a = a$ “ wahr ist, aber nicht, dass ein Identitätssatz der Form „ $a = b$ “, in dem also zwei verschiedene Eigennamen verwendet werden, wahr ist. Zwar gibt es tatsächlich einen wichtigen Unterschied zwischen diesen Typen von Sätzen. Aufgrund unser Kenntnis der Konventionen unserer Sprache werden wir normalerweise selbst dann erwarten, dass ein Satz der Form „ $a = a$ “ wahr ist, wenn wir die Bedeutung des Ausdrucks „ $a$ “ gar nicht kennen, da zwei Vorkommnisse desselben Zeichens in einem Satz sich im Allgemeinen auf dasselbe beziehen werden. Dies ist aber erstens ein kontingentes Merkmal unserer Sprache: Wir könnten uns theoretisch etwa eine Sprache vorstellen, in der Namen beim zweiten Gebrauch in einem Satz ihre Bedeutung ändern. Zweitens gilt es nicht einmal ausnahmslos, denn natürlich wird mitunter ein Satz wie „Peter traf Peter“ geäußert, in dem die Namen Verschiedenes bezeichnen. Der wichtigste Punkt ist aber ein dritter. Selbst wenn syntaktisch identische Namen sich stets auf dasselbe bezögen, selbst wenn wir uns eine disambiguierte Sprache vorstellen, in der einem (syntaktisch individuierten) Namen nie mehr als ein Gegenstand zugeordnet ist, so

würde die bloße Tatsache, dass es auch Identitätssätze der Form „ $a = b$ “ gibt, die empirische Informativität mancher Identitätssätze nur dann vollständig erklären können, wenn wir Identitätssätze generell als metalinguistische Aussagen verstehen, also so, dass mit ihnen so etwas gesagt wird wie, dass zwei Zeichen sich auf denselben Gegenstand beziehen. Würde man diese Deutung von Identitätsaussagen akzeptieren, könnte man sagen, dass sich die Wahrheit von Identitätsaussagen der Form „ $a = a$ “ aus den allgemeinen Regeln einer Sprache ableiten lässt, in der jedem Namen nur ein Gegenstand zugeordnet ist, und damit in diesem Sinne *a priori* ist, während die Wahrheit von Aussagen der Form „ $a = b$ “ sich nicht aus allgemeinen Prinzipien dieser Sprache ableiten lässt, so dass es sich daher dabei um Wahrheiten *a posteriori* über diese Sprache, über die spezifischen Bedeutungen einzelner Ausdrücke handelt.

Die metalinguistische Deutung von Identitätsaussagen ist aber nicht unbedingt plausibel, da es, jedenfalls auf den ersten Blick, nicht so scheint, dass es sich bei den üblichen Beispielen für Identitätsaussagen wie der Identifikation von Morgenstern und Abendstern, oder bei den so genannten wissenschaftlichen Identifikationen wie z.B. der von Wasser und  $H_2O$  um metalinguistische Aussagen handelt. So ist denn die metalinguistische Deutung in der philosophischen Tradition auch nur selten vertreten worden (z.B. Place 1999). Im gegenwärtigen Kontext ist sie für uns nicht interessant, weil sie gar nicht den Versuch unternimmt, das Paradox der Identität zu lösen. Denn nach der metalinguistischen Deutung wird die Identitätsaussage ja nicht durch das Bestehen einer Relation wahrgemacht, in der jeder Gegenstand notwendig mit sich selbst steht, sondern dadurch, dass zwei verschiedene Ausdrücke den selben Gegenstand bezeichnen, und dass dies der Fall ist, ist sicherlich nicht notwendig.

Bei einem Satz der Form „ $a = a$ “ sind also zwei voneinander unabhängige Formen von Wissen *a priori* im Spiel, die klar voneinander getrennt werden müssen. Eine in der beschriebenen Weise disambiguierte Sprache vorausgesetzt, können wir *a priori* sagen, dass die beiden Vorkommnisse von „ $a$ “ dasselbe bezeichnen. Gegeben dass sie dasselbe bezeichnen, können wir wiederum *a priori* wissen, dass „ $a = a$ “ wahr ist, da wir *a priori* wissen, dass jeder Gegenstand mit sich selbst identisch ist. Schließlich kann Wissen *a priori* hier noch auf eine dritte Weise ins Spiel kommen. Nach einer gängigen Konzeption ist Bedeutungswissen Wissen *a priori*. Unsere Erfahrung wird durch bestimmte sprachliche Kategorien strukturiert, und das Wissen, das für die Bedeutung dieser Kategorien konstitutiv ist, ist Wissen *a priori*. Nun scheint es plausibel zu sagen, dass einen Eigennamen zu verstehen zu wissen heißt,

welchen Gegenstand er bezeichnet. Aber daraus scheint wiederum zu folgen, dass, wer die Namen „a“ und „b“ versteht, auch allein aufgrund seines Bedeutungswissens und damit in diesem Sinne *a priori* wissen muss, dass „a = b“. Dies folgt zwar nicht aus dem allgemeinen Bedeutungswissen wie die Wahrheit von „a = a“, aber aus dem spezifischen Wissen über die Bedeutung der Ausdrücke „a“ und „b“.

Nun kann man natürlich versuchen, diese Argumentation als zu naiv zu kritisieren. Sowohl die allgemeine Unterscheidung zwischen Bedeutungs- und Weltwissen als auch diese bestimmte Anwendung dieser Unterscheidung kann sicherlich in Frage gestellt werden. Ist es tatsächlich wahr, dass, wer die Bedeutung zweier Eigennamen versteht gegebenenfalls auch *ipso facto* wissen muss, dass sie denselben Gegenstand bezeichnen? Man könnte meinen, dass Beispiele wie das von „Samuel Clemens“ und seinem Pseudonym „Mark Twain“ und vom „Morgenstern“ und „Abendstern“ oder von „Hesperus“ und „Phosphorus“, wenn man letztere Bezeichnungen, da sie nicht ‚beschreibend‘ sind – außer für Menschen, die des Griechischen mächtig sind –, als bessere Beispiele für Eigennamen vorzieht, diese Vorstellung bereits widerlegen. Zeigen sie nicht, dass man zwei Ausdrücke verstehen kann, ohne zu wissen, dass sie denselben Gegenstand bezeichnen? Wir werden später diese Frage noch eingehend diskutieren. Fürs erste zeigen die angestellten Überlegungen aber zumindest, wie schwer die Aufgabe ist zu erklären, wie Identitätsaussagen Aussagen *a posteriori* sein können, insbesondere dann, wenn man die Bedeutung von sprachlichen Ausdrücken allein in dem Bezug sieht, den sie zu dem Gegenstand herstellen, den sie bezeichnen. Es ist deshalb nicht verwunderlich, dass der wohl immer noch einflussreichste Versuch, das Paradox der Identität zu lösen, sprachlichen Ausdrücken nicht nur einen Bezug zuschreibt, sondern auch einen von diesem Bezug verschiedenen Sinn.

## 4.2 Frege: Sinn vs. Bedeutung

Dieser Vorschlag wurde zuerst von Gottlob Frege vorgebracht. Frege führt den Sinnbegriff wie folgt ein:

Wenn wir nun in der Gleichheit eine Beziehung zwischen dem sehen wollten, was die Namen „a“ und „b“ bedeuten, so schiene  $a = b$  von  $a = a$  nicht verschieden sein zu können, falls nämlich  $a = b$  wahr ist. Es wäre hiermit eine Beziehung eines Dinges zu sich



selbst ausgedrückt, und zwar eine solche, in der jedes Ding mit sich selbst, aber kein Ding mit einem anderen steht. Was man mit  $a = b$  sagen will, scheint zu sein, dass die Zeichen oder Namen „ $a$ “ und „ $b$ “ dasselbe bedeuten, und dann wäre eben von jenen Zeichen die Rede; es würde eine Beziehung zwischen ihnen behauptet. ... Damit würde dann ein Satz  $a = b$  nicht mehr die Sache selbst, sondern nur noch unsere Bezeichnungsweise betreffen; wir würden keine eigentliche Erkenntnis darin ausdrücken. Das wollen wir aber doch gerade in vielen Fällen. Wenn sich das Zeichen „ $a$ “ von dem Zeichen „ $b$ “ nur als Gegenstand (hier durch die Gestalt) unterscheidet, nicht als Zeichen; das soll heißen: nicht in der Weise, wie es etwas bezeichnet: so würde der Erkenntnisgewinn von  $a = a$  wesentlich gleich dem von  $a = b$  sein, falls  $a = b$  wahr ist. Eine Verschiedenheit kann nur dadurch zustande kommen, dass der Unterschied des Zeichens einem Unterschiede in der Art des Gegebenseins des Bezeichneten entspricht. Es seien  $a, b, c$  die Geraden, welche die Ecken eines Dreiecks mit den Mitten der Gegenseiten verbinden. Der Schnittpunkt von  $a$  und  $b$  ist dann derselbe wie der Schnittpunkt von  $b$  und  $c$ . Wir haben also verschiedene Bezeichnungen für denselben Punkt, und diese Namen („Schnittpunkt von  $a$  und  $b$ “, „Schnittpunkt von  $b$  und  $c$ “) deuten zugleich auf die Art des Gegebenseins, und daher ist in dem Satze eine wirkliche Erkenntnis enthalten. Es liegt nun nahe, mit einem Zeichen (Namen, Wortverbindung, Schriftzeichen) außer dem Bezeichneten, was die Bedeutung des Zeichens heißen möge, noch das verbunden zu denken, was ich den Sinn des Zeichens nennen möchte, worin die Art des Gegebenseins enthalten ist. Es würde danach in unserem Beispiel zwar die Bedeutung der Ausdrücke „der Schnittpunkt von  $a$  und  $b$ “ und „der Schnittpunkt von  $b$  und  $c$ “ dieselbe sein, aber nicht ihr Sinn. Es würde die Bedeutung von „Abendstern“ und „Morgenstern“ dieselbe sein, aber nicht der Sinn. (Frege 1986: 40ff.)

Frege formuliert hier das Paradox der Identität auf prägnante Weise: Wenn sich „ $a$ “ und „ $b$ “ nur rein syntaktisch unterscheiden und wir „ $a = b$ “ nicht metalinguistisch verstehen, so kann sich der „Erkenntnisgewinn“ – wie Frege es ausdrückt – durch die Bejahung von „ $a = b$ “ auch nicht vom Erkenntnisgewinn durch die Bejahung von „ $a = a$ “ unterscheiden. In einer moderneren Terminologie können wir Freges Punkt wie folgt reformulieren: Der Erkenntnisgewinn durch die Akzeptanz oder Annahme zweier Aussagen (oder Überzeugungen etc.) kann sich nur dann unterscheiden, wenn sich der propositionale oder repräsentationale Inhalt dieser Aussagen unterscheidet. Sonst handelt es sich gar nicht um zwei verschiedene Aussagen, sondern nur um zwei verschiedene Sätze, mit denen ein und dieselbe Aussage gemacht wird. Ihr Verhältnis wären dann dem des deutschen Satzes „Der Morgenstern ist der Abendstern“ zum englischen Satz „The morning star is the evening star“ vergleichbar. Die bloße Verschiedenheit der sprachlichen Ausdrücke reicht also nicht aus, um einen

semantischen Unterschied zu begründen. Ein solcher Unterschied besteht aber, wenn die Signifikanz eines sprachlichen Zeichens nicht nur in dem bezeichneten Objekt, also dem, was Frege als die Bedeutung bezeichnet, liegt, sondern auch in der Art und Weise, in der uns diese Objekte ‚gegeben‘ sind oder wir sie herausgreifen oder identifizieren. Ein und derselbe Punkt etwa kann als der Schnittpunkt verschiedener Geraden herausgegriffen werden. Ein und derselbe Stern kann uns dadurch gegeben sein, dass er morgens oder abends am Himmel erscheint. Um noch ein weiteres, alterehrwürdiges Beispiel zu nennen: Dieselbe Klasse von Tieren kann als die Klasse von Tieren mit Herz oder die von Tieren mit Nieren identifiziert werden.

Was aber ist der Sinn eines Ausdrucks? Wie wird uns nun ein Gegenstand gegeben, wie greifen wir auf ihn zu oder identifizieren ihn? Die Antwort darauf scheint einfach zu sein. Wir greifen ihn durch eine seiner Eigenschaften heraus oder durch eine Beziehung, in der er steht. Derselbe Gegenstand oder dieselbe Klasse von Gegenständen kann durch verschiedene Eigenschaften oder verschiedene Beziehungen eindeutig herausgegriffen oder identifiziert werden. Alle Tiere, die Herzen haben, haben auch Nieren. Die Eigenschaft, ein Herz zu haben und die Eigenschaft, eine Niere zu haben, sind sozusagen verschiedene Pfade, die zu derselben Klasse von Tieren führen. Die Eigenschaften, vor Sonnenaufgang im Osten zu erscheinen und nach Sonnenuntergang im Westen, sind zwei verschiedene Pfade, die zu ein und demselben Himmelskörper führen.

Damit scheint es aber, dass eine Verschiedenheit des Sinns wesentlich mit einer Verschiedenheit objektiver Gegebenheiten, objektiver Eigenschaften oder Beziehungen verknüpft ist bzw. sogar in dieser besteht. Ein Herz zu haben oder eine Niere zu haben sind genauso objektiv verschiedene Eigenschaften wie die Eigenschaften, morgens oder abends an bestimmten Stellen am Himmel zu erscheinen. Unterschiede des Sinns im Sinne Freges sind also keine bloß subjektiven Differenzen. Sie sind mit objektiven Unterschieden essenziell verbunden oder bestehen sogar in diesen, falls wir die Sinne, wie dies heutzutage oft geschieht, gar mit den jeweiligen Eigenschaften identifizieren wollen. Jedenfalls gilt dies für die Beispiele, die Frege selbst gibt und für die Beispiele, die sonst standardmäßig für die Unterscheidung zwischen Sinn und Bedeutung oder für eng verwandte Unterscheidungen wie die zwischen Intension und Extension angeführt werden.

Dieser Punkt ist sehr wichtig, weil, wie schon angedeutet wurde und wie wir später noch ausführlich sehen werden, die Verteidiger der psychophysischen Identi-

tätstheorie dazu neigen, den Unterschied zwischen verschiedenen Sinnen in einen rein subjektiven Unterschied umzudeuten. Dieser Zug ist durch die Standardbeispiele nicht gedeckt, und es ist auch schon zu erkennen, worin seine grundsätzliche Schwierigkeit besteht: Wie soll eine bloß subjektive, nicht mit objektiven Gegebenheiten verknüpfte Differenz einen *kognitiven* Unterschied zwischen „a = a“ und „a = b“ erklären können? Einen kognitiven Unterschied müssen wir aber unterstellen, wenn die Akzeptanz von „a = b“ gegenüber der von „a = a“ tatsächlich einen Erkenntnisgewinn bedeuten soll.

Unterschiede des Sinnes sind aber nicht nur wesentlich mit objektiven Differenzen verknüpft. Wir können noch einen Schritt weitergehen: Sie lassen sich vollständig durch Unterschiede des Bezugs oder, in Fregescher Terminologie, der Bedeutung, erklären. So lässt sich die Verschiedenheit des Sinns von Ausdrücken wie „Der Schnittpunkt von a und b“ und „Der Schnittpunkt von b und c“ oder „Tiere mit Herz“ und „Tiere mit Nieren“ anscheinend durch die Verschiedenheit des Bezugs von Teilen dieser Ausdrücke erklären. „A“, „b“ und „c“ beziehen sich genauso auf verschiedene Objekte wie „Herz“ und „Nieren“. Was aber ist mit dem Sinn nicht-komplexer Ausdrücke, von Eigennamen wie z.B. „Köln“ oder „Bill Clinton“?

Frege war der Meinung, dass auch der Sinn dieser Zeichen sich durch einen komplexen Ausdruck angeben lässt. So könnte der Sinn des Namens „Aristoteles“ etwa durch die Beschreibung „der Schüler Platos und Lehrer Alexander des Großen“ angegeben werden. Allerdings glaubte Frege, dass der Sinn solcher Namen in der Alltagssprache auf eine Weise schwankt, die „in einer vollkommenen Sprache nicht vorkommen“ (op. cit.: 42 Fn.) dürfte. So könnte etwa ein anderer Sprecher mit „Aristoteles“ einen Sinn verbinden, der durch eine Beschreibung „der in Stagira geborene Lehrer Alexander des Großen“ angegeben werden kann. Offensichtlich lässt sich die Verschiedenheit dieser beiden Sinne wieder auf die Verschiedenheit der Bedeutungen von Teilen der Beschreibungen zurückführen.

Frege fasste die Kategorie der Eigennamen so, dass darunter nicht nur primitive Bezeichnungen wie „Köln“ und „Aristoteles“ fallen, sondern auch komplexe Ausdrücke wie eben „der Schnittpunkt der Geraden“, „der Schüler Platos“, „der Bundeskanzler“ usw., die man seit Russell meist als definite Beschreibungen bezeichnet. Es ist aber nach dem Gesagten einsichtig, warum er das tat, nämlich weil er glaubte, dass auch der Sinn von Eigennamen mit dem von Beschreibungen, also von komplexen Ausdrücke äquivalent ist. Das heißt aber, dass nach Freges Auffassung

auch die scheinbar primitiven Eigennamen semantisch gesehen komplex sind, und dies wiederum bedeutet, dass sich auch die Verschiedenheit des Sinns von Eigennamen in der angegebenen Weise auf die Verschiedenheit der Bezugnahme von Teilen komplexer Ausdrücke zurückführen lässt. Umgekehrt folgt daraus weiter, dass sich nicht auf Freges Konzeption des Sinnes berufen kann, wer zwei semantisch primitiven Ausdrücken verschiedenen Sinn bei gleichem Bezug zuschreiben will. Der Fregesche Sinnbegriff gibt uns kein Modell dafür, wie dergleichen zu denken wäre. Auch dieser Punkt wird für die Auseinandersetzung mit der psychophysischen Identitätstheorie von großer Wichtigkeit sein.

Die Unterscheidung zwischen der Bezugnahme eines komplexen Ausdrucks und der Bezugnahme seiner Teile, derer ich mich bedient habe, bedarf aber noch der Erläuterung. Philosophische Diskussionen der Bezugnahme sind meist sehr stark an der singulären Bezugnahme orientiert. Man kann sogar sagen, dass viele Philosophen Referenz und singuläre Bezugnahme schlichtweg identifizieren. Singuläre Bezugnahme ist aber Bezugnahme auf ein Einzelding als logisches Subjekt. Man *erwähnt* eine Entität, der dann im Folgenden Eigenschaften zugeschrieben werden. (Ich übernehme hier und im Folgenden die Praxis, den Begriff der Eigenschaften so weit zu fassen, dass auch Relationen darunter fallen.) Und in diesem Sinn von singulärer Bezugnahme bezieht sich ein Ausdruck wie „Der Schüler von Plato und Lehrer Alexanders“ tatsächlich nur auf einen bestimmten Schüler und nicht z.B. auf Plato oder Alexander. Denn in einem Satz wie „Der Schüler von Plato und Lehrer Alexanders wurde in Stagira geboren“ wird natürlich nur von jenem Schüler gesagt, dass er in Stagira geboren wurde, nicht etwa von Plato oder gar von Alexander. Nichtsdestotrotz lässt sich kaum bestreiten, dass in irgendeinem Sinn von „Bezugnahme“ in dem genannten Ausdruck auch auf Plato und Alexander Bezug genommen wird.

In diesem allgemeineren Sinn von Bezugnahme bedeutet Bezugnahme im Wesentlichen dasselbe wie Repräsentation. Repräsentation, Bezugnahme ist intentionaler, geistiger, sprachlicher Kontakt mit der Realität, der Bezug auf Gegenstände, Eigenschaften, Relationen, Zustände, Ereignisse, Sachverhalte usw.. In diesem allgemeineren Sinn von Bezugnahme beziehen sich auch Relationsausdrücke wie Verben und Präpositionen auf etwas. Z.B. bezieht sich der Ausdruck „geboren in“ auf eine Relation zwischen einer Person oder einem Tier und einem Ort bzw. auf bestimmte In-

stanzen oder Vorkommnisse dieses Relationentyps. Eine solche Instanz des Relationentyps „geboren in“ ist etwa die Relation zwischen Aristoteles und Stagira.<sup>24</sup>

Wie verhält sich der Begriff der singulären Bezugnahme zu diesem allgemeineren Konzept? Innerhalb einer Nominalgruppe wie „Der Schüler von Plato“ ist, wie schon beschrieben, der besagte Schüler besonders hervorgehoben. Obwohl innerhalb dieser Nominalgruppe auch auf Plato Bezug genommen wird, ist er, nicht Plato, das logische Subjekt. Dem auf diese Weise herausgegriffenen Schüler können dann Eigenschaften und Relationen zugeschrieben werden und dies geschieht wiederum – wie sollte es anders sein – durch Bezug auf diese Eigenschaften und Relationen.

### 4.3 Russell: Referenz vs. Denotation

Bertrand Russell war der Meinung, dass Identitätssätze mit ‚echten‘, von ihm als „logisch“ bezeichneten Eigennamen, tatsächlich leere Tautologien waren und zwar gleichgültig, ob es sich um Sätze der Form „ $a = a$ “ oder „ $a = b$ “ handelt.<sup>25</sup> Ihm war also klar, dass, wie hier schon eingangs argumentiert wurde, die bloße Tatsache der Verschiedenheit der Zeichen den vermeintlichen empirischen Charakter mancher Identitätsaussagen nicht verständlich machen kann, insbesondere wenn man sich, wie Russell dies tat, die Beziehung zwischen Name und Gegenstand als eine einfache, direkte, nicht weiter analysierbare Beziehung denkt. Er glaubte ferner aus den ebenfalls eingangs schon erwähnten Gründen, dass negative Existenzsätze mit logischen Eigennamen widersprüchlich wären: Logische Eigennamen würden die Existenz eines Benannten garantieren, so dass dessen Existenz zu verneinen sinnwidrig wäre. Russell meinte allerdings, dass unsere gewöhnlichen Eigennamen keinen logische Eigennamen in seinem Sinn sind. Ähnlich wie Frege nahm er an, dass der Sinn der gewöhnlichen Eigennamen dem von Beschreibungen entspricht; er fasste sie als Abkürzungen für Beschreibungen auf. Nur in speziellen Kontexten kann man sich mit logischen Eigennamen auf Entitäten beziehen, nämlich wenn man mit diesen ‚be-

---

<sup>24</sup> Diese Übertragung der Unterscheidung zwischen Typen und ihren Token oder Vorkommnissen auf Relationen und auch auf Eigenschaften habe ich in meinem Manuskript „Having and Instantiating Properties“ weiter ausgearbeitet. In Bezug auf Eigenschaften ist dies so zu verstehen, dass etwa die Grünheit des Grases und die Grünheit der Blätter im Frühling verschiedene Instanzen des Eigenschaftstyps grün sind. Nach der gängigen Auffassung instanziiert die Relata eine Relation, und die Träger der Eigenschaft die Eigenschaft. Meiner Meinung nach verwechselt diese Auffassung die Relata von Relationen und die Träger von Eigenschaften mit deren Instanzierungen. Ein grüner Baum hat die Eigenschaft grün zu sein, aber der Baum als solcher instanziiert den Typ Baum.

<sup>25</sup> Vgl. Russell 1918. Einen hilfreichen Überblick über Russells Theorie der Beschreibungen und seine Konzeption von Identitätssätzen gibt Williams 1989.

kannt' ist in einem speziellen Sinn von Bekanntschaft, der *knowledge by acquaintance*, für die das Paradigma die Bekanntschaft mit den eigenen ‚Sinnesdaten‘ ist. Der Gegenbegriff zu *knowledge by acquaintance* ist *knowledge by description*. Durch Beschreibung können wir auf indirektere Weise Wissen von Gegenständen erlangen, die uns nicht bekannt oder präsent sind. Russells Beispiel in seinem klassischen Aufsatz *On Denoting* ist das Trägheitszentrum des Sonnensystems, über das wir eine Reihe von Aussagen machen können, ohne doch mit ihm ‚bekannt‘ zu sein. Da für Russell nur logische Eigennamen sich wirklich auf ihren Gegenstand bezogen, bestand der ganze Witz von *On Denoting* darin zu zeigen, dass sich definite Beschreibungen wie „Der Schüler von Plato“, „Der gegenwärtige König von Frankreich“ entgegen dem Anschein – und auch entgegen der Meinung von Frege – nicht auf einen Gegenstand *referieren*: stattdessen *denotieren* sie diesen Gegenstand nur.

Um diesen Gegensatz zwischen genuiner Bezugnahme und der indirekteren Relation der Denotation terminologisch möglichst klar zu markieren, werde ich von nun an die direkte, unmittelbare Bezugnahme als *Referenz* von der Denotation abgrenzen. Der Begriff der Bezugnahme wird als Oberbegriff fungieren, unter den sowohl Referenz als auch Denotation fallen können, wenn man auch darauf insistieren könnte, dass Denotation streng genommen gar keine Form der Bezugnahme ist. Wenn man davon absieht, dass nach Russells Meinung gewöhnliche Eigennamen gar keine genuinen Eigennamen sind, wäre entsprechend Bertrand Russell das Referenzobjekt des Namens „Bertrand Russell“, aber die Denotation der definiten Beschreibung „Der Autor von *On Denoting*“. Ich werde von dieser Meinung Russells auch im Folgenden absehen, da ich, wie heutzutage die meisten Autoren, der Meinung bin, dass die Beschreibungstheorie gewöhnlicher Eigennamen falsch ist, aber zumindest einige der Dinge, die Russell über logische Eigennamen gesagt hat, tatsächlich auf die gewöhnlichen Eigennamen zutreffen. Auf die Gründe für diese Auffassung werde wir bei der Diskussion von Kripkes Theorie des Benennens noch zu sprechen kommen.

Durch seine Analyse wollte Russell drei Probleme lösen, die sich aus der Annahme ergeben, dass eine Beschreibung auf einen Gegenstand referiert. Zwei davon sind die schon erwähnten Schwierigkeiten bezüglich informativer Identitätssätze und Existenzsätze. Wie kann, erstens, ein Satz wie „Scott war der Autor von *Waverly*“ informativ sein, eine Aussage *a posteriori* ausdrücken? Dies scheint ja damit unvereinbar zu sein, dass dieser Satz ein genuiner Identitätssatz der Form „a = a“ oder

„ $a = b$ “ ist, in dem also zweimal mit logischen Eigennamen auf dasselbe Objekt Bezug genommen wird. Zweitens, wie kann man dem Vorkommen von Beschreibungen in negativen Existenzsätzen wie z.B. „Der gegenwärtige König von Frankreich existiert nicht“ Sinn abgewinnen, wenn man doch nicht zuerst auf einen Gegenstand referieren und ihm dann die Existenz absprechen kann? Das dritte – in Russells Aufzählung an zweiter Stelle genannte – Problem soll hier nur der Vollständigkeit halber kurz erwähnt werden: Wie kann das logische Gesetz des ausgeschlossenen Dritten für Sätze wie „Der gegenwärtige König von Frankreich ist kahl“ gelten, obwohl weder in der Liste der kahlen noch der nicht-kahlen Dinge ein gegenwärtiger König von Frankreich auftaucht?

Der Kern von Russells Lösung besteht darin, definite Beschreibungen durch Satzfunktionen zu analysieren. Was mit einer definiten Beschreibung wie „Der Autor von *Waverly*“ gemeint ist, können wir informell wie folgt paraphrasieren: „Es gibt eine und zwar nur genau eine Entität, die *Waverly* geschrieben hat“. Dies wiederum bedeutet, dass die Satzfunktion „ $x$  schrieb *Waverly*“ für genau ein  $x$  den Wert „wahr“ annimmt. Die Bedingung, dass es sich nur um genau ein  $x$  handelt, die ich im Folgenden als die „Einzigkeitsbedingung“ bezeichnen werde, drückt Russell wie folgt aus: Wenn „ $y$  schrieb *Waverly*“ wahr ist, dann muss „ $x = y$ “ wahr sein, müssen  $x$  und  $y$  also identisch sein. Dies führt zu einer revidierten Konzeption von Identitätssätzen wie „Scott war der Autor von *Waverly*“, die Russell im Folgenden erläutert und rechtfertigt:

The usefulness of *identity* is explained by the above theory. No one outside a logic-book ever wishes to say “ $x$  is  $x$ ”, and yet assertions are often made in such forms as “Scott was the author of *Waverly*” or “thou art the man.” The meaning of such propositions cannot be stated without the notion of identity, although they are not simply statements that Scott is identical with another term, the author of *Waverly*, or that thou art identical with another term, the man. The shortest statement of “Scott is the author of *Waverly*” seems to be “Scott wrote *Waverly*; and it is always true of  $y$  that if  $y$  wrote *Waverly*,  $y$  is identical with Scott”. It is in this way that identity enters into “Scott is the author of *Waverly*”; and it is owing to such uses that identity is worth affirming.

(Russell 1905: 55)

Die Nützlichkeit des Identitätszeichens besteht also *nicht* darin, dass es uns erlaubt, von einem Gegenstand zu sagen, dass er in der Identitätsbeziehung zu sich selbst steht, sondern darin, dass es uns ermöglicht, die Einzigkeitsbedingung auszudrücken.

Wenn wir „schrieb *Waverly*“ mit „W“ symbolisieren und Scott mit „a“ können wir Russells Analyse wie folgt wiedergeben:

$$\exists x \{Wx \ \& \ [\forall y (Wy \rightarrow x = y) \ \& \ x = a]\}$$

Die definite Beschreibung wird also durch einen Existenzquantor verstanden, der ausdrückt, dass mindestens ein  $x$  die Beschreibung erfüllt und die Einzigkeitsbedingung, die festlegt, dass nicht mehr als ein  $x$  sie erfüllt. Die Einzigkeitsbedingung wird uns im Folgenden nicht mehr groß interessieren. Von zentraler Bedeutung ist aber, dass die Aussage, dass ein einziger Gegenstand eine Beschreibung erfüllt, nicht mit der Referenz auf diesen Gegenstand verwechselt werden darf. Das ist der Kern der Unterscheidung zwischen Referenz und Denotation, und so werden ja gerade die Paradoxa der Existenz und der Identität vermieden. Wir beziehen uns auf eine Eigenschaft und sagen, dass genau ein Gegenstand sie hat. Aber wir verfügen nicht über einen *Namen* dieses Gegenstandes und sind nicht mit ihm *bekannt*, jedenfalls nicht kraft der Existenzaussage, die wir machen.

Wie sind entsprechende Sätze mit zwei Beschreibungen zu verstehen? Williams beschreibt Russells Auffassung dieser Sätze wie folgt:

Russell allows that, where “Scott” and “Sir Walter” are both being used as proper names, “Scott is the same as Sir Walter” is a futile tautology. The paradox is allowed its full force where that is concerned. But in “The author of *Waverly* is the same as the author of *Marmion*” Russell sees a quite different structure. His analysis of this can be stated for this purpose without the benefit of logical symbolism. The proposition is designed to assert that being sole author of *Waverly* and being the sole author of *Marmion* are both properties of a single person. Someone both wrote *Waverly* and wrote *Marmion* (taking it for granted that no two people wrote either). The structure of the proposition is thus given by identifying the pattern “Someone both --- and .....” and seeing it as realized by filling the two blanks with the verb-phrases “wrote *Waverly*” and “wrote *Marmion*”. The paradox arose from assuming that identity propositions had the job of stating a relationship between objects (an object and itself).” (Williams 1989: 13).

Russell zeigt uns also, wie wir das Paradox vermeiden können, indem wir den Satz nunmehr so verstehen, dass damit ausgesagt wird, dass genau ein Jemand zwei verschiedene Eigenschaften hat – er hat sowohl *Waverly* als auch *Marmion* geschrieben.



Ein Identitätssatz der Form „ $x = y$ “ ist aber immer noch ein unverzichtbarer Teil seiner Analyse. Im Kontext dieser Analyse hat er die doppelte Funktion sicherzustellen, dass es nur genau einen Menschen gegeben hat, der *Waverly* geschrieben hat und dass derselbe Mensch auch *Marmion* geschrieben hat. Und die Art, wie er dies tut, lädt noch immer zu Paraphrasen ein, die einen paradoxen Klang haben. Denn dass nur ein einziger Mensch *Waverly* geschrieben hat, wird ja durch das Konditional wiedergegeben, dass, wenn  $x$  und  $y$  *Waverly* geschrieben haben,  $x$  und  $y$  identisch sind. Mit etwas bösem Willen könnte man den Inhalt dieses Konditionals so paraphrasieren: Wenn zwei Dinge *Waverly* geschrieben haben, dann sind sie miteinander identisch – also nur ein Ding.

Bevor wir dazu kommen, wie Wittgensteins Radikalisierung der Russellschen Analyse uns erlaubt, ganz ohne das Identitätszeichen und jeden Anschein von Paradoxie auszukommen, möchte ich den Gegensatz zwischen Referenz und Denotation noch vertiefend behandeln. Dies scheint aus mehreren Gründen angebracht zu sein. Erstens ist in der Sprachphilosophie seit Russell der fundamentale Gegensatz zwischen Referenz und Denotation oft nicht genug beachtet worden oder sogar gänzlich unter den Tisch gefallen. So werden Namen und Beschreibungen ohne Beachtung dieses Gegensatzes als *singuläre Termini* oder *Designatoren* oft in eine Kategorie gesteckt. Wir werden später sehen, dass die Verwischung dieses Gegensatzes für Kripkes Theorie theoretischer Identifikationen und die Debatte über die psychophysische Identitätstheorie von großer Bedeutung ist. Ein Grund für diese Nichtbeachtung der Unterscheidung liegt sicherlich darin, dass Russells Unterscheidung zwischen *knowledge by acquaintance* und *knowledge by description* mit fragwürdigen metaphysischen Auffassungen zusammenhing. Das zweite Ziel der folgenden Diskussion ist deshalb zu zeigen, dass diese Unterscheidung auch ganz unabhängig von Russells metaphysischen Annahmen intuitiv höchst plausibel und theoretisch unverzichtbar ist. Drittens schließlich möchte ich gegen Russell dafür argumentieren, dass auch definite Beschreibungen auf etwas referieren. Russell hat zwar Recht, dass „Der Autor von *Waverly*“ nicht auf Scott referiert oder „Der deutsche Bundeskanzler“ nicht auf Gerhard Schröder, selbst wenn diese Personen tatsächlich *Waverly* geschrieben haben bzw. Bundeskanzler sind. Diese Individuen werden von diesen Ausdrücken nur denotiert. Dennoch, so möchte ich behaupten, referieren diese Ausdrücke auch auf etwas.

Worauf aber sollen sich diese Beschreibungen beziehen, wenn nicht auf die Individuen, auf die sie zutreffen? Nun, die Antwort kann einfach lauten: „Der Autor von Waverly“ referiert auf den Autor von Waverly und „Der deutsche Bundeskanzler“ auf den deutschen Bundeskanzler!<sup>26</sup> Diese Antwort bedarf allerdings der Erläuterung. Nehmen wir das Beispiel des Bundeskanzlers. „Bundeskanzler“ bezeichnet einen bestimmten Typ von Rolle, die ein Mensch innehaben kann, ein Amt. „Der Bundeskanzler“ bezeichnet in einem gegebenen Kontext eine bestimmte Instanziierung dieses Rollentyps. Der Punkt ist nun, dass man zwischen der Instanziierung einer Rolle und dem Träger dieser Rolle unterscheiden muss.

Die verschiedenen Instanziierungen der Bundeskanzlerrolle sind nicht etwa Konrad Adenauer, Helmut Schmidt, Helmut Kohl, Gerhard Schröder usw. sondern, wie man sagen könnte, Konrad Adenauers Bundeskanzlersein, Helmut Schmidts Bundeskanzlersein, Helmut Kohls Bundeskanzlersein und Gerhard Schröders Bundeskanzlersein. Nun zeigt diese Aufzählung, dass es in der Tat eine enge Beziehung zwischen den Instanzen der Bundeskanzlerrolle und den Trägern dieser Rolle gibt. Es liegt nahe, die Identität der Träger als Individuationsprinzip für die Instanzen der Rolle zu benutzen, und so verfahren wir gewöhnlich auch. So wird etwa oft gesagt, Gerhard Schröder sei der 7. Bundeskanzler der Bundesrepublik Deutschland gewesen, weil er eine von 7 verschiedenen Personen ist, die bis jetzt Träger dieser Rolle gewesen sind. Es ist aber auch möglich, die Instanzen der Bundeskanzlerrolle und ähnlicher Rollen statt durch die Träger der Rolle durch die Amtszeiten zu individualisieren. So könnte man auch sagen, dass Schröder zweimal Bundeskanzler der Bundesrepublik Deutschland gewesen ist – von 1998-2002 und dann wieder von 2002-2005. Und oft werden Formulierungen dieser Art tatsächlich gebraucht. Von Churchill sagt man gewöhnlich, er sei zweimal *the Prime Minister* gewesen. Damit ist natürlich nicht gemeint, dass er zweimal Churchill war.

Auch in einem Satz wie „Der Bundeskanzler bestimmt die Richtlinien der Politik und trägt dafür die Verantwortung“ aus dem Grundgesetz (Art. 65) ist klar, dass „der Bundeskanzler“ nicht durch „Gerhard Schröder“ ersetzt werden kann und zwar nicht nur deshalb nicht, weil Gerhard Schröder nur der derzeitige Bundeskanzler ist und die Verfassung natürlich für sämtliche Bundeskanzler gilt. Der Grund besteht darin, dass nicht die Person Gerhard Schröder *per se* die Richtlinien der Politik bestimmt, selbst in dem Zeitraum nicht, in dem sie tatsächlich Bundeskanzler ist,

---

<sup>26</sup> So lässt sich auch die Vorstellung einer ‚kontingenten Identität‘ z.B. von 9 und der Anzahl der Planeten überwinden.

sondern nur Gerhard Schröder *als Bundeskanzler*, in Ausübung seines Amtes, im Kontext der verfassungsgemäßen Ordnung und des Gefüges der durch sie bestimmten, den Staat ausmachenden Institutionen. Dass Unterscheidungen dieser Art nötig sind, zeigt sich auch daran, dass mitunter solche Fragen aufgeworfen werden, wie ob jemand etwas als Träger des einen oder anderen Amtes oder gar als Privatperson etwas Bestimmtes gesagt oder gemacht hat. Hat Gerhard Schröder die Opposition als Bundeskanzler oder als SPD-Vorsitzender angegriffen? Hat X als Minister an der Veranstaltung teilgenommen oder als Parteipolitiker? Von der Beantwortung dieser Frage kann abhängen, ob sein Flug zu Recht dem Staat in Rechnung gestellt wurde – und damit, ob er sein Amt behält.

Der wichtigste Punkt ist aber, dass einen Ausdruck wie „der Bundeskanzler“ zu verstehen nicht erfordert, die Person zu *kennen*, die Bundeskanzler ist. Wir würden nicht allein deshalb, weil jemand nicht weiß, wer Bundeskanzler ist, sagen, er verstünde diesen Ausdruck nicht. Er muss nur verstehen, was für ein Amt das ist. Das Kennen einer Person – nicht unbedingt das persönliche – oder anderer Objekte ist eine grundlegende intentionale Beziehung, auf der der Gebrauch von Eigennamen und vielen anderen sprachlichen Ausdrücken zumindest in den zentralen Fällen basiert. Diese grundlegende Intuition Russells ist von seiner spezifischen Metaphysik unabhängig und scheint mir bewahrenswert. Sie wird eine wichtige Rolle in der Auseinandersetzung mit der materialistischen Metaphysik der Identitätstheorie spielen, die implizit oder explizit versucht, Ausdrücke, mit denen wir etwas uns Bekanntes bezeichnen, durch Be- oder Umschreibungen umzuinterpretieren, um dann das, was wir zu kennen meinen, stillschweigend zu eliminieren. Ich werde das Kennen auch als *transparente* Bezugnahme bezeichnen, während die Denotation eine *opake* Form der Bezugnahme ist – wenn man sie denn überhaupt so bezeichnen möchte. Sie gibt uns einen Pfad zu einem Objekt, einen Weg, es kennen zu *lernen*, aber keinen direkten Zugang zu ihm selbst.

#### **4.4 Wittgenstein: Die Elimination des Identitätszeichens**

Wittgenstein radikalisierte Russells Analyse von Identitätssätzen. Im *Tractatus* zeigte er einen Weg, jede Verwendung des Identitätszeichens zu vermeiden. Den Grundgedanken seiner Lösung formuliert er so:

Gleichheit des Gegenstands drücke ich durch Gleichheit des Zeichens aus, und nicht mit Hilfe eines Gleichheitszeichens. Verschiedenheit der Gegenstände durch Verschiedenheit der Zeichen. (Tractatus-logico-philosophicus 5.53)

Dass also etwa derselbe Gegenstand zwei verschiedene Eigenschaften hat, können wir, wie wir dies ja gemeinhin auch tun, einfach dadurch ausdrücken, dass wir zweimal denselben Namen verwenden. Wir sagen etwa, dass Wittgenstein den *Tractatus* geschrieben hat und dass Wittgenstein *Philosophische Untersuchungen* geschrieben hat. Durch die Verwendung desselben Namens können wir denselben Gegenstand durch den Wandel der Zeit hindurch verfolgen. Zwei Namen für denselben Gegenstand zu haben, eröffnet nicht die Möglichkeit informativer Identitätssätze, sondern ist bestenfalls ein überflüssiger Luxus. Statt einen Identitätssatz zu bilden, könnten wir stets einfach einen der beiden Eigennamen aus unserer Sprache eliminieren – oder sogar beide zugunsten eines dritten.

Analog können auch die Variablen im Prädikatenkalkül behandelt werden. Dazu muss die gängige, bis heute übliche, so genannte inklusive Interpretation der Variablen durch die so genannte exklusive Interpretation ersetzt werden (vgl. Hintikka 1956). Die exklusive Interpretation schließt aus, dass verschiedene Variablen innerhalb des Bereichs eines Quantors dieselben Werte annehmen können. D.h. eine Verschiedenheit der Variablen drückt nunmehr eine Verschiedenheit der Werte aus, die sie annehmen können, genau wie die Verschiedenheit der Konstanten (Eigennamen) eine Verschiedenheit der benannten Gegenstände ausdrückt. Wittgenstein führt an einer Reihe von Beispielen vor, wie dies funktioniert, ausgehend noch einmal von Beispielen mit Eigennamen:

Ich schreibe also nicht „ $f(a,b).a=b$ “, sondern „ $f(a,a)$ “ (oder „ $f(b,b)$ “). Und nicht „ $f(a,b).\sim a=b$ “, sondern „ $f(a,b)$ “.

Und analog: Nicht „ $(\exists x,y).f(x,y).x=y$ “, sondern „ $(\exists x).f(x,x)$ “; und nicht „ $(\exists x,y).f(x,y).\sim x=y$ “, sondern „ $(\exists x,y).f(x,y)$ “.

(Also statt des Russellschen „ $(\exists x,y).f(x,y)$ “

„ $(\exists x,y).f(x,y).v.(\exists x).f(x,x)$ “.)

Statt „ $(x):fx.\supset x=a$ “ schreiben wir also z.B. „ $(\exists x).fx.\supset fa:\sim(\exists x,y).fx.fy$ “.

Und der Satz: „Nur Ein x befriedigt f( )“ lautet:

„ $(\exists x).fx:\sim(\exists x,y).fx.fy$ .“

(Tractatus 5.531-5.5321)

Das letzte Beispiel ist die Formalisierung der Einzigkeitsbedingung. Da nun verschiedene Variablen innerhalb des Skopus eines Quantors nur mehr verschiedene Werte annehmen können, lässt sich mittels „ $(\exists x).fx: \sim(\exists x,y).fx.fy$ “ ausdrücken, dass nur ein Gegenstand  $f$  ist. Für diese Formalisierung wäre die Paraphrase „Es gibt mindestens ein Ding, das  $f$  ist, und es gibt keine zwei Dinge, für die beide gilt, dass sie  $f$  sind“ angebracht, die keinen paradoxen Klang hat. Wir vermeiden die Formulierung, die zu der Paraphrase „Wenn es zwei Dinge sind, sind es in Wirklichkeit nur eins“ einlädt. Diese Paraphrase tut zwar der inklusiven Interpretation insofern Unrecht, als in dieser die Verwendung zweier verschiedener Variablen ja eben nicht die Existenz zweier verschiedener Gegenstände als Werte dieser Variablen bedeutet. Die Variablen können sowohl den selben als auch verschiedene Werte annehmen. Deshalb ist es ja nötig, mittels des Identitätssatzes auszudrücken, dass es sich nur um einen Gegenstand handelt. Aber die böswillige Paraphrase hat doch einen Witz, weil sie vorführt, wie natürlich es ist, die Verschiedenheit der Zeichen als Zeichen der Verschiedenheit zu interpretieren. Die inklusive Interpretation lässt dieses Ausdrucksmittel ungenutzt und ist deshalb gezwungen, das Identitätszeichen zu verwenden. Wittgenstein dagegen zeigt uns, dass ein eigenes Zeichen für die Identität in Form eines Relationsausdrucks überflüssig ist. Identität und Verschiedenheit des Bezeichneten kann hinreichend durch Identität und Verschiedenheit der Zeichen ausgedrückt werden. Deshalb sollte ein ideale, auf das Wesentliche verknappte Sprache ohne ein Identitätszeichen auskommen. Für Wittgenstein ist dieses Resultat gleichbedeutend damit, dass es eine Relation der Identität auch gar nicht gibt. Ich glaube, dass Wittgenstein damit auch Recht hat. Wenn wir alles, was wir mit den so genannten Identitätsaussagen sagen wollen, auch ohne diese ausdrücken können, gibt es keinen Grund zu der Annahme einer Identitätsrelation. Die Identitätsrelation ist genauso überflüssig wie ihr Zeichen. Für die im Folgenden entwickelten Überlegungen ist es aber nicht unbedingt wesentlich, dass man diese Konklusion akzeptiert. Wesentlich ist nur das Resultat, dass insofern die bisher betrachteten so genannten Identitätssätze empirischen Gehalt haben, dieser sich nicht dadurch erklären lässt, dass von einem Gegenstand gesagt wird, er sei mit sich selbst identisch, sondern nur dadurch, dass ein und demselben Gegenstand verschiedene Eigenschaften zugeschrieben werden. Dass dies sich auch gänzlich ohne Verwendung des Identitätszeichens ausdrücken lässt, ist nur das Tüpfelchen auf dem „i“ dieser Argumentation.

## 4.5 Kripke: Kontingenz *a priori* und Notwendigkeit *a posteriori*

Der letzte große Einschnitt im Nachdenken über Identitätssätze und den Bezug von Namen – Eigennamen sowohl als Namen für Stoffe und andere ‚natürliche Arten‘ – wird von Saul Kripkes *Naming and Necessity* markiert. Da auch die gegenwärtige Diskussion über das Leib-Seele-Problem im Allgemeinen und die psychophysische Identitätstheorie im Besonderen stark durch Kripkes Überlegungen geprägt ist, werde ich mich ausführlich kritisch mit Kripkes Hauptthese, dass es apriorisches Wissen kontingenter Tatsachen und empirisches Wissen metaphysisch notwendiger Identitätsrelationen gibt. Einleitend ein paar kurze Bemerkungen zum Stand der Debatte in den 1950er und 1960er Jahren, also in der Zeit vor der Publikation von *Naming and Necessity* (1971).

Wittgensteins Auffassung, dass es eine Identitätsrelation gar nicht gibt, dass Identitätssätze bestenfalls tautologisch sind und ohne Verlust aus einer logisch angemessenen Sprache eliminiert werden können, wurde kaum beachtet.<sup>27</sup> Kontra Wittgenstein und Russell etablierte sich die Meinung, es könne ‚kontingente Identitäten‘ (*contingent identities*) geben – wobei dieser Begriff zwischen Identitätsrelation und -aussage changierte. Die frühen psychophysischen Identitätstheoretiker etwa gingen davon wie selbstverständlich aus. Ein bis heute andauernder Problemverlust scheint damals eingesetzt zu haben. Auch W.V.O. Quine (1960: 114ff) etwa lehnte Wittgensteins Auffassung von Identität genauso ab wie Freges Sinnbegriff, die Frege-Russell-Theorie gewöhnlicher Eigennamen und die metalinguistische Deutung und sah dennoch kein Problem darin, die Existenz empirischer Identitätsaussagen zu behaupten, ohne eine eigene Antwort auf das Paradox der Identität zu suchen:

We may tag the planet Venus, some fine evening, with the proper name “Hesperus“. We may tag the same planet again, some day before sunrise, with the proper name “Phosphorus“. When we discover that we have tagged the same planet twice our discovery is empirical. And not because the proper names were descriptions.

(zit. nach Kripke 1980: 100)

Auch bei Kripke, der zwar seine Existenz anerkennt, aber ansonsten nicht nennenswert darauf reagiert, hat man den Eindruck, dass er es nicht sehr ernst nimmt. Kripke

---

<sup>27</sup> Dies ist bis heute so. Der Artikel über Identität in der *Routledge Encyclopedia of Philosophy* (Williamson 1998) etwa erwähnt Wittgenstein gar nicht. Williams spricht treffend von „Wittgenstein’s curiously neglected denial that identity is a relation“ (op. cit.: vii).

kritisierte aber, wie vorher schon Ruth Barcan Marcus (1963), die vorherrschende Auffassung ‚kontingenter Identitäten‘, obwohl auch er die Existenz solch hölzerner Eisen für die Fälle zugesteht, in denen die entsprechenden Aussagen so genannte *nicht-starre Designatoren* enthalten, Beschreibungen, die nicht ‚in allen möglichen Welten den selben Gegenstand bezeichnen.‘ Da Kripke die Identitätsrelation für notwendig hält und zwar für ‚metaphysisch notwendig‘, kommt ihm an dieser Stelle die Ambiguität des Begriffs ‚kontingenter Identitäten‘ zur Hilfe. Identitätsaussagen mit starren Designatoren, d.h. Eigennamen und Namen ‚natürlicher Arten‘ wie Gold, Licht, Wasser, Wärme usw. – aber auch bestimmten Beschreibungen – sollen dagegen stets notwendig sein. Die neben seiner Kritik an der Beschreibungstheorie der Eigennamen einschneidendste von Kripke vorgeschlagene Revision ist seine Dissoziation der Notwendigkeit und des *a priori* auf der einen Seite und der Kontingenz und des *a posteriori* auf der anderen. Damit stellt sich Kripke gegen die gesamte neuzeitliche philosophische Tradition. Das *a priori* und das *a posteriori* sind für Kripke epistemische Kategorien. Sie beziehen sich darauf, wie Wahrheiten erkannt werden, durch die Erfahrung oder unabhängig von der Erfahrung. Die Kategorien des Notwendigen, des Kontingenten, essentieller Eigenschaften usw. sollen dagegen metaphysische Kategorien sein. Sie sind anwendbar auf Entitäten in der Welt, auf Tatsachen oder Eigenschaften, unabhängig davon, welchen epistemischen Zugang wir zu diesen haben. Dass etwas notwendig ist, leitet sich nicht davon her, wie wir davon wissen. Wir können auch einen Zugang *a posteriori* dazu haben. Z.B. können wir *a posteriori* erkennen, dass es das Wesen von Wasser ist, H<sub>2</sub>O zu sein. Die Identität von Wasser und H<sub>2</sub>O ist eine notwendige, aber empirisch erkannte Tatsache. Umgekehrt kann eine kontingente Gegebenheit *a priori* erkannt werden.

#### **4.6 Kontingenz *a priori***

Beginnen wir mit Kripkes These, dass metaphysisch kontingente Tatsachen auch *a priori* erkannt werden können. Kripke führt diese These unter Bezug auf Wittgensteins Diskussion des Pariser Urmeters in § 50 der *Philosophischen Untersuchungen* ein:

Man kann von *einem* Ding nicht aussagen, es sei 1 m lang, noch, es sei nicht 1 m lang, und das ist das Urmeter in Paris. – Damit haben wir diesem aber natürlich nicht irgend-

eine merkwürdige Eigenschaft zugeschrieben, sondern nur seine eigenartige Rolle im Spiel des Messens mit dem Metermaß gekennzeichnet.

Kripke kommentiert dies wie folgt:

This seems to be a very ‘extraordinary property’, actually, for any stick to have. I think he must be wrong. If the stick is a stick, for example, 39.37 inches long (I assume we have some different standard for inches), why isn’t it one meter long? Anyway, let’s suppose that the is wrong that the stick is one meter long. Part of the problem that is bothering Wittgenstein is, of course, that this stick serves as a standard of length and so we can’t attribute length to it. Be this as it may (well, it may not be), is the statement “stick S is one meter long”, a necessary truth? Of course its length might vary in time. We could make the definition more precise by stipulating that one meter is to be the length of S at a fixed time  $t_0$ . Is it then a necessary truth that stick S is one meter long at time  $t_0$ ? Someone who thinks that everything one knows *a priori* is necessary might think: “This is the definition of a meter. By definition, stick S is one meter long at  $t_0$ . That’s a necessary truth.” But there seems to me to be no reason so to conclude, even for a man who uses the stated definition of ‘one meter’. For he’s using this definition not to give the meaning of what he called the ‘meter’, but to fix the reference. ... There is a certain length which he wants to mark out. He marks it out by an accidental property, namely that there is a stick of that length. Someone else might mark out the same reference by another accidental property. But in any case, even though he uses this to fix the reference of his standard of length, a meter, he can still say, “if heat had been applied to this stick S at  $t_0$ , then at  $t_0$  stick S would not have been one meter long.” (1980: 55)

Die hier entscheidende, nicht ausklammerbare Frage ist tatsächlich, wie die Tatsache, dass der Urmeter als Längenmaß oder Paradigma einer bestimmten Länge fungiert, zu verstehen ist und welche Auswirkungen sie hat. Machen wir uns zuerst klar, dass auch Kripkes Darstellung voraussetzt, dass das Urmeter als Paradigma oder Standard fungiert.

Denn wie sollten wir sonst, wie Kripke unterstellt, *a priori* wissen können, dass dieser Stab einen Meter lang ist? Der nächste wichtige Punkt ist, dass Kripke Wittgenstein missversteht, wenn er ihm die These unterstellt, man könne gar nicht *sagen*, das Urmeter sei ein Meter lang, und es sei entsprechend sogar nicht einmal 1 m lang! Es ist verständlich, dass Kripke diese These sehr merkwürdig findet, aber es ist nicht Wittgensteins These. Wittgensteins These ist, dass wenn wir sagen, dass dieser besondere Stab, das Urmeter, 1 m lang ist, dies keine *Aussage* ist. (Elizabeth Anscom-



bes Übertragung der *Untersuchungen* (1953) befördert dieses Missverständnis, da dort „aussagen“ mit „say“ übersetzt wird.)

Wir betrachten also den Fall, dass jemand, in einem angemessenen Kontext auf das Urmeter deutend, einen Satz äußert wie „Dies ist 1 Meter“ oder „Dies ist 1 Meter lang“. Dabei muss dieser Satz aber so verstanden werden, dass er weder das Resultat einer Messung der Länge des Stabes *berichtet* – die ja nur mittels eines anderen Maßstabs oder sonstigen Standards möglich wäre – *noch*, dass dieser Stab bzw. seine Länge das Paradigma der Längeneinheit „Meter“ für einen bestimmten Sprecher oder für eine Gruppe von Sprechern ist.

Wenn die Akzeptanz des Satzes auf einer Messung mittels eines anderen Maßstabes beruht, ist es unproblematisch, dass mit ihm eine Aussage gemacht wird, aber dann handelt es sich auch zweifelsohne um eine Aussage *a posteriori*, und das philosophische Problem, um das es hier geht, wird verfehlt. Aus diesem Grund ist auch Kripkes Hinweis, dass der Stab 39.37 Zoll lang ist, für das eigentliche Problem irrelevant, weil er einen anderen Standard ins Spiel bringt. (Ferner wirft er die weitere Schwierigkeit auf, wie denn ein Satz wie „1 m = 39.37 Zoll“ zu verstehen ist. Drückt dies das Resultat einer bloßen Messung aus und ist damit eine Aussage *a posteriori* oder ist dies wiederum eine Festlegung *a priori*? Entsprechend wäre auch hier wieder zu fragen, ob der Satz Wissen von einer kontingenten Tatsachen ausdrückt.)

Dasselbe gilt, wenn wir den Satz als bloßen Bericht einer Messpraxis verstehen würden, so wie ihn etwa ein Anthropologe oder Wissenschaftstheoretiker meinen könnte. In dieser Lesart könnte man den Satz etwa so paraphrasieren: „Manche Menschen verwenden diesen Stab als Maßstab für eine Längeneinheit, die sie „Meter“ nennen.“ Auch hier wird unzweifelhaft eine Aussage *a posteriori* gemacht, die auf der Beobachtung der Sitten, Gebräuche und Institutionen einer bestimmten Gruppe beruht. Als *a priori* kann man den Satz nur in einer Lesart verstehen, in der er so etwas ausdrückt wie „Dieser Stab soll als Maßstab einer Länge dienen, die ich hiermit als „Meter“ bezeichne“. D.h. mit einer Äußerung des Satzes „Dies ist ein Meter“, die so gemeint ist, wird dem Stab die Rolle als Maßstab, als Paradigma einer bestimmten Länge, *zugewiesen* oder *verliehen*; gleichzeitig wird diese Länge *benannt* oder *getauft* und damit der Ausdruck „Meter“ als Bezeichnung einer Längeneinheit *definiert*.<sup>28</sup>

---

<sup>28</sup> Damit dies geschehen kann, muss natürlich, wie Wittgenstein es ausdrücken würde, „bereits viel in der Sprache vorbereitet sein“; aber wir wollen hier den entsprechenden Hintergrund, vor dem z.B. klar

Die kursiv gesetzten Ausdrücke sind Indikatoren dafür, dass es sich hier tatsächlich nicht um eine Aussage handelt. Sprechakttheoretisch gesehen gehört die Äußerung nicht wie Aussagen in die Kategorie der Behauptungen, sondern in die der Deklarationen (Searle 1975).<sup>29</sup> Der Sprecher *erklärt* einen bestimmten Gegenstand zum Maßstab einer Längeneinheit und einen bestimmten Laut zum Namen dieser Längeneinheit. Er *kreiert* bestimmte Tatsachen: der Stab ist nunmehr ein Maßstab und der Laut ein Name der Längeneinheit.

Hiermit sollte auch klar sein, worin der Fehler von Kripkes Konzeption von Aussagen *a priori* über kontingente Tatsachen liegt: allein dadurch, dass der Sprecher den Stab zum Maßstab für eine bestimmte Länge erklärt und einen Laut zum Namen dieser Länge, erwirbt er kein Wissen von einer kontingenten Tatsache. Er weiß nun, dass der Stab ein Maßstab ist, und dass die Länge „Meter“ heißt; aber dieses Wissen kann Kripke nicht meinen, und es ist auch nicht *a priori*. Kripke muss ein Wissen von der Länge des Stabes im Sinn haben, von der kontingenten Tatsache, dass der Stab so-und-so lang ist; aber tatsächlich hat der Sprecher in der betrachteten Situation nichts über die Länge des Stabes erfahren (Dummett 1973). Sein Sprechakt drückt kein Wissen über die Länge dieses Stabes aus.

Kripke hat also der Vorstellung apriorischen Wissens von kontingenten Tatsachen mit seinem Beispiel keinen Sinn geben können. Verstanden als Deklaration, bedeutet der Satz „Dies ist 1 Meter“ etwas anders, als wenn man eine Aussage *a posteriori* damit macht. Nur in letzterer Interpretation ist sie auf eine kontingente Tatsache gerichtet. Die Deklaration ist dagegen eine Erklärung einer Absicht, eine Festlegung darauf, einen Stab und einen Laut in bestimmter Weise zu gebrauchen. Wir werden nun sehen, dass auch Sätze wie „Hesperus ist Phosphorus“ mehr mit einer Deklaration dieser Art gemein haben, als Kripkes diametrale Opposition zwischen Kontingenz *a priori* und Notwendigkeit *a posteriori* uns erwarten lässt.

---

ist, dass unser Satz tatsächlich als eine Definition der Länge des Stabes gemeint ist, als gegeben annehmen und ihn an dieser Stelle auch theoretisch nicht weiter untersuchen.

<sup>29</sup> Man könnte einwenden, dass es hier um Satzsemantik geht und nicht um Sprechakte. Es scheint mir aber ziemlich offensichtlich zu sein, dass die sprechakttheoretischen Kategorien hier die relevanten sind. Auf die damit verbundenen größeren theoretischen Fragen kann ich an dieser Stelle nicht eingehen.

## 4.7 Notwendigkeit *a posteriori*

Betrachten wir nun die Art von Aussagen, die Kripke als notwendig *a posteriori* auffasst, beginnend mit dem altherwürdigen Beispiel von Morgenstern/Phosphorus und Abendstern/Hesperus. Wie schon erwähnt, glaubt Kripke, dass Identitätsaussagen, wenn sie überhaupt wahr sind, notwendig wahr sein müssen – jedenfalls wenn es sich um Identitätsaussagen mit starren Designatoren handelt. Aber warum sollen wir uns nicht einen Fall vorstellen können, in dem Hesperus nicht Phosphorus gewesen wäre?

Let's suppose we refer to the same heavenly body twice, as 'Hesperus' and 'Phosphorus'. We say: Hesperus is that star over there in the evening; Phosphorus is that star over there in the morning. Actually, Hesperus is Phosphorus. Are there really circumstances under which Hesperus wouldn't have been Phosphorus? Supposing that Hesperus is Phosphorus, let's try to describe a possible situation in which it would not have been. Well, it's easy. Someone goes by and he calls two *different* stars 'Hesperus' and 'Phosphorus'. It may even be under the same conditions as prevailed when we introduced the names 'Hesperus' and 'Phosphorus'. But are those circumstances in which Hesperus is not Phosphorus or would not have been Phosphorus? It seems to me that they are not.

Now, of course I'm committed to saying that they're not, by saying that such terms as 'Hesperus' and 'Phosphorus', when used as names, are rigid designators. They refer in every possible world to the planet Venus. Therefore, in that possible world, the planet Venus is the planet Venus and it doesn't matter what any other person has said in this other possible world. How should *we* describe this situation? He can't have pointed to Venus twice, and in the one case called it 'Hesperus' and in the other 'Phosphorus', as we did. If he did so, then 'Hesperus is Phosphorus' would have been true in that situation too. He pointed maybe neither time to the planet Venus – at least one time he didn't point to the planet Venus, let's say when he pointed to the body he called 'Phosphorus'. Then in that case we can certainly say that the name 'Phosphorus' might not have referred to Phosphorus. We can even say that in the very position when viewed in the morning that we found Phosphorus, it might have been the case that Phosphorus was not there – that something else was there, and that even, under circumstances it would have been called 'Phosphorus'. But that is still not a case in which, a possible counterfactual situation in which, 'Hesperus' and 'Phosphorus' weren't names of the things they in fact are names of. Someone, if he did determine their reference by identifying descriptions, might even have used the very identifying descriptions we used. But still that's not a case in which Hesperus wasn't Phosphorus. For there could not have been such a case, given that Hesperus is Phosphorus." (Kripke 1980: 102f)

Kripkes Argumentation läuft also wie folgt: Wann immer wir versuchen, eine Situation zu beschreiben, in der Hesperus nicht Phosphorus gewesen wäre, beschreiben wir tatsächlich nur eine Situation, in der sich die Ausdrücke „Hesperus“ und „Phosphorus“ auf etwas anderes bezogen hätten als auf das, auf das sie sich tatsächlich beziehen. Aber dies ist keine Situation, in der Hesperus nicht Phosphorus gewesen wäre. Denn es geht doch darum, so drückt Kripke es an anderer Stelle aus (109 Fn. 51), ob sich in *unserer Sprache* eine solche kontrafaktische Situation beschreiben lässt. Wir beschreiben kontrafaktische Situationen, aber wir tun dies in der Sprache, die wir tatsächlich sprechen. Dass, wenn sich einer oder beide der Namen „Phosphorus“ und „Hesperus“ auf etwas anderes bezogen hätten, der Satz „Hesperus ist nicht Phosphorus“ wahr gewesen wäre, ist zwar wahr, aber irrelevant für die Frage, ob Hesperus auch nicht Phosphorus hätte gewesen sein können. Denn diese Frage ist in unserer gewöhnlichen Sprache formuliert, nicht in der kontrafaktischen Sprache, in der einer oder beide der Namen „Phosphorus“ und „Hesperus“ sich auf etwas anderes beziehen und der Satz „Hesperus ist nicht Phosphorus“ wahr ist.

Diese Überlegungen haben sicherlich eine gewisse Scheinbarkeit für sich, führen aber doch zu einer Reihe merkwürdiger Resultate. So scheint intuitiv klar zu sein, dass hier nicht wirklich eine Möglichkeit ausgeschlossen worden ist, sondern nur eine bestimmte Bezeichnung derselben. Kripke scheint uns nur sozusagen verbieten zu wollen, die Namen „Hesperus“ und „Phosphorus“ bei der Beschreibung der entsprechenden Szenarien zu benutzen. Natürlich wäre es aber doch möglich gewesen, dass an dieser Stelle zwei verschiedene Planeten oder andere Himmelskörper gewesen wären. Weiter sagt Kripke, den Fall, dass Hesperus nicht Phosphorus wäre, könne es nicht geben, *gegeben dass Hesperus nicht Phosphorus ist*. Den selben Sachverhalt kommentiert er an anderer Stelle so:

Some of the problems which bother people in these situations, as I have said, come from an identification, or as I would put it, a confusion, between what we can know *a priori* in advance and what is necessary. Certain statements – and the identity statement is a paradigm of such a statement on my view – if true at all must be necessarily true. One does know *a priori*, by philosophical analysis, that if such an identity statement is true it is necessarily true. (109)

Es ist merkwürdig, dass in diesem Fall die Frage nach der Wahrheit der Aussage und die nach der Notwendigkeit dieser Wahrheit zwei ganz verschiedene Fragen zu sein

scheinen. Die tatsächliche Wahrheit der Aussage erscheint als eine Bedingung ihrer notwendigen Wahrheit. Schließlich sagt Kripke selbst, dass wir die Notwendigkeit der Wahrheit *a priori* erkennen können. Dies legt nahe, dass wir ihre *tatsächliche* Wahrheit *a posteriori* erkennen. Dabei behauptet Kripke doch eigentlich, dass die Notwendigkeit der Aussage metaphysisch begründet ist und *a posteriori* erkannt wird. Es scheint, dass auch hier die metaphysischen und epistemologischen Kategorien sich nicht in der von Kripke behaupteten Weise trennen lassen.

Die an dieser Stelle entscheidende Frage ist aber meiner Meinung nach, ob die Sprecher nach der Entdeckung der Identität von Hesperus und Phosphorus *noch die gleiche Sprache sprechen* oder nicht. Meinten sie vorher dasselbe mit „Hesperus“ und „Phosphorus“ wie danach? Ich möchte nun im Folgenden dafür argumentieren, dass die Sprecher vor der Entdeckung eine andere Sprache gesprochen haben, und dass sich das Moment der Entdeckung am besten als ein empirisch motivierter Übergang von einer Sprache zur anderen verstehen lässt.

#### **4.8 ‚Informative‘ Identitätsaussagen als empirisch motivierte Redefinitionen**

*Each of our beliefs must be possible for a being who is a brain in a vat because each of us is precisely a brain in a vat; the vat is a skull and the ‘messages’ coming in are coming in by way of impact on the nervous system. The necessity of this internalism is masked from us in many discussions by the adoption of a third-person point of view. By adopting a God’s eye view we think we can see what Ralph’s real beliefs are even if he can’t. But what we forget when we try to construct a belief that is not entirely in Ralph’s head is that we have only constructed it in our head.*

– John Searle, Intentionality

Das Merkwürdige an Kripkes Auffassung ist, dass ihr zufolge „Hesperus“ und „Phosphorus“ sich schon zu einer Zeit auf dasselbe Objekt bezogen haben sollen, als die Sprecher noch glaubten, oder gar zu wissen meinten, dass Hesperus *nicht* Phosphorus ist. Denn sie meinten ja, es mit zwei verschiedenen Himmelskörpern zu tun zu haben. D.h. obwohl die Sprecher Hesperus und Phosphorus für zwei verschiedene Himmelskörper halten und obwohl sie auch die Absicht haben, mit „Hesperus“ und „Phosphorus“ auf diese Himmelskörper Bezug zu nehmen, sollen diese zwei Ausdrücke sich doch auf dasselbe beziehen.

Dies bedeutet, dass, welche Sprache die Sprecher sprechen, von Dingen abhängt, von denen sie nichts wissen. Sie scheinen selber die Sprache, die sie sprechen, nicht zu verstehen. „Hesperus“ und „Phosphorus“ beziehen sich in ihrer Sprache auf ein und denselben Gegenstand; sie verstehen diese Ausdrücke aber so, dass sie sich auf zwei verschiedene Gegenstände beziehen. Dass externalistische Positionen wie die von Kripke solche Folgen haben, ist oft bemerkt worden. Mir scheinen diese Folgen inakzeptabel zu sein. Einen sinnvollen Begriff der Bezugnahme, der eine solche Kluft aufreißt zwischen dem, was Sprecher denken und meinen, und dem, was ihre Worte bedeuten, kann es meiner Ansicht nach nicht geben.<sup>30</sup> Oder, anders ausgedrückt: Es muss einen Begriff der Bezugnahme geben, der eine solche Kluft nicht aufreißt, der den Inhalt des Geistes und seine Referenz nicht durch externe Gegebenheiten bestimmt, die dem Subjekt *unbekannt* sind.<sup>31</sup> Der Externalismus macht die Bezugnahme opak. Sie ist durch Sachverhalte bestimmt, von denen die Sprecher nichts wissen. Es muss aber ein Konzept der Bezugnahme geben, nach dem sie transparent ist; nach dem auf etwas als das Bezug genommen wird, als das es auch erkannt wird; nach dem es z.B. nicht möglich ist, das man nur eines meint, wo man doch zwei zu erkennen glaubt. Und wenn man einmal von einem solchen Begriff der Bezugnahme ausgeht, wird man schnell sehen, dass der opake Begriff der Bezugnahme weder nötig noch sinnvoll ist.

Wir können diese Überlegung mittels des folgenden – von Kripke stammenden – argumentativen Manövers anschaulicher machen.<sup>32</sup> Schenken wir doch Kripke

---

<sup>30</sup> Manche Autoren haben darauf durch eine Übertragung der externalistischen Lehren von der Sprache auf den Geist reagiert. Demnach bestünde keine Kluft mehr zwischen den Inhalten des Geistes und denen der Sprache, zwischen geistiger und sprachlicher Repräsentation und Referenz. Mir scheint dieses Manöver aber die mangelnde Plausibilität der externalistischen Position, um nicht zu sagen ihre Absurdität, zu potenzieren statt sie abzumildern oder gar aufzuheben.

<sup>31</sup> Ich möchte betonen, dass das Anstößige hier *nicht* in der These besteht, dass zur Bestimmung des Inhalts des Geistes auf Gegebenheiten Bezug genommen werden muss, die dem Geist buchstäblich extern sind, also räumlich außerhalb des Geistes zu lokalisieren sind. Diese These scheint mir vielmehr richtig zu sein. Anstößig ist vielmehr die Vorstellung, zur Bestimmung der Inhalt des Geistes sei der Bezug auf Gegebenheiten wesentlich, die dem Geist des jeweiligen Subjekts in dem viel stärkeren Sinn extern sind, dass sie außerhalb seiner Reichweite liegen, ihm *unbekannt* sind. Eine erschöpfende Diskussion des Externalismus kann im Rahmen dieser Arbeit nicht geleistet werden. Ich werde aber später an einigen Stellen noch einmal auf die Thematik zurückkommen. Eine tief greifende Analyse und Kritik verschiedener externalistischer Theorien, Argumentationen und Begriffsbildungen einschließlich der so genannten Kausaltheorie der Referenz findet sich bei Searle (1983: Kap. 8 und 9).

<sup>32</sup> Kripke bot in *Naming and Necessity* Zweifeln an der Existenz einer Relation der Selbstidentität an, Ihnen den Begriff „identity“ zu schenken, um damit die jeweils vorgeschlagenen Substitute zu bezeichnen, und behauptete, dies würde an der Sachlage aber nichts ändern, da all die ursprünglichen Probleme wieder in Bezug auf eine Relation auftauchen würden, die er nunmehr als „schmidentity“ bezeichnet, die aber natürlich nichts anderes ist als die alte Identitätsrelation unter neuem Namen. Dieser Zug wird aber einen Skeptiker nicht überzeugen, der das Konzept einer Identitätsrelation und insbesondere das von empirisch informativen Aussagen, die von Instanzen dieser Relation wahr gemacht werden, für inkohärent hält. (Eine ausführlichere Kritik von Kripkes Argumentation findet sich

und anderen Externalisten den Begriff der Referenz und führen wir stattdessen den Begriff „Schmeferenz“ ein! Schmeferenz ist eine transparente Form der Bezugnahme, so dass unter anderem gilt, dass eine Person mit zwei Ausdrücken nur dann auf denselben Gegenstand schmeferiert, wenn sie auch intendiert, mit diesen Ausdrücken auf denselben Gegenstand zu schmeferieren. Für die Schmeferenz gilt, dass es keine Kluft zwischen den geistigen Einstellungen einer Person und ihren sprachlichen Äußerungen gibt, sondern letztere erstere zum Ausdruck bringen. Es kann wohl kaum bezweifelt werden, dass dies ein sinnvolles Konzept ist. Dagegen scheint der Externalist nur *seine eigene Sprache* in den interpretierten Sprecher hineinzuprojizieren. So versteht er ihn nicht. Bleiben wir also bei einem transparenten Konzept der Referenz.

Eine Schwierigkeit liegt hier sicherlich darin, dass wir dazu neigen, Referenz als einen faktiven Begriff zu interpretieren. D.h. wir möchten Referenz so verstehen, dass wir, genau wie wir nicht etwas wissen können, was nicht der Fall ist, wir auch nicht auf etwas referieren können, was gar nicht existiert. Es gibt hier aber nicht zwei Himmelskörper, sondern nur einen. Wie kann man auf zwei Himmelskörper Bezug nehmen, wenn dort nur einer ist? Mir scheint aber, dass die Zurückweisung einer irreduziblen Unterscheidung zwischen Sinn und Bedeutung<sup>33</sup> nicht bedeuten kann, dass die Bezugnahme gar nicht scheitern kann. Anders gesagt: Es muss möglich sein, etwas Falsches zu meinen. Eine falsche Aussage verfehlt auch ihr Objekt, die vermeintliche Tatsache, warum soll einem Namen dies nicht möglich sein? Vor der Entdeckung der Venus meinten die Menschen, dass es verschiedene Himmelskörper waren, die morgens und abends an bestimmten Stellen am Firmament sichtbar waren. Diese irrige Voraussetzung, so möchte ich insistieren, bestimmt ihre Verwendung von „Hesperus“ und „Phosphorus“ und damit das, was sie mit diesen Ausdrücken *meinen*. Was sie damit meinen, ist nun aber nichts anderes als worauf sie sich damit

---

bei Williams (op. cit.: 80ff)). Nun könnte man natürlich meinen, dass man entsprechend auch zeigen könnte, dass der Begriff der Schmeferenz inkohärent oder aus anderen Gründen sinnlos ist, und dass deshalb diese Anwendung des Kripkeschen Manövers auch nichts einbringt. Dazu kann ich hier nur sagen, dass eine solche Argumentation prinzipiell sicherlich denkbar ist, dass ich aber nicht die leiseste Ahnung habe, wie sie aussehen sollte. Mir scheint im Gegenteil, dass überhaupt nur ein Antirealist, ein globaler Skeptiker oder Idealist, kohärent bestreiten kann, dass es eine transparente Referenzbeziehung zwischen seinen sprachlichen oder nichtsprachlichen Repräsentationen und der Welt gibt. (Und damit, dass er dies kohärent bestreiten kann, ist nur gemeint, dass ein solches Bestreiten mit seiner Position kohärent sein mag, nicht dass diese Position in sich kohärent ist.)

<sup>33</sup> Ich gehe hier davon aus, dass die Diskussion der Beispiele für den Fregeschen Sinnbegriff gezeigt hat, dass die Sinnunterschiede zwischen Beschreibungen auf Unterschiede der Referenz der Komponenten dieser Beschreibungen reduziert werden kann.

*beziehen*. Einen Gegenstand meinen heißt dasselbe wie sich auf ihn zu beziehen, oder jedenfalls werde ich diese Ausdrücke hier so verwenden.

Entsprechend muss es also einen Sinn von „beziehen“ geben, in dem man sich auf etwas beziehen kann, das nicht existiert, oder jedenfalls nicht als das existiert, als das es gemeint wird. Die Zurückweisung der Unterscheidung zwischen Sinn und Bedeutung impliziert nicht die Identifikation des gemeinten Objekts mit der Referenz oder Bedeutung des sprachlichen Ausdrucks (oder der mentalen Repräsentation). Die Bedeutung besteht neben der Fähigkeit zur Repräsentation in der Intention, sich auf einen Gegenstand zu beziehen, nicht in diesem Gegenstand selbst, und diese Intention kann scheitern, weil es den Gegenstand nicht gibt.

Der zweite Punkt ist der, dass die Menschen natürlich auch vor der Entdeckung der Venus mit ihrer Verwendung der Namen „Hesperus“ und „Phosphorus“ in irgendeinem Sinne in intentionalem Kontakt mit der Realität waren und sich damit auf diese Realität bezogen. Sie wendeten diese Namen ja auf etwas an, das sie am Himmel wahrnahmen. Wir können uns selbst dann auf die Realität beziehen, im intentionalen Kontakt mit ihr sein, wenn diese Bezugnahme, wie in diesem Fall, unauflöslich in einen Kontext irriger Voraussetzungen und Annahmen eingebettet ist. (Wäre dies nicht möglich, wären wir wohl gar nicht in der Lage, uns auf etwas in der Welt zu beziehen, da alle unsere Bezugnahmen wesentlich in ein holistisches Netzwerk von Einstellungen eingebunden sind, von denen sicherlich immer einige falsch sind.) Wie dies in dem gegenwärtigen Fall genauer zu verstehen ist, wird bald noch deutlicher werden.

Mir scheint deshalb, dass es einen sinnvollen Begriff von Referenz gibt, nach dem sich die Sprecher vor der Entdeckung der Venus mit „Hesperus“ und „Phosphorus“ nicht auf dasselbe beziehen, und dass dies tatsächlich der maßgebliche Referenzbegriff ist. Entsprechend müssen wir sagen, dass die Sprecher vor dieser Entdeckung, was diesen Punkt betrifft, ‚eine andere Sprache sprechen‘ – jedenfalls wenn wir hier von Kripkes Vorschlag zur Individuation von Sprachen ausgehen. Es ist also nicht so, dass die Sprecher mit der Entdeckung der Venus erst herausfinden, was sie mit ihren Worten meinen und gemeint haben – einen Gegenstand statt zweien, wie es ihnen schien – und welche Sprache sie sprechen, wie die externalistische Auffassung es nahe legt; die angemessene Beschreibung ist die, dass sie in Reaktion auf die Entdeckung ihre Sprache, ihr Vokabular der Situation anpassen. Mir scheint, der Schlüssel zum Verständnis von Sätzen wie „Hesperus ist Phosphorus“ liegt darin, sie als



empirisch motivierte *Redefinitionen* der in ihnen vorkommenden Begriffe anzusehen. Empirische Entdeckungen *motivieren* diese Redefinitionen und damit den Übergang zu einer neuen Sprache, ohne dass diese Sätze in der neuen Sprache doch selbst den Inhalt dieser Entdeckungen ausdrücken können. Um besser zu verstehen, was damit gemeint ist, wird es nützlich sein, den Begriff eines dynamischen *a priori* einzuführen.

#### **4.9 Empirisch motivierte Redefinitionen und das dynamische *a priori***

Das dynamische *a priori* ist nur in einem schwachem und damit vergleichsweise unproblematischem Sinn *a priori*. Strukturen, die in diesem Sinn *a priori* sind, sind weder angeboren noch haben sie einen transzendentalen Status. Der dynamische Charakter dieses *a priori* bedeutet, dass die Strukturen sowohl selbst das Resultat von Erfahrung sein, als auch wiederum durch Erfahrung verändert werden können. Im gegenwärtigen Kontext geht es dabei um die Strukturen, die durch sprachliche Kategorien einschließlich der Eigennamen konstituiert werden. Mit diesen sprachlichen Kategorien interpretieren Menschen die Wahrnehmungserfahrung. Sie sehen etwa einen Körper am Himmel und drücken mit dem Urteil „Dies ist Hesperus“ aus, dass sie diesen Körper als den *wiedererkennen*, den sie schon bei verschiedenen Gelegenheiten wahrgenommen haben. Diese Kategorien enthalten substanzielle Festlegungen darüber, was der Fall ist oder sind mit solchen verbunden: Festlegungen etwa auf die Existenz bestimmter Arten von Entitäten. Im Einklang mit Wittgensteins Entscheidung, die Gleichheit durch Gleichheit des Zeichens auszudrücken, wird der Gebrauch von zwei Eigennamen normalerweise eine Festlegung auf die Existenz zweier damit benannter Entitäten ausdrücken. In dem zur Debatte stehenden Fall drückt er eine Festlegung auf die Existenz zweier Himmelskörper aus.

Der Punkt ist nun, dass das System sprachlicher Kategorien vor der Entdeckung der Venus, das System, das die Namen „Hesperus“ und „Phosphorus“ als Namen zweier verschiedener Himmelskörper enthält, die Möglichkeit der Identität von Hesperus und Phosphorus genauso ausschließt wie unsere Sprache ihre Verschiedenheit. Wie Kripke anerkennt (114 Fn56), muss, wenn die Identität notwendig ist, die Verschiedenheit es gleichermaßen sein. Genau dieselben Argumente, die Kripke für die Notwendigkeit der Identität von Hesperus und Phosphorus anführt, können wir

entsprechend vom Standpunkt der Menschen vor ihrer Entdeckung auch für die Notwendigkeit ihrer Verschiedenheit anführen. Insbesondere können wir im gleichen Stil argumentieren, dass eine Identität von Hesperus und Phosphorus nicht denkbar ist, da wir auch das, was der Fall hätte sein können, in unserer Sprache beschreiben müssen und nicht in einer kontrafaktischen Sprache, die vielleicht andere Menschen in einer qualitativ identischen epistemischen Situation gesprochen hätten. Und in unserer Sprache beziehen „Hesperus“ und „Phosphorus“ sich ja auf zwei verschiedene Himmelskörper! Was wir also als eine Situation zu beschreiben versucht sind, in der Hesperus mit Phosphorus identisch wäre, ist tatsächlich nur eine Situation, in der ein Planet an der Stelle der beiden Himmelskörper gewesen wäre, die wir benannt haben.

Vor der Entdeckung der Venus sprechen die Menschen also eine Sprache, in der „Hesperus“ und „Phosphorus“ sich auf zwei verschiedene Entitäten beziehen. Vom Standpunkt dieser Sprache ist die Verschiedenheit von Hesperus und Phosphorus notwendig. Es gehört zur kategorialen Struktur dieser Sprache, zu ihrem Grundgerüst, dass diese beiden verschieden sind, und in dem Sinn, in dem etwas, das zur kategorialen Struktur einer Sprache gehört, *a priori* ist, ist auch das (vermeintliche) Wissen von der Verschiedenheit von Hesperus und Phosphorus *a priori*. Zu wissen, dass Hesperus von Phosphorus verschieden ist, gehört zum Bedeutungswissen für die Sprecher dieser Sprache. Jemand, der nicht verstünde, dass sie mit diesen Ausdrücken zwei verschiedene Himmelskörper meinen, würde ihre Sprache nicht verstehen. Wir haben aber ebenfalls bereits betont, dass es sich nur um ein *a priori* in einem schwachen Sinn handelt. Die kategoriale Struktur der Sprache ist selbst durch Erfahrung entstanden und spiegelt diese wieder. Sie enthält substanzielle Annahmen über die Struktur der Welt, die selbst wiederum durch die Erfahrung in Frage gestellt und widerlegt werden können. In unserem Fall stellt sich die Annahme, dass es zwei verschiedene Himmelskörper (Sterne, Planeten) sind, die morgens und abends am Firmament erscheinen, als falsch heraus.

Nach der Standarddeutung in der philosophischen Literatur, die ich bisher auch zugrunde gelegt habe, war die Reaktion darauf, Hesperus und Phosphorus zu identifizieren. Ich möchte gleich zeigen, dass dies gar nicht richtig ist, dass in der Standardsprache Hesperus und Phosphorus *de facto* gar nicht identifiziert worden sind, sondern vielmehr Hesperus und Phosphorus bzw. der Morgenstern und der Abendstern – sofern diese Ausdrücke überhaupt noch gebraucht werden – als verschiedene

*Stadien* desselben Planeten rekonzeptualisiert wurden. Und dass die Sprecher diese Entscheidung getroffen haben, ist sicherlich kein Zufall. Aber die Identifikation von Hesperus und Phosphorus ist doch eine mögliche Reaktion. Wie also ist sie zu verstehen?

Ich glaube, die Hauptschwierigkeit, die Sätze wie „Hesperus ist Phosphorus“ aufzuwerfen, liegt darin, dass einerseits ihre Akzeptanz klar durch empirische Belege begründet und motiviert ist, dass es aber andererseits unmöglich scheint zu verstehen, wie der Inhalt der entsprechenden empirischen Aussagen durch die so genannten Identitätssätze ausgedrückt werden soll – jedenfalls dann, und diese Annahme teile ich hier mit Kripke, wenn die involvierten Namen nicht mit Beschreibungen synonym sind. Es ist natürlich leicht möglich, den Inhalt der Entdeckung anzugeben, indem man etwa sagt, wir hätten entdeckt, dass es dort nur einen Himmelskörper gab statt zwei. Aber wenn die Bedeutung der Namen nur in dem Bezug besteht, den sie zu den benannten Gegenständen herstellen, ist der Identitätssatz tautologisch, auch wenn er die Form „ $a = b$ “ hat statt „ $a = a$ “. Die Bedeutung der Namen und die des Identitätszeichens zu kennen heißt zu wissen, dass er wahr ist.

So betrachtet erscheint der Identitätssatz als notwendig, aber – und darauf habe ich in Abgrenzung zu Kripke insistiert – damit natürlich auch als *a priori*. Den empirischen Charakter von „ $a = b$ “ können wir dagegen nur unter Bezug auf den Kontrast zwischen einem Begriffssystem verstehen, in dem „ $a \neq b$ “ akzeptiert wird und dem neuen Begriffssystem (oder der neuen Sprache), in dem nunmehr „ $a = b$ “ gilt. Die Verschiedenheit von  $a$  und  $b$  erscheint aber nicht als Möglichkeit innerhalb des letzteren Begriffssystems, genauso wenig wie ihre Identität in ersterem. Und es ist auch nicht so, dass der Übergang zu dem anderen Begriffssystem doch irgendwie die Möglichkeit der Identität oder Verschiedenheit von  $a$  und  $b$ , *so wie sie in dem ursprünglichen Begriffssystem gemeint wurden*, zeigt. Dies gilt, egal ob wir von einem Begriffssystem mit „ $a \neq b$ “ zu einem mit „ $a = b$ “ übergehen oder umgekehrt. (Der umgekehrte Fall ist aber die künstlichere Konstruktion, da wir normalerweise nicht zwei Namen für einen Gegenstand haben.) D.h. es zeigt nicht, dass etwa derjenige einfach Unrecht hat, der à la Kripke insistiert, es sei doch notwendig, dass  $a \neq b$ , wir müssten doch, solange wir dieselbe Sprache sprechen (wollen), auch „ $a \neq b$ “ akzeptieren. Es ist durchaus vernünftig zu sagen, dass für denjenigen, der  $a$  und  $b$  unterscheidet, der also „ $a \neq b$ “ akzeptiert und für denjenigen, der „ $a = b$ “ akzeptiert, „ $a$ “ und „ $b$ “ nicht dasselbe bedeuten, und dass sie damit insoweit eine andere Sprache sprechen statt

einfach nur in derselben Sprache verschiedene Meinungen zu äußern. Dennoch kann der Übergang von der einen zur anderen Sprache empirisch motiviert sein. Empirische Gründe können uns dazu bringen, die Bedeutung unserer sprachlichen Zeichen zu ändern, unsere Sprache der Welt anzupassen. Die Begriffe, die Sprache, die wir wählen, enthält substantielle Festlegungen in Bezug darauf, was der Fall ist. Dass ich zwei Namen für Himmelskörper einführe, wird normalerweise meine Überzeugung reflektieren, dass ich es mit zwei verschiedenen Himmelskörpern zu tun haben. Diese Überzeugung kann durch empirische Gründe in Frage gestellt werden, und dies wiederum kann ein Grund dafür sein, die sprachlichen Kategorien zu ändern. In unserem Fall besteht eine Möglichkeit – und ich betone, dass dies nur eine Möglichkeit ist unter anderen, die ich gleich diskutieren werde – darin, die Bedeutungen von „a“ und „b“ so zu ändern, dass sie nunmehr das gleiche bedeuten, also „a = b“ zu akzeptieren. Man kann sagen, dass damit „a“ und „b“ zu *einem* Namen *erklärt* werden. Sie bleiben zwar syntaktisch verschieden, aber diese Verschiedenheit hat nun keine Bedeutung mehr, so dass man sie semantisch als einen Namen ansehen kann.

Ich betone, dass sie zu einem Namen *erklärt* werden, um damit den deklarativen, definitiven, apriorischen Charakter dieses Sprechakts hervorzuheben. Es handelt sich um eine Redefinition der Namen „a“ und „b“, nicht um eine Aussage, die mit den Namen in ihrer alten Bedeutung gemacht wird. Die Signifikanz, der Witz dieser Redefinition ergibt sich aus der Tatsache, dass „a“ und „b“ vorher eine andere Bedeutung hatten und aus den empirischen Gründen für die Redefinition. Diese empirischen Gründe motivieren die Redefinition, sie erzwingen sie aber weder, noch nimmt der Sprechakt auf einen empirischen erkannten Sachverhalt Bezug. Er legt vielmehr einen anderen Begriffsrahmen zur Beschreibung empirisch erkannter Sachverhalte fest.

Mir scheint, dass diese Auffassung den Doppelcharakter von Identitätsaussagen am besten verständlich macht. In dem Maße, in dem wir den Begriffsrahmen als gegeben und fixiert annehmen, sind die Identitäts- bzw. Verschiedenheitsaussagen notwendig, aber entsprechend sind sie auch *a priori*. Sie legen einen Teil der grundlegenden Struktur dieses Rahmens fest. Gleichzeitig ist der Begriffsrahmen, den sie partiell verkörpern, durch empirische Gründe, empirisch erkannte Tatsachen motiviert. Insoweit es sich um empirisch erkannte Tatsachen handelt, sind diese aber auch nicht notwendig. Es ist nicht notwendig, dass nur ein Himmelskörper morgens und abends am Firmament auftaucht statt zweien. Das Konzept des dynamischen, durch

die Erfahrung motivierten *a priori* kann dieser Dualität Sinn abgewinnen. Kripke dagegen paart einfach mehr oder minder gewaltsam die Notwendigkeit, die die Stabilität des Begriffssystems voraussetzt, mit dem empirischen Aspekt der Identitätsaussage, der nur über die Revidierbarkeit des Begriffssystems und die Kontingenz der Annahmen verständlich ist, die seine Struktur motivieren. So kommt er zu seiner Entdeckung der Notwendigkeit *a posteriori*, die wir als ein *factum brutum* hinnehmen sollen. Ich meine hier eine adäquatere Erklärung gegeben zu haben, die die Postulierung einer irreduziblen Kategorie der Notwendigkeit *a posteriori* vermeidet.

Ein weiteres Argument für die gegebene Deutung ergibt sich aus der Tatsache, dass die Identifikation von Hesperus und Phosphorus nicht notwendig ist. Wie schon angedeutet, ist sie nur eine mögliche Reaktion auf die gemachte empirische Entdeckung und nicht einmal die tatsächliche Reaktion der Sprecher. Wenn dies richtig ist, stützt es die Auffassung, dass die Identifikation in einem wesentlichen Sinn *a priori* ist. Denn es bedeutet, dass sie nicht durch die Erfahrung erzwungen wird. Es ist der begrifflichen Souveränität und Freiheit des Sprechers anheim gestellt, wie er sich verhalten, wie er seinen Begriffsrahmen gestalten will, und dieses Merkmal ist oft mit Apriorizität verbunden worden. Gleichzeitig erläutert dies weiter, was damit gemeint ist, dass die Wahl des Begriffsrahmens empirisch motiviert ist, ohne doch einen empirisch erkennbaren Sachverhalt auszudrücken: Es gibt empirische Gründe, aber diese sind nicht *zwingend*. Sie lassen dem Sprecher die Wahl zwischen verschiedenen Begriffssystemen.

Es mag auf den ersten Blick paradox erscheinen, dass das *a priori* somit einerseits wesentlich mit Notwendigkeit verbunden sein soll, andererseits mit der Freiheit, ein bestimmtes Begriffssystem zu wählen. Aber tatsächlich liegt hier kein Widerspruch vor. Die Freiheit ist die Freiheit, ein bestimmtes Bezugssystem zu *wählen*. Die Notwendigkeit, das Apriorische besteht darin, dass dieses Bezugssystem durch die Akzeptanz bestimmter grundlegender Sätze festgelegt ist. Dass etwa Hesperus nicht Phosphorus ist, definiert – unter anderem – dieses Bezugssystem. Diese Festlegung aufzugeben bedeutet *ipso facto* ein anderes Begriffssystem, eine andere Sprache zu wählen. Der Fall ist zu dem eingangs diskutierten Beispiel des Metermaßes analog. In diesem Fall hatte der Sprecher die Freiheit, einen geeigneten Stab beliebiger Länge zum Metermaß zu erklären. Dass der gewählte Stab einen Meter lang ist, ist dann eine Festlegung *a priori*; diese Festlegung definiert eine bestimmte Längeneinheit, ein bestimmtes System der Längenmessung. Hätte der Sprecher einen länge-

ren oder kürzeren Stab gewählt, hätte er eine andere Längeneinheit, ein anderes System definiert.

Aber welche Wahl soll nun ein Sprecher im Hesperus/Phosphorus-Beispiel haben? Natürlich ist die Situation nicht in jeder Hinsicht analog, da es nicht um eine Neudefinition der involvierten Begriffe handelt. Die Begriffe haben ja bereits Bedeutung. Es handelt sich – wenn überhaupt – um eine Redefinition, um eine Anpassung des begrifflichen Rahmens an die neuen Erkenntnisse. Im Allgemeinen wird man eine solche Anpassung möglichst konservativ gestalten, also so, dass der begriffliche Rahmen und die Menge der vom Sprecher akzeptierten Sätze möglichst wenig verändert werden. Betrachten wir also genauer, durch welche Verwendungskontexte und Annahmen die Bedeutung von „Hesperus“ und „Phosphorus“ vor der Entdeckung der Venus bestimmt wird, also vor der Entdeckung, dass es ein und derselbe Himmelskörper ist, der morgens im Osten und abends im Westen am Firmament erscheint.

Da nicht viel bekannt zu sein scheint über die Umstände der Entdeckung der Venus einschließlich des astronomischen Hintergrunds, den Beobachtungen und/oder Theorieänderungen, die zu derselben geführt haben, nehme ich mir hier die Freiheit, einen solchen Hintergrund, insofern er für die hier diskutierten Fragen wesentlich ist, einfach zu erfinden. Ich denke aber, dass es sich dabei durchaus um eine realistische, wenn auch äußerst vereinfachte Fiktion handelt, nicht um eine bloße Fabrikation zur Unterstützung einer vorgefassten These.

Zwei Aspekte dieses Hintergrundes vor allem scheinen mir wesentlich zu sein:

1. Zwei *Beobachtungspraxen*, in denen in entsprechenden Wahrnehmungskontexten Äußerungen wie „Dies dort ist Hesperus (Phosphorus) gemacht werden“ oder „Heute war Hesperus sehr früh zu sehen“ gemacht werden. Dies konstituiert schon das *Unterscheiden* zwischen Hesperus und Phosphorus.

2. Verknüpfung der Namen mit einer *Typspezifikation*. Die Bezeichnungen „Abendstern“ und „Morgenstern“ legen nahe, dass Hesperus und Phosphorus seinerzeit als Sterne konzeptualisiert wurden – oder jedenfalls als etwas, das im damaligen astronomischen Denken unserem gegenwärtigen Begriff eines Sterns möglichst nahe kommt. Um die Situation etwas zu vereinfachen, werde ich aber im Folgenden stipu-

lieren, dass Hesperus und Phosphorus von vornherein als Planeten aufgefasst wurden.

Der oder die Sprecher gehen davon aus, dass Hesperus und Phosphorus verschieden sind. Stellen wir uns vor, dass diese Verschiedenheit anfangs nicht als Hypothese behandelt, sondern als selbstverständlich unterstellt wird. Dies ist sicherlich auch der Normalfall in einer Praxis der Verwendung von Namen – nicht nur von Eigennamen, sondern auch von Gattungs- oder Stoffnamen. In einem solchen Normalfall ist die angemessene Beschreibung einfach die, dass die Sprecher eine bestimmte *Unterscheidung machen* – es handelt sich um ein Strukturmoment ihres sprachlichen und sonstigen intentional diskriminierenden Handelns eher als um eine *Überzeugung*, die hier ausgedrückt wird. Dieses Machen der Unterscheidung manifestiert sich primär in den bereits beschriebenen Beobachtungspraxen. Darin, dass die Sprecher morgens urteilen „Dies ist Hesperus“ und abends „Dies ist Phosphorus“ statt wiederum „Dies ist Hesperus“, liegt bereits ein Unterscheiden. (Noch einmal, es ist richtig, das Handeln der Sprecher als ein Unterscheiden zu charakterisieren, obwohl nach unserem heutigen Wissen diese Unterscheidung in einer grundlegenden Hinsicht illusorisch ist.) Wir können uns aber auch vorstellen, dass die Sprecher mitunter dieses Unterscheiden explizit machen, eine entsprechende Überzeugung herausbilden und etwa äußern „Hesperus und Phosphorus sind verschieden“. Eine solche Äußerung könnte man etwa in einem Kontext finden, in dem ein Kind, das die Verwendung von „Hesperus“ und „Phosphorus“ in der Sprachgemeinschaft noch nicht verstanden hat, abends sagt „Guck mal, Hesperus!“. Die Antwort könnte lauten: „Nein, dies ist Phosphorus. Hesperus leuchtet zwar genauso hell wie Phosphorus, ist aber ein anderer Himmelskörper“. Den Status einer Meinung im landläufigen Sinn bekommt diese Unterscheidung aber erst, wenn sich Zweifel melden, ob man es hier tatsächlich mit zwei Himmelskörpern zu tun hat. Jemand stellt die alternative Hypothese auf, dass es sich nur um einen Himmelskörper handelt, und diese Hypothese setzt sich langsam durch.

Welche Gründe für diese Hypothese sprechen, was letztlich für diesen Wandel sorgt, brauchen wir uns hier nicht auszumalen. Wesentlich ist allein, dass ein solcher Wandel stattfindet. Uns interessiert, welche Auswirkungen er für unser Begriffssystem hat, das also wesentlich durch die irriige Annahme oder Unterstellung geprägt ist, dass wir es mit zwei verschiedenen Planeten zu tun haben. Es gibt nun im Wesentli-

chen drei Möglichkeiten, unser Begriffssystem an die neuen Erkenntnisse anzupassen.

1. Die Sprecher könnten „Hesperus“ und „Phosphorus“ nunmehr einfach als zwei bloß syntaktisch verschiedene, aber semantisch identische Namen für einen Planeten verwenden, also, im materialen Modus gesprochen, Hesperus und Phosphorus schlichtweg identifizieren.
2. Die Sprecher könnten zu dem Schluss kommen, dass es Hesperus und Phosphorus gar nicht gibt. Der Grund für diese Entscheidung wäre, dass mit „Hesperus“ und „Phosphorus“ zwei verschiedene Planeten gemeint waren, man es tatsächlich aber nur mit einem zu tun hatte. Die Tatsache, dass diese Namen mit einem so grundlegenden Irrtum wesentlich verbunden sind, dass sie diesen geradezu verkörpern, würde als ein so großer Makel angesehen, dass man sie sozusagen aus dem Verkehr ziehen muss, jedenfalls aus dem ernsthaften, Bezug nehmenden sprachlichen Verkehr, indem man die Existenz von Hesperus und Phosphorus verneint.
3. Man könnte Hesperus und Phosphorus als zwei Planeten*stadien* rekonzeptualisieren. Demnach wäre die Venus Phosphorus, wenn sie morgens im Osten und Hesperus, wenn sie abends im Westen am Firmament erscheint.

Wie schon zur Genüge bekannt, wird in der philosophischen Literatur allgemein fraglos unterstellt, dass wir die 1. Variante gewählt haben. Ich möchte nun argumentieren, dass unser tatsächliches Begriffssystem am besten so interpretiert wird, dass wir uns für eine Mischung der 2. und der 3. Variante entschieden haben. Betrachten wir etwa folgende Passage aus einer populären Enzyklopädie:

Neben Sonne und Mond ist die Venus von der Erde aus gesehen das hellste Objekt am Himmel. Sie wird allgemein auch als Morgenstern (griechisch Phosphorus, lateinisch Lucifer) bezeichnet, wenn sie vor Sonnenaufgang im Osten erscheint, und als Abendstern (griechisch Hesperus), wenn sie nach Sonnenuntergang im Westen zu sehen ist.<sup>34</sup>

---

<sup>34</sup> Vgl. Microsoft Encarta Enzyklopädie Professional 2003 (CD-ROM), Artikel „Venus“. Ich habe die in der Encarta verwendeten Schreibungen „Hesperos“ und „Phosphoros“ der in dieser Arbeit gebrauchten Schreibweise angepasst.



Heben wir zuerst das Selbstverständliche hervor und konstatieren, dass es heutzutage Standard ist, die Venus als „Venus“ zu bezeichnen. Die Ausdrücke „Morgenstern/Phosphorus“ und „Abendstern/Hesperus“ werden nur noch in speziellen Kontexten gebraucht, etwa in poetischen, historischen – oder philosophischen. „Venus“ ist der wissenschaftliche Name der Venus, in jedem Nachschlagewerk werden sich die eigentlichen Sachinformationen über den Planeten unter diesem Stichwort finden usw. usf.. In dem Maße, in dem „Venus“ die anderen Ausdrücke abgelöst hat, hat sich somit in gewissem Sinne die 2. Variante durchgesetzt. Wir würden zwar normalerweise nicht unbedingt sagen, dass Phosphorus/Morgenstern und Hesperus/Abendstern nicht existieren, aber wir haben doch etwas nah Verwandtes getan, nämlich das mit einer falschen Voraussetzung verknüpfte Vokabular außer in einigen Spezialbereichen schlichtweg aus der Sprache eliminiert.

Zweitens werden „Hesperus“ und „Phosphorus“ nunmehr als Namen für Stadien der Venus verwendet. Es soll zugestanden werden, dass die Formulierung oben noch nicht von letzter Klarheit ist. Es heißt dort, dass die Venus als „Morgenstern“ etc. bezeichnet wird, obwohl man streng genommen sagen müsste, dass ein *Stadium* der Venus als „Morgenstern“ bezeichnet wird. Wir pflegen uns normalerweise eben nicht explizit darüber Rechenschaft abzulegen, ob wir mit bestimmten Ausdrücken nun Planeten oder ihre Stadien meinen – der Begriff „Objektstadium“ ist ja nicht zufällig ein philosophischer Kunstausdruck. Was mit der Rede von Stadien gemeint ist, wird aber hier durch die Antezedentien der Konditionale ausgedrückt: Die Venus wird als „Phosphorus“ bezeichnet, wenn sie vor Sonnenaufgang erscheint – und eben nicht, wenn sie nach Sonnenuntergang zu sehen ist – dann wird sie als „Hesperus“ bezeichnet. Man könnte versucht sein, an dieser Stelle einzuwenden, dass Phosphorus eben doch mit der Venus identisch, nur nach Sonnenuntergang nicht mehr als Phosphorus bezeichnet wird. Aber warum sollte man dies sagen? Dies führt eine sehr merkwürdige Kluft ein zwischen dem, was etwas ist, und dem, als das es bezeichnet wird. Wenn dies am Abendhimmel doch Phosphorus wäre, warum sollten wir es dann nicht so bezeichnen? Mir scheint deshalb die vorgeschlagene Deutung die plausibelste Interpretation des gewöhnlichen Sprachgebrauchs zu sein – dass sie etwas klarer und expliziter als das Interpretierte ist, liegt in der Natur der Sache.

Es lässt sich auch leicht verständlich machen, warum sich dieser Sprachgebrauch durchgesetzt hat. Erstens wäre es witzlos, zwei Namen für denselben Gegenstand zu haben. Dies wäre eine stetige Quelle der Verwirrung, und es ist deshalb

wahrscheinlich, dass dieser Zustand nicht stabil wäre. Es gibt eine starke Tendenz dahin, das Unterscheidungspotential, das durch die Existenz verschiedener Ausdrücke gegeben ist, auch zu nutzen oder ansonsten den Gebrauch des überflüssigen Ausdrucks einzustellen. (Wann soll ich den diesen Ausdruck verwenden, wann den anderen? Sobald ein Unterschied in der Verwendung systematisch mit einem Merkmal des Kontextes zusammenhängt, kann sich die Bedeutung des Ausdrucks leicht dahin entwickeln, dass dieses Merkmal auch bezeichnet wird. Studien sprachlichen Bedeutungswandels bieten dafür reichlich Anschauungsmaterial.) Dies erklärt auch, warum Synonymie so selten ist – wenn es sie denn streng genommen überhaupt gibt.

Deshalb ist es eine sinnvolle Entscheidung, einerseits das Paar „Hesperus/Phosphorus“ in den wichtigsten Kontexten durch „Venus“ zu ersetzen und andererseits in den sprachlichen Nischen, in denen das veraltete Begriffssystem noch überlebt, Hesperus und Phosphorus weiterhin als verschieden zu behandeln. Und dies kann, wie gesehen, leicht geschehen, indem Hesperus und Phosphorus als Planetenstadien rekonzeptualisiert werden. So wird die Übereinstimmung mit unseren früheren Selbsten maximiert und die eingangs erhobene Forderung der Konservativität erfüllt. D.h. die Namen bleiben mit ihren ursprünglichen Beobachtungskontexten verknüpft und die meisten Sätze, in denen sie vorkommen, können weiterhin akzeptiert werden – wenn sie auch jetzt etwas anderes verstanden werden. Viele, wenn auch nicht alle der Eigenschaften, die wir zwei Planeten „Hesperus“ oder „Phosphorus“ zuschrieben, können wir nunmehr zwei Planetenstadien zuschreiben. Stellen wir uns z.B. vor, jemand habe im Zuge elementarer astronomischer Beobachtungen jeden Tag die Zeit notiert, zu der Hesperus und Phosphorus von seinem Standpunkt aus sichtbar waren. All diese Notizen könnten übernommen werden. Es würde genügen, eine Anmerkung des Inhalts zu machen, dass Hesperus und Phosphorus nunmehr als zwei Stadien desselben Planeten aufgefasst werden.

Diese Überlegungen zeigen auch, dass es wesentlich ist, dass ein Eigenname mit einer Typspezifikation verbunden ist. Die Änderung der Typspezifikation von „Planet“ zu „Planetenstadium“ erlaubt uns, Hesperus und Phosphorus weiterhin als verschieden anzusehen, viele Sätze mit diesen Namen weiterhin zu akzeptieren und damit unser Begriffssystem auf konservative, unaufwendige Weise zu ‚reparieren‘, an die Realität anzupassen. Ähnlich erwies es sich an einem bestimmten Punkt der Geschichte der Astronomie als nötig, die Venus als Planeten statt als Stern zu rekonzeptualisieren – von dieser Komplexität hatten wir ja bis jetzt in unserer Diskussion

abgesehen. Der Grund, warum ein Name mit einer Typspezifikation verbunden sein muss, ist einfach der, dass, je nachdem mit welcher Typspezifikation er verbunden ist, andere Aussagen mit diesem Namen wahr werden. (Und dies liegt wiederum einfach daran, dass er dann eine andere Entität bezeichnet. Hesperus das Planetenstadium ist offensichtlich von Hesperus dem Planeten verschieden.) Nun kann ein Akt des Benennens natürlich in mancherlei Hinsicht unbestimmt sein – und wird es auch. Es könnte etwa jemand nur so etwas gedacht oder gesagt haben wie „Diesen Himmelskörper (oder gar nur: dieses Ding da) will ich „X“ nennen!“, ohne sich darüber Rechenschaft abzulegen, ob denn jetzt z.B. ein Stern oder ein Planet gemeint sein soll, ja ohne über solche Begriffe überhaupt zu verfügen. Dies macht nichts, die Spezifikation als Himmelskörper oder auch nur als Ding ist bereits eine basale Typspezifikation, die bei Bedarf präzisiert werden kann. Je nachdem z.B., ob ich die Konzeption *dieses Gegenstands* vor mir etwa präzisiere als *dieses Möbelstück*, dieser Tisch, dieses Stück Holz, habe ich verschiedene Entitäten mit potentiell verschiedenen raumzeitlichen Grenzen herausgegriffen. Das Stück Holz etwa könnte, wenn ich mir daraus einen Stuhl baue, aufhören ein Tisch zu sein, wäre aber immer noch ein Möbelstück. Aber irgendeine Typspezifikation, sei sie auch noch so vage und implizit, muss vorhanden sein, weil sonst völlig unbestimmt ist, *was* eigentlich benannt worden ist. Es kann z.B. nicht völlig unbestimmt und willkürlich sein, was als Wiedererkennen des benannten Gegenstands zählen soll. Es kann z.B. nicht sein, dass ich beim nächsten Mal auf eine Blume zeige und sage „Da ist ja X wieder!“ – es sei denn, für mich handelt es sich dabei um zwei Instanzen einer ganz besonderen, exotischen Typspezifikation.

Bedeutet dies nun, dass wenn sich die mit einem Namen verbundene Typspezifikation ändert, wenn etwa Hesperus und Phosphorus als Planetenstadien rekonzeptualisiert werden, diese Namen ihre Bedeutung ändern? Diese Frage scheint mir sehr schwierig zu sein, und ich werde hier nicht in der Lage sein, sie vollständig zufriedenstellend zu beantworten. Aber weil sie so wichtig ist, ein paar knappe Überlegungen dazu. Zuerst möchte ich die zunächst sicher paradox klingende Behauptung aufstellen, dass Bedeutungen sich ändern, aber dennoch die selbe Bedeutung bleiben können. "Wenn dies merkwürdig erscheint," so kommentierte Hilary Putnam (1991) diese These einmal, "so liegt das daran, dass wir nicht gewohnt sind, Bedeutungen als historische Entitäten aufzufassen, in dem Sinn in dem Menschen oder Nationen historische Entitäten sind. Ich, Hilary Putnam, hatte lockiges blondes Haar, als ich klein

war. ... Jetzt habe ich ausgedünntes graues Haar. Dennoch bin ich die selbe Person." Das Bild von Bedeutungen als historischen, also irreduzibel zeitlichen Entitäten, die wie z.B. Personen ihre Identität bewahren können, obwohl sich zumindest manche ihrer Eigenschaften ändern, drängt sich aber auf, wenn wir die zeitlich ausgedehnte Praxis der Verwendung eines Eigennamens, etwa des Namens einer Person, betrachten. Wir passen nicht nur unser ‚Dossier‘ über die Person ihrer tatsächlichen Entwicklung an und erhalten neue Informationen über ihre Vergangenheit, Gegenwart und Zukunft. Auch unsere Fähigkeit zum Wiedererkennen der Person wird durch persönliche Begegnungen mit ihr oder durch Bilder, Tondokumente etc. neu kalibriert und damit aktualisiert. (Jemanden, den ich zehn Jahre nicht mehr getroffen habe, werde ich oft im ersten Moment nur mühsam wiedererkennen, danach ist meine Erinnerung, mein Wiedererkennungsvermögen aufgefrischt worden.) Schließlich sind wir auch mitunter gezwungen, falsche Meinungen über die Person zu korrigieren. Aber immer noch meinen wir dieselbe Person. Der Kontinuität der Person entspricht die Kontinuität unserer Praxis des Verfolgens der Entwicklung dieser Person und der Entwicklung unseres tatsächlichen oder vermeintlichen Wissens über sie. Wir haben uns getäuscht, wir hatten eine falsche Überzeugung, aber es war eine falsche Überzeugung über dieselbe Person, nicht über eine andere. Es scheint nun, dass wir uns im Extremfall auch darüber täuschen können, mit was für einer Art von Entität wir es zu tun hatten. So könnte ich etwa anfangen, einen Vogel zu füttern, der immer wieder auf meinen Balkon geflogen kommt, und den ich „Harry“ nenne. Harry ist eine Meise, aber in meiner ornithologischen Ignoranz halte ich ihn für einen Spatzen – bis eine Freundin mich aufklärt. Damit hat sich die Bedeutung des Namens für mich in dem Sinne geändert, das meine Konzeption des Benannten sich in einer wichtigen Hinsicht verändert hat. Aber es wäre äußerst unplausibel zu behaupten, dass ich nicht mehr die selbe Entität mit dem Namen meine. Eine wichtige Überzeugung über das Referenzobjekt hat sich geändert, aber die neue Überzeugung ist eine Überzeugung über das Objekt, das ich schon immer gemeint habe. Ähnlich verhält es sich mit der Rekonzeptualisierung eines Sterns als Planeten oder eines Planeten als Planetenstadium. Wichtig ist dabei auch, dass ich mich in all diesen Fällen, wenn eine präzisierte Typspezifikation sich als irrtümlich erweist, auf basalere Typspezifikationen zurückziehen kann: d.h. ich kann etwa sagen, ich hätte doch diesen Vogel, diesen Himmelskörper oder diese Himmelserscheinung gemeint. Es ist wesentlich, dass ein Eigenname implizit oder explizit mit einer Typspezifikation verknüpft ist, aber es

muss nicht unbedingt stets dieselbe Typspezifikation sein. Innerhalb gewisser Grenzen sind auch Typspezifikationen revidierbar.

Der im gegenwärtigen Zusammenhang entscheidende Punkt ist der, dass die Änderung der Typspezifikation einer Entität zwar einschneidend ist, aber bei weitem nicht so einschneidend wie die Identifikation mit einer Entität, von der sie unterschieden wurde. Letzterer Fall scheint nicht mit dem Fall einer Person analog zu sein, die sich durch die Zeit hinweg ändert, sondern mit dem Szenario, dass eine Person zu einer anderen Person wird, aber doch sie selbst bleibt. Dies ist aber unmöglich, und genau darin liegt die fatale Schwierigkeit des Konzepts der Identitätsaussage *a posteriori*. „Hesperus“ ist unter anderem in Opposition zu „Phosphorus“ definiert, und deshalb würde ihre Identifikation eben bedeuten, Hesperus mit etwas anderem zu identifizieren und damit etwas Unmögliches. So kann diese so genannte Identifikation nur als ein Übergang von einem Begriffssystem, in dem die beiden unterschieden werden, zu einem Begriffssystem verstanden werden, in dem diese Unterscheidung aufgehoben wird. Ein solcher Übergang ist zwar möglich, aber er bedeutet nicht die Identifikation der ursprünglich gemeinten Objekte, sondern eine Transformation in eine andere Sprache, in der die Zeichen nunmehr nur noch ein Objekt meinen, so dass sie, jedenfalls nach einer Definition von Namen, nunmehr auch nur noch zwei verschiedene syntaktische Instanzen desselben Namens sind.

Ich werde später versuchen, das erarbeitete einfache Modell der Revision eines Begriffssystems auch auf den Fall der Revision der Opposition zwischen Körper und Geist zu übertragen. Im nächsten Kapitel wird es nun vorerst darum gehen, die Resultate dieses Kapitels auf theoretische Identifikationen einschließlich der psychophysischen Identitätstheorie anzuwenden. Ich werde verschiedene Versuche analysieren, dem Argument gegen empirisch informative Identitätsaussagen zu entgehen und zeigen, dass sie allesamt scheitern.

## 5 Die Identität des Bewusstseins

*Everything is what it is, and not another thing*

Joseph Butler, Sermons

Im Einklang mit der Hauptargumentation dieser Arbeit wollen wir uns nun der psychophysischen Identitätstheorie im Kontext anderer so genannter wissenschaftlicher oder theoretischer Identifikationen wie z.B. „Wasser ist H<sub>2</sub>O“, „Wärme ist Molekularbewegung“ oder „Licht ist ein Strom von Photonen“ zuwenden, wobei wir mit letzteren beginnen. Nach dem prinzipiellen Argument gegen empirisch informative Identitätssätze scheint bereits klar, dass auch der Versuch, die wissenschaftlichen Identifikationen als genuin empirische Aussagen zu verstehen, zum Scheitern verurteilt ist. Wir wollen aber die Argumente und begrifflichen Konstruktionen der Identitätstheoretiker nicht einfach ignorieren. Sie zu analysieren wird unser Verständnis sowohl von Identitätsaussagen, als auch das der materialistischen Metaphysik, die den wissenschaftlichen Identifikationen zugrunde liegt, vertiefen. Dabei werden wir uns vorerst weiter an Kripkes Darstellung orientieren. Dies scheint angemessen, da Kripkes Konzeption von theoretischen Identifikationen als notwendig *a posteriori* sehr einflussreich gewesen ist und die Debatte bis auf den heutigen Tag prägt. Gleichzeitig hat Kripke ein viel diskutiertes Argument *gegen* die psychophysische Identitätstheorie vorgelegt. Ich werde nun zuerst zeigen, dass die wissenschaftlichen Identifikationen in Wirklichkeit Eliminationen sind und dann, dass Kripkes Argument gegen die psychophysische Identitätstheorie darauf beruht, dass er sich – vernünftigerweise – weigert, das Bewusstsein auch noch zu eliminieren. Darauf folgt eine Analyse der verschiedenen Versuche der Identitätstheoretiker, eine Identifikation von Bewusstseinsentitäten dennoch verständlich zu machen.

### 5.1 Wissenschaftliche Identifikationen als notwendig *a posteriori*

Kripke erläutert seine Konzeption in der folgenden Passage:

We have ... discovered about light that light is a stream of photons; alternatively it is a form of electromagnetic radiation. Originally we identified light by the characteristic internal visual impressions it can produce in us, that make us able to see. Heat, on the

other hand, we originally identified by the characteristic effect on one aspect of our nerve endings or our sense of touch. (129)

Wir haben nach Kripkes Meinung also herausgefunden, dass Licht mit Photonenströmen identisch ist. Wie ist es möglich, dass es sich dabei um eine genuine Entdeckung *a posteriori* handelte? Die wesentliche Voraussetzung von Kripkes Konzeption ist seine Auffassung der Semantik solcher Begriffe wie „Licht“ und „Wärme“: ursprünglich haben wir demnach Licht als die Ursache „charakteristischer innerer visueller Eindrücke“ identifiziert und Wärme als die Ursache unter anderem von taktilen Eindrücken. Dies eröffnet anscheinend die Möglichkeit, nunmehr empirisch herauszufinden, was diese Ursache ist: Ströme von Photonen oder Molekularbewegung. Aber bedeutet dies nicht, dass wir einen kontingenten Sachverhalt herausgefunden haben? Könnte nicht auch etwas anderes als Photonenströme unsere visuellen Eindrücke verursachen? Könnten nicht auch umgekehrt Photonenströme existieren, ohne visuelle Eindrücke zu verursachen? Wäre das nicht ein Fall, in dem sie nicht mit Licht identisch wären? Gegen diesen nahe liegenden Einwand verteidigt Kripke sich so:

Imagine a situation in which human beings were blind or their eyes didn't work. They were unaffected by light. Would that have been a situation in which light did not exist? It seems to me that it would not. It would have been a situation in which our eyes were not sensitive to light. ... So it seems to me that such a situation would be a situation in which there was light, but people could not see it. So, though we may identify light by the characteristic visual impression it produces in us, this seems to be a good example of fixing a reference. We fix what light is by the fact that it is whatever, out in the world, affects our eyes in a certain way. ... (129f)

...we are talking about this phenomenon in the world, and not using 'light' as a phrase *synonymous* with 'whatever gives us the visual impression – whatever helps us to see'; for there might have been light and it not helped us to see; and even something else might have helped us to see. The way we identified light *fixed a reference*. (130)

Die Antwort ist also folgende. Mit „Licht“ meinen wir stets das – was immer es auch sein mag –, das in der (tatsächlichen) Welt unsere Augen in bestimmter Weise affiziert und charakteristische visuelle Eindrücke verursacht. Anders ausgedrückt: „Licht“ bezieht sich stets auf das Denotat der definiten Beschreibung „Das x, so dass gilt: x verursacht (tatsächlich) diese charakteristischen visuelle Eindrücke“. Deswegen zeigt die Existenz bloß möglicher Welten, in denen etwas anderes diese visuellen

Eindrücken verursacht, oder in denen es diese visuellen Eindrücke gar nicht gibt, nicht, dass die Identitätsaussage „Licht = Photonenströme“ nicht notwendig wahr ist. Denn „Licht“ bezeichnet eben nur die *tatsächliche* Ursache, und wenn Photonenströme diese Ursache sind, also Photonenströme.

Kripke sagt, dass die Beschreibung die Bezugnahme von „Licht“ festlegt (*fixes the reference*), aber nicht mit ihr synonym ist. Allerdings ist diese Behauptung Kripkes bestenfalls insofern plausibel, als sie sich auf die Beschreibung ohne das oben eingeklammerte „tatsächlich“ bezieht. Wenn wir das „tatsächlich“ – im modallogischen Kontext oft als so genannter Aktualitätsoperator interpretiert – hinzufügen, erhalten wir eine Beschreibung, die nur den Gegenstand herausgreift, der in der tatsächlichen Welt die Beschreibung erfüllt, und dies entspricht genau Kripkes Charakterisierung der Bedeutung von „Licht“. Es ist nicht zu sehen, welche darüber hinausgehende Bedeutung „Licht“ nach Kripkes Auffassung noch haben soll.

Eine Beschreibung, deren Bezugnahme durch Hinzufügung eines Ausdrucks wie „tatsächlich“ starr auf ihr Denotat in der tatsächlichen Welt festgelegt wird, wird oft auch als ‚rigidifizierte‘ (*rigidified*) Beschreibung bezeichnet (vgl. z.B. Chalmers 1996), da eine solche Beschreibung ein starrer Designator im Sinne Kripkes ist, also ‚in allen möglichen Welten‘ denselben Gegenstand bezeichnet. Demnach können wir also sagen, dass nach Kripkes Auffassung Ausdrücke für natürliche Arten wie z.B. „Licht“, „Wärme“, „Wasser“ und „Gold“ mit starren Beschreibungen eines bestimmten Typs synonym sind. Anders ausgedrückt: Kripke behandelt diese Ausdrücke als *Abkürzungen* dieser Beschreibungen, ähnlich wie Frege, Russell und andere Eigennamen als Abkürzungen von Beschreibungen aufgefasst haben. Ein Unterschied besteht, wenn überhaupt, nur darin, dass die von Kripke verwendeten Beschreibungen starr sind.<sup>35</sup> Ob dieser Unterschied in dieser Form tatsächlich besteht, scheint mir sehr fraglich zu sein, denn auch Frege und Russell verstanden definite Beschreibungen

---

<sup>35</sup> Diese Aussage mag überraschen, da Kripke doch der große Kritiker der Beschreibungstheorie ist, und tatsächlich beschreibt er selbst seine Position nicht so und kritisiert (im erst nach der Erstveröffentlichung von *Naming and Necessity* entstandenen Vorwort zur Buchausgabe) auch eine entsprechende Behandlung von Eigennamen. Was Termini für natürliche Arten angeht, scheinen mir die oben zitierten Passagen im Text schon eindeutig zu sein, im folgenden Haupttext werden aber auch noch weitere Belege angeführt. Eine ähnliche Interpretation von Kripke mit den technischen Mitteln der so genannten zweidimensionalen Semantik findet sich bei Chalmers (1996). Auch Searle (1983) vertritt die These, dass Kripkes und Putnams Semantik von Begriffen für natürliche Arten nicht so verschieden von der Beschreibungstheorie ist wie diese Autoren behaupten, sondern und letztlich nur eine Variante derselben darstellt. Im Gegensatz zu diesen Autoren geht es mir aber hier nicht darum, eine modifizierte Version der Beschreibungstheorie zu verteidigen. Die Stoßrichtung meiner Kritik an Kripke ist vielmehr die, dass er in seiner Zurückweisung der Beschreibungstheorie nicht konsequent genug ist.



gen natürlich so, dass sie diejenigen bezeichnen, die sie *tatsächlich* erfüllen. Eine ausführlichere Diskussion dieses Punktes würde uns aber zu modallogischen Fragen führen, die für die Hauptargumentation dieser Arbeit nebensächlich sind. Ich werde sie mir deshalb hier ersparen und die Vorstellung starrer Beschreibung *for the sake of argument* akzeptieren.

Ein sicheres Anzeichen dafür, dass hier die opake Form der Bezugnahme vorliegt, als die wir das Denotieren von Beschreibungen charakterisiert haben, ist die Präsenz solcher Phrasen wie „was auch immer“ in Kripkes Spezifikation der Bedeutung von „Licht“. Wir beziehen uns damit auf die Ursache der charakteristischen visuellen Eindrücke, *was auch immer* diese Ursache sein mag. (Mit „der König von Frankreich“ meine ich den König von Frankreich, *wer auch immer* König von Frankreich sein mag.) Es handelt sich also, Russells Analyse entsprechend, um einen Existenzsatz. Wir glauben, dass es etwas gibt, auf das die Beschreibung zutrifft, aber was dieses etwas an sich ist, wissen wir nicht, es ist uns als solches unbekannt. Damit handelt es sich also nicht um genuine Referenz, denn diese setzt ja Bekanntschaft mit dem Bezugsobjekt voraus. In unserem alltäglichen Verständnis solcher Ausdrücke wie „Licht“ und „Wärme“ gehen wir davon aus, dass wir durch die Wahrnehmung Zugang zu diesen Phänomenen haben. Wir können sie sehen und fühlen. Von diesem Verständnis ausgehend, macht es keinen Sinn, so etwas zu sagen wie: „Ich meine mit „Licht“ Licht, was immer das auch sein mag“. Kripke beraubt also „Licht“ seines gewöhnlichen Sinnes, weil die Beschreibung, mit der er es gleichsetzt, unsere visuelle Erfahrung als bloße Wirkung äußerer, uns als solcher unbekannter Phänomene behandelt, nicht als etwas, das uns als solches Zugang zur Realität gibt. Dies hat auch wieder zur Folge, dass die Sprecher zumindest vor der Entdeckung der Ursache ihrer visuellen Eindrücke nicht wissen, welche Sprache sie sprechen. Sie sprechen eine Sprache, deren Ausdrücke sich z.B. auf Photonenströme, auf Molekularbewegung oder auf H<sub>2</sub>O-Moleküle beziehen, ohne doch die mindeste Ahnung von diesen Dingen zu haben. Sie wissen buchstäblich nicht, wovon sie reden. Dies scheint mir wiederum keine akzeptable Konzeption des Sprechens einer Sprache und von Referenz zu sein. Genuine Referenz muss transparent sein, d.h. sie muss uns einen intentionalen Zugang zu einer Entität als solcher eröffnen. Das heißt es, sie zu meinen, oder sich auf sie zu beziehen.

Dass die gegebene Interpretation von Kripkes Position richtig ist, zeigt sich auch in der folgenden, zusammenfassenden Passage Kripkes:

What characteristically goes on in these cases of, let's say, 'heat is molecular motion'? There is a certain referent which we have fixed, for the real world and for all possible worlds, by a contingent property of it, namely the property that it's able to produce such and such sensations in us. Let's say it's a contingent property of heat that it produces such and such sensations in people. It's after all contingent that there should ever have been people on this planet at all. So one doesn't know *a priori* what physical phenomenon, described in other terms – in basic terms of physical theory – is the phenomenon which produces these sensations. We don't know this, and we've discovered eventually that this phenomenon is in fact molecular motion. When we have discovered this, we've discovered an identification which gives us an essential property of this phenomenon. We have discovered a phenomenon which in all possible worlds will be molecular motion – which could not have failed to be molecular motion, because that's what the phenomenon *is*. (132f)

Wir suchen also nach etwas, das die *kontingente* Eigenschaft hat, bestimmte visuelle Eindrücke zu produzieren, wir suchen nach der Ursache dieser Eindrücke. Dies bedeutet aber nicht mehr und nicht weniger als dass wir – im formalen Modus gesprochen – nach etwas suchen, auf das die Beschreibung „x verursacht (tatsächlich) bestimmte Empfindungen“ zutrifft. Entsprechend ist der Gegenstand unserer Entdeckung, dass die Bewegung von Molekülen diese Empfindungen verursacht. In der Notation des Prädikatenkalküls können wir dies wie folgt ausdrücken. „Wärme“ kürzt die Beschreibung (1) ab, wobei ich im Folgenden zur Vereinfachung die Einzigkeitsbedingung weglasse. (2) drückt dann die Identifikation der Ursache als Molekularbewegung (M) aus. Dieser Satz ist aber logisch äquivalent mit (3). (3) ist offensichtlich ein prädikativer Satz und kommt ohne Identitätszeichen aus.

- (1)  $(\exists x) x$  verursacht (tatsächlich) bestimmte Empfindungen
- (2)  $(\exists x) x$  verursacht (tatsächlich) bestimmte Empfindungen &  $x = M$
- (3) M verursacht (tatsächlich) bestimmte Empfindungen

(3) identifiziert die entdeckte Tatsache. Also folgt aus Kripkes eigener Darstellung, dass der Gegenstand unserer Entdeckung eine kontingente Tatsache, auf die wir mit einem prädikativen Satz bezugnehmen – nicht mit einem Identitätssatz. Dass eine solche Tatsache *a posteriori* erkannt wird, ist natürlich unproblematisch, aber der Vorstellung, dass eine Instanz der Identitätsbeziehung *a posteriori* erkannt werden

kann, ist immer noch kein Sinn gegeben worden. Ein weiterer scheinbar informativer Identitätssatz hat sich als prädikativer Satz herausgestellt. Kripke kommt der Satz wie ein Identitätssatz vor, weil er die Namen natürlicher Arten nur implizit mit Beschreibungen gleichsetzt und sich selbst darüber nicht klar ist. Dabei mag helfen, dass in unserem gewöhnlichen Verständnis diese Namen tatsächlich nicht mit Beschreibungen synonym sind. Aber dieses Verständnis beruht eben darauf, dass wir uns normalerweise einen von Beschreibungen der von Kripke und anderen verwendeten Art unabhängigen Zugang zu diesen Phänomenen zuschreiben: Wir können sie wahrnehmen. Dieses Verständnis entleert Kripke seines Sinns, in dem er die Wahrnehmungserfahrung als bloße Wirkung der äußeren Realität behandelt, nicht als eigenständigen Zugang zu ihr. Damit werden die Namen mit bloßen Beschreibungen synonym. Weil Kripke sich darüber aber nicht im Klaren ist und sie weiterhin als Namen behandelt und weil er richtig sieht, dass ein Identitätssatz mit Namen sich nicht auf eine kontingente Tatsache beziehen kann, scheint es ihm, dass wir in den so genannten wissenschaftlichen Identifikationen *a posteriori* eine notwendige Tatsache, eine Instanz der Identitätsrelation erkannt haben. Er erliegt damit einer Illusion der Notwendigkeit. Eine solche, nicht eine Illusion der Kontingenz, liegt hier in Wirklichkeit vor.

## 5.2 Die Illusion der Notwendigkeit

Ich möchte diesen Punkt noch kurz verstärken, indem ich zeige, dass sich Kripkes Behandlung der wissenschaftlichen Identifikationen völlig mit seiner Interpretation des Meterbeispiels parallelisieren lässt, das er ja als eine Instanz des kontingenten *a priori* aufgefasst hatte. Der entsprechende Satz, der sich als *a priori* verstehen lässt, lautet wie folgt:

- (1) Licht verursacht diese charakteristischen visuellen Eindrücke.

Dies gilt natürlich nur dann, wenn wir den Satz in dem im letzten Kapitel ausführlich beschriebenen Sinn als Definition, als deklarative Bedeutungsfestlegung *a priori* verstehen, also so, dass er etwa wie folgt glossiert werden kann:

- (2) Dasjenige, was (tatsächlich) diese bestimmten visuellen Eindrücke verursacht, dieses  $x$ , was immer es auch sei, nenne ich hiermit „Licht“!

So verstanden ist der Satz (1) *a priori* in dem Sinn, in dem auch der Satz „Dies ist 1 m lang“, bezogen auf das Urmeter, als *a priori* interpretiert werden kann. In keinem dieser beiden Fälle liegt Wissen von einer kontingenten Tatsache vor. Indem ich mich entschieße, eine Länge „Meter“ nenne, erfahre ich genau sowenig eine kontingente Tatsache über sie, wie ich über die – mir ansonsten unbekannte – Ursache bestimmter meiner Eindrücke etwas erfahre, indem ich mich darauf festlege, sie als „Licht“ zu bezeichnen. Betrachten wir nun die mit dem Satz (3) gemachte Aussage:

- (3) Photonenströme verursachen diese bestimmten visuellen Erfahrungen.

Diese Aussage ist zweifelsfrei eine Aussage *a posteriori*. Sie wird aber, wie wir gesehen haben, auch von einer kontingenten Tatsache wahrgemacht. Es ist nicht notwendig, dass Photonenströme die Ursache der charakteristischen visuellen Eindrücke sind. Die Ursache könnte auch eine andere sein. Notwendig ist nur in einem bestimmten Sinne, dass wir nun Photonenströme als „Licht“ bezeichnen. Es ist notwendig, weil wir uns mit (1) in der Interpretation (2) darauf festgelegt haben. Wir haben uns nach Kripkes Deutung *a priori* darauf festgelegt, die Ursache bestimmter Empfindungen als „Licht“ zu bezeichnen. Wir wissen nun ferner *a posteriori*, was diese Eindrücke verursacht, nämlich Photonenströme. Damit sind wir darauf festgelegt, Photonenströme als Licht zu bezeichnen. Und dies heißt wiederum, dass wir darauf festgelegt sind, den Satz „Licht = Photonenströme“ zu akzeptieren, denn dieser Satz drückt nur im materialen Modus aus, was „Licht“ bezeichnet Photonenströme“ im formalen Modus ausdrückt.

Was ergibt sich nun aus diesen Überlegungen für die Notwendigkeit dieses Satzes? Insofern „Licht“ hier nur als Abkürzung der Beschreibung fungiert, die „Photonenströme“ denotiert, bedeutet der Satz nur soviel wie unser Satz (3). Insofern der Satz aber auch die deklarative definitonische Festlegung darauf widerspiegelt, Licht als Abkürzung für die Beschreibung zu verwenden, hat er einen apriorischen Aspekt. Diesem Aspekt korrespondiert aber keine außersprachliche Tatsache in der Welt, keine kontingente und schon gar keine notwendige. Die Erklärung, einen Aus-

druck als Abkürzung einer Beschreibung zu verwenden, ist weder mit der Erkenntnis einer kontingenten, noch der einer notwendigen Tatsache verbunden. Kripkes Argumentation erreicht eine gewisse Scheinbarkeit nur dadurch, dass er diese beiden Aspekte in verwirrender Weise vermengt bzw. gar nicht als solche erkennt und unterscheidet. So missversteht er das *a priori* der definitiven Deklaration völlig als die vermeintliche Notwendigkeit einer offensichtlich kontingenten Tatsache.

Warum behandelt Kripke das Meterbeispiel so anders als die theoretischen Identifikationen, obwohl diese Fälle doch nach der hier gegebenen Argumentation so wesentliche Parallelen aufweisen? Ich glaube, dass eine Kombination der folgenden Faktoren dafür verantwortlich ist. Erstens ist beim Meterbeispiel explizit eine definite Beschreibung involviert ( $1\text{m} = \text{die Länge des Stabs zum Zeitpunkt } t$ ), während Kripke Termini wie „Licht“, „Wärme“ und „Wasser“ nur implizit mit (starr) Beschreibungen gleichsetzt. Zweitens sind bei den theoretischen Identifikationen tatsächlich empirische Entdeckungen im Spiel, bei der bloßen Festlegung der Längeneinheit dagegen nicht. (Um auch hier Parallelität herzustellen, genügt es aber, einen Satz wie „ $1\text{m} = 39.37 \text{ Zoll}$ “ zu betrachten – vorausgesetzt, dieser drückt keine weitere definitive Festlegung aus, sondern berichtet das Resultat einer Messung mit einem unabhängigen Maßstab. Diesen Satz müsste man nach Kripkes Maßstäben auch als notwendig *a posteriori* auffassen. Auf den Einwand, dass der Stab doch auch länger oder kürzer hätte sein können, würde man analog zu Kripkes Argumentation bei den wissenschaftlichen Identifikationen antworten, dass man mit „Meter“ eben die *tatsächliche* Länge des Stabes meint. D.h. man würde die Bedeutung dieses Ausdrucks mit der rigidifizierten Version der definitiven Beschreibung gleichsetzen. Der Satz würde also Wissen *a posteriori* von einer kontingenten Tatsache ausdrücken, gleichzeitig würde das Moment der definitiven Festlegung des Begriffs „Meter“ in seine Bedeutung hineinspielen. Wir kennen nun die Länge dieses Stabes, haben Wissen *a posteriori* darüber erlangt; gleichzeitig haben wir uns *a priori* darauf festgelegt, diese Länge, welche es auch immer sei, als „Meter“ zu bezeichnen.) Der dritte und sicherlich wichtigste Grund ist aber der, dass Kripke von der Tradition der physikalistischen Metaphysik geprägt ist, der zufolge die Physik das *Wesen* zumindest der äußeren Realität erfasst. Vor dem Hintergrund dieses Essentialismus werden die wissenschaftlichen Entdeckungen als metaphysische Einsichten in das Wesen der äußeren Realität verstanden. Das Wesen von Wasser, seine Identität, besteht darin,  $\text{H}_2\text{O}$  zu sein, das von Wärme, Molekularbewegung zu sein usw.. Dagegen würde

wohl niemand ohne weiteres auf die Idee verfallen, das Wesen des Meters bestehe darin, 39.37 Zoll lang zu sein.

Wir haben schon eingangs herausgearbeitet, dass diese Metaphysik die manifeste äußere Welt eliminiert. Die Diskussion von Kripkes Konzeption der so genannten wissenschaftlichen Identifikationen bestätigt dies. Die Identifikationen erscheinen nur deshalb als möglich, weil die Wahrnehmung, die uns nach unserem gewöhnlichen Selbstverständnis einen eigenständigen Zugang zur Realität gibt, stillschweigend als bloße Wirkung der physikalischen Welt rekonzeptualisiert wird. Damit werden die Begriffe manifester physischer Phänomene ihres ursprünglichen Sinns entleert, denn dieser war wesentlich durch die Verbindung mit der Wahrnehmung bestimmt. Wir meinen mit „Licht“ und „Wärme“ uns durch die Wahrnehmung bekannte Phänomene, nicht irgend etwas, was immer es auch sei, das die bloße Ursache der entsprechenden Bewusstseinszustände ist. Kripkes Interpretation dieser Begriffe beinhaltet also bereits die Elimination der manifesten physischen Welt.

Weiter hatten wir schon eingangs gesehen, dass uns vor diesem Hintergrund nur die Wahl bleibt, entweder das Bewusstsein auch noch zu eliminieren, was als unverständlich erscheint, weil wir das, was ich als den „qualitativen Überschuss“ der manifesten Welt gegenüber der physikalischen bezeichnet hatte, dann nirgendwo mehr unterbringen können, oder dem Bewusstsein einen metaphysischen Sonderstatus zu geben. Kripke folgt dieser zweiten Strategie. Aus seinen Überlegungen zu den vermeintlichen Identifikationen manifester physischer Entitäten entwickelt er ein Argument gegen die psychophysische Identitätstheorie, dem ich mich nun zuwende.

### **5.3 Kripkes Argument gegen die psychophysische Identitätstheorie**

Auf der Grundlage unserer bisherigen Überlegungen zu Identitätsaussagen im Allgemeinen und zu Kripkes Deutung derselben im Besonderen lässt sich leicht eine Anleitung zur Konstruktion eines solchen Argumentes erstellen. Wir müssen nur im Fall von Bewusstseinsentitäten ihre Elimination mittels der Redefinition ihrer Begriffe als Abkürzungen von Beschreibungen, etwa Beschreibungen ihrer kausalen Rolle, zurückweisen – vernünftigerweise natürlich – und Bewusstseinsbegriffe stattdessen als direkt referierende Ausdrücke behandeln. Es wird uns dann leicht fallen, eine wesentliche Asymmetrie zwischen den (vermeintlichen) Identifikationen der makrophysischen Phänomene mit ihren mikrophysikalischen Korrelaten und der Identifika-

tion von Bewusstseinsentitäten mit ihren neuronalen Korrelaten nachzuweisen. Dann wird offenkundig sein, dass das nicht gelingt, was vorher auch nur scheinbar gelungen ist, nämlich die empirische Informativität der in Frage stehenden Aussagen mit der Notwendigkeit der Identitätsbeziehung zu versöhnen.

Kripkes Argumentation entspricht genau dieser Beschreibung. Die Asymmetrie, auf die Kripke verweist, betrifft das von ihm so genannte Wegerklären des Anscheins oder der Illusion der Kontingenz (vgl. z.B.: 150f). Dass uns theoretische Identifikationen wie die von Wärme und Molekularbewegung als kontingent erscheinen, ist dadurch zu erklären, dass wir dazu neigen, ein Szenario zu betrachten, in dem etwas anderes als Molekularbewegung unsere Wärmeempfindung verursacht hätte und dieses fälschlich als einen Fall zu beschreiben, in dem Wärme etwas anderes als Molekularbewegung gewesen wäre. Eine entsprechende Erklärung ist aber bei psychophysischen Identitätsaussagen, z.B. bei einer Identifikation von Schmerzen mit einem bestimmten Typ von Gehirnzuständen nicht möglich, und dies liegt an einem wesentlichen Unterschied in der Funktionsweise der involvierten Begriffe:

In the case of the identity of heat with molecular motion the important consideration was that although 'heat' is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us the sensation S. It is thus possible that a phenomenon should have been rigidly designated in the same way as a phenomenon of heat, with its reference also picked out by means of the sensation S, without that phenomenon being heat and therefore without its being molecular motion. Pain, on the other hand, is not picked out by one of its accidental properties; rather it is being picked out by the property of being pain itself, by its immediate phenomenological property. Thus pain, unlike heat, is not only rigidly designated by 'pain' but the reference of the designator is determined by an essential property of the referent. Thus it is not possible to say that although pain is necessarily identical with a certain physical state, a certain phenomenon can be picked out in the same way we pick out pain without being correlated with that physical state. If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon *is* pain. (153)

Was Kripke hier als den Gegensatz zwischen dem Herausgreifen über eine kontingente und dem Herausgreifen über eine essenzielle Eigenschaft charakterisiert, haben wir als den Gegensatz zwischen Denotation und (direkter) Referenz beschrieben. Dass Kripke *de facto* makrophysische Begriffe als Abkürzungen von Beschreibungen behandelt, wurde bereits gezeigt. Dass Schmerzen über die Eigenschaft, Schmerzen

zu sein, herausgegriffen werden, ist bestenfalls eine etwas merkwürdige Umschreibung dafür, dass wir auf Schmerzen eben direkt *als Schmerzen*, als einen bestimmten, uns bekannten Bewusstseinszustand referieren. (Schmerzen haben nicht die *Eigenschaft*, Schmerzen zu sein, sondern sind eben einfach als solche Schmerzen.) Wie immer die Differenz aber auch beschrieben wird, das Resultat ist, dass der Anschein der Kontingenz des Verhältnisses von Bewusstseins- und Gehirnzuständen bestehen bleibt, nicht als Illusion wegerklärt werden kann, oder jedenfalls nicht auf analoge Weise. Wir werden aber später noch andere Versuche kennen lernen, den Anschein als Illusion zu entlarven.

Weiterhin bestätigt diese Passage noch einmal, was wir oben bereits herausgearbeitet haben, nämlich dass Kripkes Illusion der Kontingenz tatsächlich eine der Notwendigkeit ist. Die Relation zwischen Molekularbewegung und Wärmeempfindungen ist nach Kripkes eigenem Bekunden kontingent: Es ist eine akzidentelle Eigenschaft der ersteren, dass sie letztere verursacht. Nur diese kontingente Relation kann aber als Gegenstand der relevanten Erkenntnis *a posteriori* konstruiert werden. Darüber hinaus bleibt nur die apriorische Festlegung darauf, die Ursache der Wärmeempfindungen, was immer diese auch sein mag, als „Wärme“ zu bezeichnen. Diese apriorische Festlegung kann, solange sie nicht als solche erkannt wird, eine Illusion der Notwendigkeit erzeugen.

Die Kritik an Kripkes Darstellung theoretischer Identifikationen zusammenfassend und pointierend, können wir sagen, dass ihr Grundgebrehen darin besteht, mit der Opposition zwischen starren und nichtstarren Designatoren den eigentlich wesentlichen Gegensatz zwischen Benennen und Beschreiben, Referenz und Denotation zu verdecken. Dieser Gegensatz ist deshalb so wesentlich, weil, wie wir im Anschluss an Russell argumentiert haben, das Verständnis von Namen Bekanntschaft mit dem Benannten voraussetzt, das einer Beschreibung, selbst einer starren Beschreibung, aber nicht. Referenz ist transparent, Denotation opak. Man muss, um die Beschreibung zu verstehen, nicht wissen, wer oder was sie erfüllt, sie gibt uns einen Weg zu einem Gegenstand, nicht diesen selbst. Deshalb kann auch die Entdeckung dieses Gegenstandes in unproblematischer Weise als eine Erkenntnis *a posteriori* aufgefasst werden. Es ist die Erkenntnis, dass eine Entität eine bestimmte Eigenschaft hat, die Eigenschaft, auf die die Beschreibung referiert, während sie die Entität selbst nur denotiert. Gegenstand dieser Erkenntnis ist die Tatsache, dass die Entität diese Eigenschaft hat, nicht die angebliche Identitätsbeziehung zwischen dieser Entität und



ihr selbst. Dies gilt egal ob die Entität diese Eigenschaft als einzige hat oder nicht und egal, ob die Beschreibung nur die Entität denotiert, die tatsächlich die Beschreibung erfüllt oder nicht, also ob sie rigidifiziert ist oder nicht.

Dass also die vermeintliche Identitätsaussage zwischen einem Namen und einer Beschreibung ihrer logischen Tiefenstruktur nach in Wirklichkeit nur ein ganz normaler prädikativer Satz ist, übersieht Kripke, weil er den entscheidenden Gegensatz zwischen Referenz und Denotation durch den zwischen einem starren Designator, der in allen möglichen Welten denselben Gegenstand bezeichnet und einem nichtstarken Designator, für den das nicht gilt, verdeckt. Diese Konfusion wird dadurch befördert, dass er die Namen manifester makrophysischer Entitäten wie „Licht“, „Wärme“, „Wasser“, „Gold“ usw. nur implizit und entgegen seiner ‚offiziellen‘ Theorie mit (rigidifizierten) Beschreibungen gleichsetzt – er meint sie tatsächlich als genuine Namen zu behandeln. So scheint es ihm, dass z.B. „Wärme ist Molekularbewegung“ eine notwendige Identitätsaussage *a posteriori* ist, während wir gezeigt haben, dass „Wärme“ in Kripkes Deutung die Beschreibung „das x, so dass x (tatsächlich) Wärmempfindungen verursacht“ abkürzt, so dass, insofern hier eine Aussage *a posteriori* vorliegt, deren Gegenstand statt einer Instanz der Identitätsbeziehung die Tatsache ist, dass Molekularbewegung Wärmeempfindungen verursacht. Aber darüber hinaus spielt hier noch das Moment der definitiven, deklarativen Festlegung *a priori* hinein, „Wärme“ als Abkürzung der Beschreibung bzw. als Name für das zu verwenden, was immer es auch sei, das die Beschreibung erfüllt. Dieses Moment ist, wie gezeigt wurde, für die Illusion der Notwendigkeit der Aussage verantwortlich.

Ferner haben wir klargestellt, dass die von Kripke angenommene Deutung der Namen manifester makrophysischer Entitäten nicht das ursprüngliche, naive, den entsprechenden wissenschaftlichen Entdeckungen vorhergehende Verständnis dieser Begriffe ist, sondern vielmehr eines, das auf einer durch diese Entdeckungen motivierten metaphysischen Haltung beruht bzw. diese zum Ausdruck bringt. Seit der Neuzeit hat die Hauptströmung der westlichen Philosophie den naiven Wahrnehmungsrealismus verabschiedet; Kripkes Deutung bringt dies zum Ausdruck, indem die manifesten makrophysischen Entitäten als bloße Ursachen von Bewusstseinszuständen behandelt werden. Damit erweist sich die scheinbare Identifikation dieser manifesten Entitäten in Wirklichkeit als ihre Elimination aus einer physikalistischen Ontologie. Nach der naiven Auffassung dagegen sind die relevanten Bewusstseins-

zustände transparent: Sie eröffnen uns einen intentionalen Zugang zur Welt, zu den entsprechenden manifesten Entitäten, wir nehmen diese als solche wahr und sind somit mit ihnen bekannt. Deshalb funktionieren nach der naiven, alltäglichen Deutung die Ausdrücke für manifeste makrophysische Entitäten auch tatsächlich wie genuine Namen. Sie benennen primär das, was uns durch die Wahrnehmung bekannt ist, statt nur etwas als uns unbekannte Ursache bestimmter Empfindungen zu umschreiben.

Schließlich haben wir gesehen, dass Kripke ein Argument gegen die psychophysische Identitätstheorie konstruieren konnte, indem er sich vernünftigerweise weigerte, entsprechend eine Elimination der Bewusstseinsentitäten zu akzeptieren, etwa mittels einer analogen Redefinition der Bewusstseinsbegriffe als Beschreibungen, z.B. nach Art des Funktionalismus. Es war dann leicht, eine grundlegende Asymmetrie zwischen psychophysischen und anderen theoretischen Identifikationen nachzuweisen.

Kripkes Konzeption notwendiger Identitätsaussagen *a posteriori* ist äußerst einflussreich gewesen, aber nur manche Philosophen haben dieses Argument akzeptiert (z.B. Searle 1992 und Chalmers 1996). Viele haben stattdessen zwar die These akzeptiert, dass die (vermeintlichen) theoretischen Identifikationen notwendig *a posteriori* sind, aber sich bemüht zu zeigen, dass Kripkes Auffassung dieser Aussagen entgegen seiner Meinung doch auch auf den Leib-Seele-Fall übertragbar ist. Wenden wir uns nun diesen Versuchen und damit der Debatte über die psychophysische Identitätstheorie zu.

#### **5.4 Kritik der psychophysischen Identitätstheorie**

Eine Theorie widerlegen kann man nur vor dem Hintergrund gemeinsamer, von Vertretern und Kritikern der Theorie geteilten Annahmen. Die nachstehenden Überlegungen gehen davon aus, dass die folgenden Annahmen als geteilt unterstellt werden können. Erstens, dass die Identitätstheorie sich substanziell vom Eliminativismus unterscheiden soll. Zweitens, dass die in Frage stehenden psychophysischen Identitätsaussagen empirisch informativ sind. Drittens, dass die relevanten, durch Identitätsaussagen miteinander verbundenen Begriffe, trotz gleicher Bezugnahme verschiedene Bedeutung in einem bestimmten, zu erläuterndem Sinn von Bedeutung

haben. Viertens, dass diese Bedeutungsunterschiede *kognitive* Differenzen darstellen, oder mit solchen zumindest wesentlich verknüpft sind.

Rechtfertigen wir kurz die Unterstellung dieser Annahmen, die auch als Adäquatheitsbedingungen für die Identitätstheorie aufgefasst werden können. Natürlich ist der Eliminativismus sowohl historisch als auch systematisch eng mit der Identitätstheorie verwandt. Mitunter ist er als ihre „Verschwindensform“ bezeichnet worden (Rorty 1965: 189). Aber dennoch sind diese beiden Theorien verschieden, jedenfalls ihrem Anspruch nach. Dies zeigt sich schon darin, dass der Eliminativismus manchmal als die angemessene Konsequenz aus dem diagnostizierten Scheitern der Identitätstheorie verstanden und propagiert worden ist (z.B. Feyerabend 1963). Die Attraktion der Identitätstheorie besteht aber gerade darin, dass sie die Existenz des Bewusstseins, statt sie zu leugnen, mit einem radikalen ontologischen Physikalismus vereinbar zu machen verspricht. Aus dieser Perspektive ist ihre Tendenz, in den Eliminativismus abzukippen, die wir schon in Bezug auf die anderen theoretischen Identifikationen bemerkt haben, ein fundamentales Problem, wenn nicht gar das fundamentale Problem der Identitätstheorie. Schließlich haben wir uns mit dem Eliminativismus bereits separat auseinandergesetzt und werden später noch einmal zu ihm zurückkehren.

Die zweite dritte und vierte Annahme sind sehr eng miteinander verwandt und können daher in einem Aufwasch behandelt werden. Unter natürlichen Annahmen ist die Bedeutungsverschiedenheit der involvierten Begriffe eine Folgerung aus der empirischen Informativität der Aussagen. Und dass es sich um kognitive Differenzen handeln muss, ergibt sich daraus, dass die relevanten Identitätsaussagen Erkenntnisse ausdrücken sollen. Dass Bewusstseins- mit Gehirnentitäten identisch sind, soll ja eine empirische Erkenntnis sein. Die klassischen Formulierungen der Identitätstheorie<sup>36</sup> enthalten all diese Annahmen. Die Identitätstheorie wurde von Beginn ihrer Geschichte an als Alternative bzw. Ergänzung zu Theorien vorgeschlagen, die eine begriffliche Analyse von Bewusstseinsbegriffen bzw. von mentalistischen Begriffen vornehmen. Die ursprüngliche Idee war, eine behavioristische Analyse intentionaler Zustände durch eine Identitätsbehauptung bezüglich der ‚Qualia‘, also der sensorischen Aspekte des Bewusstseins zu ergänzen (z.B. Smart 1959). Da es sich

---

<sup>36</sup> Vgl. z.B. Place 1956, Smart 1959, Feigl 1958 und Armstrong 1968. Ich führe Place 1956 auf, weil er oft als das Gründungsdokument der modernen Identitätstheorie gesehen wird (z.B. auch von Smart 1958), aber bei genauer Lektüre zeigt sich, dass, wie Place auch selbst sagt (z.B. Place 1999: 229), in diesem Aufsatz die Beziehung zwischen Bewusstsein und Gehirn tatsächlich als Kompositionsbeziehung interpretiert wird.

bei der begrifflichen Analyse intentionaler Zustände auch letztlich um eine begriffliche Reduktion oder definatorische Identifikation von intentionalen Zuständen mit Verhaltensdispositionen handelt, kann man noch genauer auch sagen, dass die Identitätstheoretiker eine begrifflich begründete, auf Analytizität und damit nach einer weithin akzeptierten Deutung auf Synonymie basierende Identitätsbehauptung, durch eine empirisch gestützte, nichtbegriffliche Identitätsbehauptung *a posteriori* ergänzt wollten. Damit ist klar, dass in letzterem Fall keine Synonymiebeziehung zwischen den involvierten Begriffen besteht. Ferner hat die Diskussion von Kripke bereits gezeigt, dass eine Explikation der Identitätstheorie, die nicht in irgendeiner Form einen Bedeutungsunterschied der involvierten Begriffe annimmt, um ihren empirischen Charakter zu erklären, wohl aussichtslos sein dürfte. Entsprechend postulieren auch die jüngeren Verteidigungen der Identitätstheorie, wie wir im Einzelnen noch sehen werden, Bedeutungsunterschiede der einen oder anderen Form zwischen Bewusstseins- und Gehirnbegriffen. Schließlich hat sicherlich nie ein Identitätstheoretiker bestritten, dass die psychophysischen Identitätsaussagen Erkenntnisse ausdrücken sollen.

Die allgemeine Diskussion von Identitätsaussagen sollte bereits klargemacht haben, worin die grundlegende Schwierigkeit der Identitätstheorie liegt: Insofern die vermeintlichen Identitätsaussagen informativ waren, erwiesen sie sich letztlich als ihrer logischen Tiefenstruktur nach als prädikative Aussagen, durch die einer Entität eine oder mehrere Eigenschaften zugeschrieben werden. Oder es stellte sich heraus, dass die vermeintlichen Identitätsaussagen in Wirklichkeit deklarative Redefinitionen von Begriffen sind, die zwar durch empirische Erkenntnisse motiviert sind, aber selbst keinen empirischen Gehalt transportieren. Schließlich wurden durch diese Redefinitionen die von den Begriffen in ihrer ursprünglichen Bedeutung – also in ihrer Bedeutung vor der Redefinition – gemeinten Entitäten eliminiert statt identifiziert.

Die Schwierigkeit, die vier Adäquatheitsbedingungen gleichzeitig zu erfüllen, kann man sich auch durch folgende Gegenüberstellung der Identitätstheorie und des Eliminativismus klarmachen. Die Identitätstheorie verspricht, die vom Eliminativismus gelegnete Realität des Bewusstseins zu wahren. Dies scheint auf den ersten Blick eine fundamentale und sehr wichtige *ontologische* Differenz zwischen diesen Theorien zu sein. Bei näherer Betrachtung wird aber unklar, worin diese eigentlich bestehen soll. Der Eliminativist behauptet, so wollen wir annehmen, dass in Wirklichkeit nur die mikrophysikalischen Entitäten existieren, der Identitätstheoretiker,

dass alles, was überhaupt existiert, mit mikrophysikalischen Entitäten identisch ist. (Ich betrachte hier nur diese globale Formulierungen der eliminativistischen und identitätstheoretischen These. Für lokale Beschränkungen der Thesen können leicht entsprechende Varianten des nun folgenden Arguments formuliert werden.) Demnach ist aber für beide die mikrophysikalische Beschreibung der Welt *ontologisch vollständig*. Worin soll also die Differenz zwischen ihren Auffassungen bestehen? Es scheint nun, dass diese Differenz gar keine Differenz in Bezug darauf sein kann, wie die außersprachliche Welt beschaffen ist, was der Fall ist, sondern nur in Bezug auf die Sprache. Der Identitätstheoretiker glaubt im Gegensatz zum Eliminativisten, dass auch manche Ausdrücke, die nicht zum mikrophysikalischen Vokabular gehören, erfolgreich auf mikrophysikalische Entitäten bezugnehmen. Weiterhin kann die in Frage stehende sprachliche Differenz keine bloß syntaktische Differenz, keine bloße Differenz der Zeichen sein, denn nach der zweiten Adäquatheitsbedingung muss ja die empirische Informativität der Aussage erklärt werden, und dies bedeutet nach der dritten Adäquatheitsbedingung, dass die Zeichen verschiedene Bedeutung haben müssen. Damit ist der Identitätstheoretiker aber auf die Vorstellung von *bloß subjektiven* Differenzen der Bedeutung festgelegt, also auf Unterschiede in der Bedeutung von Ausdrücken, die nicht in Differenzen im Bezug auf objektiv verschiedene Entitäten bestehen oder auf solche zurückführbar sind.<sup>37</sup> Anders ausgedrückt: Es muss nach dieser Vorstellung möglich sein, dass etwa ein Wissenschaftler eine Entdeckung macht, etwas, wie wir sagen würden, Neues herausfindet, ohne dass in irgendeinem Sinne eine bis dato unbekannte Entität in der Welt entdeckt wird. Er würde etwa entdecken, dass „Schmerz = Gehirnzustand X“ – wobei X hier durch einen Ausdruck ersetzt werden müsste, der sich auf das neuronale Korrelat von Schmerz bezieht – ohne etwas Neues in der Welt zu entdecken, selbst dann, wenn er vorher bereits zwar über den Begriff des Schmerzes verfügte, aber nicht nur nicht wusste, dass Gehirnzustand X mit Schmerz korreliert ist, sondern auch nicht – so würden wir wenigstens normalerweise den Fall beschreiben – dass Gehirnzustand X überhaupt existiert.

Normalerweise würden wir etwa sagen: „Wissenschaftler Y wusste vorher nichts von der Existenz des Gehirnzustands X, er hat ihn erst bei der Suche nach dem neuronalen Korrelat von Schmerz entdeckt. Er entwickelte bei dieser Suche eine neue Methode zur Untersuchung des Gehirns und entdeckte damit einen bisher unbe-

---

<sup>37</sup> Einen ähnlichen Punkt macht McGinn 2002.

kannten Typ neuronaler Oszillation. Dieser stellte sich als das neuronale Korrelat des Schmerzes heraus.“ Der Identitätstheoretiker dagegen muss den Fall so beschreiben, dass nicht eine Entität in der Welt entdeckt wurde, sondern nur eine neue Weise, sich auf dieselbe Entität zu beziehen, auf die wir uns mit dem Begriff „Schmerz“ schon immer bezogen haben. Wir verfügen jetzt nur über einen neuen Zugang zu dieser Entität.

Ich hoffe, dem Identitätstheoretiker nicht Unrecht zu tun, wenn ich sage, dass seine These doch äußerst merkwürdig und künstlich zu sein scheint. Die Vorstellung, eine neue Erkenntnis zu erwerben, eine Entdeckung zu machen, ohne doch etwas Neues zu entdecken, wirkt inhärent mysteriös. Wie kann man dennoch meinen, dass sie sinnvoll ist? Auf der Grundlage unserer allgemeinen Diskussion von Identitätsaussagen können wir annehmen, dass sie nur dann als sinnvoll erscheinen kann, wenn die Bezugnahme eines oder beider involvierter Begriffe implizit oder explizit als eine opake Form der Bezugnahme konstruiert wird. Wir würden demnach etwa mit „Schmerz“ auf opake Weise auf einen Gehirnzustand Bezug nehmen, also so, dass sich uns die vermeintliche Identität des Bezugsobjekts mit Gehirnzustand X nicht eröffnet. Mit „Gehirnzustand X“ dagegen würden wir vermutlich auf transparente oder Weise auf den Gehirnzustand Bezug nehmen – auf den Schmerz jedoch wohl wieder auf opake Weise. Die Auseinandersetzung verschiebt sich dann auf die Frage, ob dieses oder ein ähnliches Konzept der opaken Bezugnahme sinnvoll ist. Wir haben bereits im letzten Kapitel gegen die Vorstellung der opaken Bezugnahme argumentiert und werden diese Argumentation in diesem Kapitel fortführen.

Im letzten Kapitel haben wir auch bereits zwei verschiedene Konzeptionen opaker Bezugnahme kennen gelernt. Einmal die opake Bezugnahme durch eine Beschreibung im Sinne Russells. Diese Bezugnahme ist opak, weil das Verstehen der Beschreibung damit vereinbar ist, dass wir das Objekt, das sie erfüllt und das somit von der Beschreibung denotiert wird, nicht kennen. Die andere Konzeption opaker Bezugnahme wurde durch Kripkes Auffassung des Hesperus/Phosphorus-Falls exemplifiziert. Hier handelt es sich um Namen, die sich direkt auf ihren Gegenstand beziehen sollen, nicht als das Denotat einer Beschreibung. Dennoch ist auch diese Form der Bezugnahme opak, weil nach Kripkes Darstellung die Namen schon vor der Entdeckung der Venus dasselbe Objekt bezeichnen, obwohl die Sprecher meinen, sich damit auf Verschiedenes zu beziehen. Wir werden sehen, dass alle Versuche zur

Verteidigung der Identitätstheorie die Bezugnahme eines oder beider involvierter Begriffe auf die eine oder die andere Weise als opake Bezugnahme konstruieren.

Beginnen wir mit den Versuchen, die die Bezugnahme als die einer oder mehrerer Beschreibungen konstruieren.

## **5.5 1. Verteidigungsstrategie: opake Bezugnahme durch Beschreibungen**

Obwohl es wohl fast so viele Vorschläge zur Lösung des Problems der Informativität von Identitätsaussagen gibt wie Identitätstheoretiker, scheinen doch alle eine im weitesten Sinne Fregesche Strategie zu verfolgen. D.h. der Identitätstheoretiker beruft sich darauf, dass die verschiedenen Begriffe jeweils mit verschiedenen Arten des Gegebenseins verbunden sind, also dem, was Frege als „Sinn“ bezeichnete, wobei eine Fülle weiterer Bezeichnungen benutzt werden, z.B. „Zugänge“, „Gegebenheitsweisen“, „modes of presentation“, „modes of access“ und „ways of knowing“. Ich werde im Folgenden meist von „Zugängen“ sprechen.

Dabei ist es sicherlich keine Frage, dass Bewusstseins- und Gehirnbegriffe in irgendeinem Sinn mit verschiedenen Zugangsweisen verbunden sind. Jeder wird zustimmen, dass wir von Bewusstseinsentitäten auf andere Weise wissen als von Gehirnentitäten. Das Problem besteht ja umgekehrt gerade darin, die Verschiedenheit der Zugänge mit einer Identität des Gemeinten vereinbar zu machen. Nun haben wir bei der Diskussion des Fregeschen Sinnbegriffs bereits gesehen, dass nach einer natürlichen Deutung dieses Begriffs und der Beispiele, die in diesem Zusammenhang üblicherweise verwendet werden, Unterschiede in der Art des Gegebenseins wesentlich mit unterschiedlichen Eigenschaften der jeweils gemeinten Gegenstände verbunden sind.<sup>38</sup> Wir erinnern uns an Freges Beispiel, dass derselbe Punkt durch die Eigenschaft, Schnittpunkt der Geraden  $a$  und  $b$  und durch die Eigenschaft, Schnittpunkt der Geraden  $b$  und  $c$  zu sein, herausgegriffen werden kann. Schließlich haben wir im Anschluss an Russell argumentiert, dass die über eine Eigenschaft vermittelte Bezugnahme als die Denotation einer Beschreibung in der Analyse Russells zu ver-

---

<sup>38</sup> Heutzutage werden Gegebenheitsweisen oft sogar direkt mit den jeweils relevanten Eigenschaften identifiziert (vgl. z.B. McGinn 2001, McLaughlin 2001, Papineau 2002). Mir scheint es zwar angemessener zu sein, den Sinn nicht unmittelbar mit der relevanten Eigenschaft zu identifizieren, sondern mit einer mit dem Begriff verbundenen Repräsentation dieser Eigenschaft, aber da daran für die folgende Debatte nichts hängt, werde ich mich mitunter auch dieser Terminologie bedienen und der Einfachheit halber so reden, als sei der Sinn mit der in Frage stehenden Eigenschaft identisch.

stehen ist und dass die Beschreibung auf diese Eigenschaft referiert. Das heißt wir meinen den Gegenstand als das  $x$ , das die Beschreibung erfüllt. (Von der Einzigkeitsbedingung sehe ich hier und im Folgenden wieder ab, da sie für das Argument keine Rolle spielt.)

Es ist aber klar, dass der Fregesche Sinnbegriff in dieser Interpretation nicht herangezogen werden kann, um das Konzept bloß subjektiv verschiedener Zugänge zu stützen, auf das die Identitätstheorie nach dem oben gegebenen Argument festgelegt ist. Wenn wir den Sinnbegriff in natürlicher Weise anwenden, wird das Identifikationsproblem nur auf die Ebene der Zugänge verschoben. Denn die natürlichste Interpretation ist sicherlich die, dass sich Bewusstseinsbegriffe über Bewusstseinsbeziehungen auf ihren Gegenstand beziehen und physische Begriffe über entsprechende physische Eigenschaften. Demnach würde die Verschiedenheit der Objekte, auf die zugegriffen wird, nur durch die Verschiedenheit der Eigenschaften ersetzt, die die verschiedenen Zugriffe vermitteln. Die Grundschwierigkeit ist wieder die altbekannte: was zuerst wie eine informative Identitätsaussage aussieht, entpuppt sich bei näherem Hinsehen als eine Aussage, in der einem Objekt eine Eigenschaft zugeschrieben wird oder, wie in diesem Fall, zwei. Es wird von ein und derselben Entität  $a$  gesagt, dass sie etwa die Eigenschaft  $B$  hat und die Eigenschaft  $G$ . Aber  $a$  ist sowohl von  $B$  verschieden als auch von  $G$  und die beiden Eigenschaften wiederum voneinander. Entsprechend sind die involvierten Tatsachen, also dass  $a$   $B$  ist und dass  $a$   $G$  ist, ebenfalls objektiv verschieden.<sup>39</sup> Der Himmelskörper hat sowohl die Eigenschaft, am Morgen im Osten am Firmament aufzutauchen als auch die Eigenschaft, am Abend im Westen aufzutauchen und dies sind zwei verschiedene Tatsachen über ihn. Die Person kann ich sowohl durch ihre charakteristische Stimme herausgreifen als auch durch ihren Gang. Dass sie diese Stimme hat, ist eine andere Tatsache, als dass sie diesen Gang hat.

Entsprechend würde im psychophysischen Fall bestenfalls eine Verschiedenheit psychischer und physischer Substanzen oder Ereignisse durch eine Verschiedenheit psychischer und physischer Eigenschaften ersetzt. Dies wird den Identitätstheoretiker aber nicht zufrieden stellen. Er möchte auch die Eigenschaften miteinander

---

<sup>39</sup> Eine Konzeption von nicht objektiv, sondern nur rein subjektiv, durch assoziierte Gegebenheitsweisen verschiedenen Tatsachen, findet sich bei Lycan 1996. Demnach könnten etwa in einem Sinn von Tatsache – dem, der die Gegebenheitsweisen einschließt – „Dies ist Wasser“ und „Dies ist  $H_2O$ “ sich auf verschiedene Tatsachen beziehen, obwohl die Referenzobjekte in der Welt identisch sind. Lycans Konzeption ist ein Beispiel für die Vorstellung bloß subjektiver Differenzen der Gegebenheitsweise, um die es hier geht.



identifizieren. Entsprechend ist von frühen Kritikern der Identitätstheorie oft eingewendet worden, dass diese bestenfalls zu einer Zurückweisung des Substanzdualismus führt oder zu einer Identifikation von Bewusstseinsereignissen mit Gehirnereignissen, eine Identität der jeweiligen Eigenschaften aber nicht verständlich machen kann.<sup>40</sup> Die bekannteste Fassung dieser Kritik ist der Max Black zugeschriebene Einwand Nr. 3 in J.J.C. Smarts Aufsatz „Sensations and brain processes“ (1959), der kanonischen frühen Formulierung der Identitätstheorie:

Even if Objections 1 and 2 do not prove that sensations are something over and above brain processes, they do prove that the qualities of sensations are something over and above the qualities of brain processes. That is, it may be possible to get out of asserting the existence of irreducibly psychic processes, but not out of asserting the existence of irreducibly psychic *properties*. For suppose we identify the Morning Star with the Evening Star. Then there must be some properties which logically imply that of being the Morning Star, and quite distinct properties which entail that of being the Evening Star. Again, there must be some properties (for example, that of being a yellow flash) which are logically distinct from those in the physicalist story. (Smart 1959: 59)

Smarts Reaktion auf diesen Einwand bestand darin, Bewusstseinstermine in einer Begrifflichkeit zu analysieren, die er als „gegenstandsneutral“ (*topic-neutral*) bezeichnete. Die Aussage, ein oranges Nachbild erlebt zu haben, wird entsprechend wie folgt paraphrasiert:

*There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange“* (op. cit.: 60; Kursivierung von Smart)

Die Neutralität dieser Analyse besteht nach Smart darin, dass im kursivierten Teil nur gesagt wird, dass dort etwas vor sich geht, das von der gleichen Art ist („is like“) wie etwas, das unter bestimmten spezifizierten Umständen vor sich geht. Aber es wird nicht angegeben, um welche Art es sich handelt. Wir könnten die Formulierung auch wie folgt erweitern: „Es geht dort etwas von einer bestimmten Art vor; von der gleichen Art, *welche Art es auch immer sein mag*, geht auch unter den und den Umständen etwas vor.“ Die Angemessenheit der Wendung „welche Art es auch immer

---

<sup>40</sup> Vgl. z.B. Shaffer 1961 und Stevenson 1960. Shaffer bezeichnete, obwohl er für die Nichtidentität der Eigenschaften argumentiert, seine Position dennoch als eine Variante der Identitätstheorie. Dies scheint aber nicht mehr als eine terminologische Idiosynkrasie zu sein.

sein mag“ ist wieder ein sicherer Indikator dafür, dass hier eine opake Form der Bezugnahme vorliegt. Smart konstruiert die Bedeutung von Bewusstseinsbegriffen ganz explizit so, dass sie definite Beschreibungen nach der Russellschen Analyse abkürzen: Es gibt dort ein Vorkommnis einer Art, die eine bestimmte Beschreibung erfüllt, aber welche Art dies ist, bleibt unspezifiziert. Die Form der Bezugnahme ist Denotation, nicht Referenz.

Smarts Analyse ist in dieser Form noch nicht geeignet, Bewusstseinsbegriffe vollständig zu eliminieren – sein Analysans enthält z.B. noch den Begriff des Sehens –, aber sie inspirierte eine Variante des Funktionalismus, die die Identitätstheorie durch eine durchgängige funktionalistische Analyse aller mentalistischen Begriffe zu stützen versucht (z.B. Lewis 1966, Armstrong 1968). Demnach werden mentale Entitäten durch Beschreibungen ihre kausalen Beziehungen untereinander und zu ‚Inputs‘ und ‚Outputs‘ spezifiziert. So genannte Ramsey-Sätze sollen es dabei möglich machen, sukzessive alle Vorkommnisse mentaler Begriffe zu eliminieren.<sup>41</sup> Damit wird das ganze Bewusstseinsvokabular durch Existenzsätze ersetzt, die mentale Entitäten denotieren sollen. Die in diesen Existenzsätze vorkommenden Prädikate referieren nur auf nicht-mentale, physische Entitäten, auf die äußeren Ursachen und Wirkungen geistiger Phänomene. Dann wäre es nach dieser Konzeption die Aufgabe der empirischen Wissenschaft, herauszufinden, welche physiologischen oder anderen physischen Entitäten die jeweiligen kausalen Rollen spielen, also von den Beschreibungen denotiert werden. Wir könnten also dann *a posteriori* herausfinden, welche physischen Entitäten mit mentalen Entitäten identisch sind. Wir wüssten dann von welcher Art die Vorkommnisse sind – von einer physischen. Das Problem informativer Identitätsaussagen wäre anscheinend gelöst.

Es ist aber offensichtlich, dass es sich hier um eine verkappte Form des Eliminativismus handelt. Nach dem gewöhnlichen Verständnis von Bewusstseinsbegriffen meinen wir mit diesen nicht irgendwelche, uns ansonsten unbekannte Phänomene, die bestimmte kausale Rollen innehaben, sondern wir meinen diese Phänomene zu kennen, zu wissen, um welche Art von Phänomenen es sich handelt. Deshalb scheint es auch denkbar zu sein, dass etwas anderes als Schmerz die kausale Rolle von Schmerz spielen könnte, während nach der funktionalistischen Analyse diese Vorstellung sinnlos sein muss.

---

<sup>41</sup> Der Originalaufsatz ist Ramsey 1931. Zu den technischen Details vgl. auch Field 1980.

Smart begründet die Adäquatheit seiner ‚gegenstandsneutralen‘ Analyse durch ein Argument, das für die materialistische Denkweise so charakteristisch ist, dass es kurz zu betrachten sich lohnt. Er verweist darauf, dass doch z.B. die Berichte eines ungebildeten griechischen Bauern über seine Empfindungen neutral sein müssten zwischen einer dualistischen und seiner materialistischen Metaphysik (op. cit.: 60). Dazu ist nun zuerst zu anmerken, dass Überlegungen dieser Art Smart und andere Identitätstheoretiker offensichtlich nicht darin hindern, die Empfindungsbegriffe von Menschen, die noch nie in ihrem Leben etwas von neuronalen Entitäten gehört haben, so zu interpretieren, dass sie sich auf diese beziehen! Da ist es sicherlich vergleichsweise noch plausibler, diesen Menschen eine Anlage zur dualistischen Metaphysik zuzuschreiben. Noch bemerkenswerter ist aber, dass Smart implizit die bloße Tatsache, dass Menschen geistige und physische Phänomene gemeinhin voneinander *unterscheiden* und damit mit den entsprechenden Begriffen verschiedenes meinen, mit einer dualistischen Metaphysik gleichsetzt. Dies zeigt die Tendenz in der Leib-Seele-Debatte zu denken, dass man zwischen Monismus und Dualismus zu wählen hat. Weil er zu Recht meint, dass der Dualismus als metaphysische Theorie inadäquat ist, fühlt der Monist sich dazu gedrängt, äußerst unplausible begriffliche Reduktionen und Konstruktionen zu akzeptieren, und umgekehrt meinen diejenigen, die die mangelnde Plausibilität dieser Konstruktionen erkennen, deshalb Dualisten werden zu müssen. Deshalb besteht der Schlüssel zur Lösung oder Auflösung des Leib-Seele-Problems darin zu sehen, dass die Unterscheidung zwischen physischen Phänomenen und Bewusstseinsphänomenen selbst neutral ist gegenüber den metaphysischen Theorien des Dualismus und Monismus. Dagegen greifen die Identitätstheoretiker, weil sie die bloße Unterscheidung mit dem Dualismus gleichsetzen oder meinen, dass letzterer aus erster unweigerlich folgt, zu dem verzweifelten Mittel, zu behaupten, das mentale Vokabular, das diese Unterscheidung doch verkörpert, sei selbst ihr gegenüber neutral, spezifiziere seine Entitäten nur auf eine ‚neutrale‘ Weise.

Die ‚gegenstandsneutrale‘ funktionalistische Analyse mentalistischer Begriffe entspricht genau Kripkes Redefinition makrophysischer Begriffe. Ein Terminus, der seinem ursprünglichem Sinn nach auf ein uns bekanntes Phänomen referiert, wird als Beschreibung redefiniert, deren prädikativer Teil auf die kausale Rolle dieses Phänomens referiert, den Träger dieser Rolle aber nur als ein unbekanntes bloßes *x* denotiert. Dies soll es ermöglichen, das Mikrokorrelat dieser Makroentität als den Trä-

ger dieser Rolle zu identifizieren. Die Makroentität selbst dagegen wird stillschweigend eliminiert, was dadurch verdeckt wird, dass, anders als die redlicheren Versionen des Eliminativismus es vorsehen, das entsprechende Vokabular beibehalten wird. Es wird nur seiner ursprünglichen Bedeutung entleert. Zur Elimination von physischen und geistigen Makroentitäten werden also die gleichen intellektuellen Konstruktionen gebraucht. (Entsprechend ließe sich etwa auch der Dispositionalismus in Bezug auf Farben treffend als „Farbbehaviorismus“ zu bezeichnen.) Und entsprechend ist, wie schon angedeutet wurde, und wie wir später noch genauer sehen werden, die gleichzeitige Aufhebung beider Eliminationen der einzige Weg zu einer befriedigenden Lösung oder Auflösung des Leib-Seele-Problems jenseits von Monismus und Dualismus zu kommen.

Die Diskussion der Verteidigungen der Identitätstheorie, die den Sinn des Bewusstseinsvokabular als den von Beschreibungen verstehen, die ihre Gegenstände durch Bezug auf bestimmte ihrer Eigenschaften herausgreifen und so das Problem der Informativität dieser Aussagen lösen wollen, ist hiermit abgeschlossen. Entweder sind diese Eigenschaften selbst Bewusstseins-eigenschaften, womit das eigentliche Ziel der Identitätstheorie nicht erreicht und stattdessen die Frage aufgeworfen wird, was denn der Träger dieser Eigenschaften an sich selbst eigentlich ist, oder es handelt sich um nicht-mentale Eigenschaften wie die, eine bestimmte kausale Rolle zu haben, womit es sich um eine verkappte Variante des Eliminativismus handelt. Letzteres wird nach der extensiven Kritik am Funktionalismus in den letzten Jahrzehnten auch von Vertretern der Identitätstheorie zunehmend so gesehen, zumindest insofern es um die sensorischen Aspekte des Bewusstseins geht. Die folgende Aussage von John Perry kann als repräsentativ für diese Entwicklung gelten:

Smart's strategy was to appeal to topic-neutral concepts of physical properties. The prevailing application of that strategy has been to find this neutrality in causal-role or functional properties that a brain state might have. I have argued that this version of the topic-neutral strategy is a mistake when applied to the phenomenal side of the mind, to subjective characters, qualia, raw feels, that is, to experience. We need instead the neutrality of demonstrative/recognitional concepts. (205)

Statt also einzusehen, dass die Vorstellung, die Bewusstseinsbegriffe seien ‚neutral‘, verfehlt ist, bemühen viele Identitätstheoretiker sich nunmehr, die Vorstellung einer neutralen beschreibenden Bezugnahme durch die einer direkten, aber dennoch neut-

ralen Referenz zu ersetzen, für die die Bezugnahme von demonstrativen oder Rekognitionsbegriffen als Modell dienen soll. Indem sie so versuchen, die direkte Bezugnahme als opak zu konstruieren, wählen sie damit die zweite der oben unterschiedenen Hauptstrategien zur Lösung des Informativitätsproblems.

## **5.6 2. Verteidigungsstrategie: direkte opake Bezugnahme**

Bewusstseinsbegriffe werden jetzt also nicht mehr als verkappte Beschreibungen aufgefasst, sondern als direkt referierende Ausdrücke. In der Literatur finden sich eine Reihe verschiedener Versionen dieser Strategie. Sie werden an demonstrative oder allgemein an indexikalische Begriffe (Ismael 1999, O’Dea 2002) assimiliert, an Rekognitionsbegriffe (Loar 1997, Carruthers 2003 Tye 2003) oder als Mischformen dieser Typen von Begriffen aufgefasst (Perry 2002, McLaughlin 2002). Eine weitere Variante, die so genannte Zitattheorie („quotational account“; Papineau 2002, Block im Erscheinen) besagt, dass Bewusstseinsentitäten selber eine Konstituente der jeweiligen mentalen Repräsentationen sind.<sup>42</sup> Beim Bezug auf Schmerzen soll etwa eine Instanz von Schmerzen diese Bezugnahme wesentlich mitvermitteln. Manche Autoren begnügen sich schließlich zur Erklärung der Verschiedenheit des Sinns von Bewusstseinsbegriffen und physischen Begriffen mit dem Hinweis, dass diese verschiedene begriffliche Rollen in unserem Denken haben (Hill 1997, Lycan 1996).

Letzteres stellt aber keine wirklich eigenständige Auffassung dar. Dass diese Begriffe in irgendeinem Sinne verschiedene begriffliche Rollen einnehmen, wird niemand bestreiten wollen. Die Aufgabe besteht ja darin, diese Verschiedenheit der begrifflichen Rollen so zu explizieren, dass sie mit einer Identität des Gemeinten vereinbar wird, und dies erfordert, wie wir gesehen haben, die Vorstellung einer ‚neutralen‘, opaken Form der Bezugnahme. Der bloße Hinweis auf die Verschiedenheit der begrifflichen Rollen leistet dazu nichts. Schließlich sind alle normalen Fälle der Verschiedenheit begrifflicher Rollen auch mit einer Verschiedenheit der jeweiligen Bezugnahme verbunden.

Die Zitattheorie krankt daran, dass es schwer ist zu verstehen, was damit gemeint sein soll, dass die Bewusstseinsentitäten Konstituenten der jeweiligen Begriffe sind. Sie sind ihre Bezugsobjekte. Was leistet es, zu sagen, sie seien Teile derselben? Die Zitattheorie scheint vor allem durch die Beobachtungen inspiriert zu sein, dass

---

<sup>42</sup> Diese Taxonomie verschiedener Positionen ist an die von Chalmers (unveröffentlicht) angelehnt.

eine vollgültige, eigenständige Bezugnahme auf Bewusstseinsentitäten nur demjenigen möglich ist, der sie aus eigenem Erleben kennt, der ‚weiß, wie es ist‘, in diesen Zuständen zu sein (Nagel 1974), und dass die eigenen Bewusstseinszustände beim Bezug auf die Bewusstseinszustände anderer die Rolle von Paradigmen spielen. Für das Verständnis dieser Form der Bezugnahme ist es wichtig, dass das Subjekt etwa die Fähigkeit besitzt, sich in andere ‚hineinzusetzen‘, ihre Bewusstseinszustände sympathetisch zu imaginieren (Nagel 1974, Hill 1997, Papineau 2002). Wenn dies alles ist, was damit gemeint sein soll, dass die Bewusstseinsentitäten Konstituenten ihrer Begriffe sind, ist dagegen nichts einzuwenden. Im Gegenteil, es bestätigt die hier vertretene Auffassung, dass der Begriff der Bekanntschaft für das Verständnis von Bezugnahme zentral ist. Letztlich besagt es nur, dass die Bezugnahme auf Bewusstseinsentitäten zumindest in den zentralen Fällen voraussetzt, dass das Subjekt mit diesen aus eigenem Erleben bekannt ist. Was darüber hinaus noch damit gemeint sein soll, dass das Bezugsobjekt eine Konstituente des Begriffs ist, bleibt unklar. Dies zu sagen, trägt zum Verständnis dieser Bezugnahme nichts bei. Insbesondere setzt die Fähigkeit, die eigenen Bewusstseinszustände beim Bezug auf die Bewusstseinszustände anderer als Paradigma zu benutzen, bereits die Fähigkeit voraus, sich auf die eigenen Bewusstseinszustände zu beziehen und kann diese damit nicht erklären. Schließlich macht die Zitattheorie auch nicht verständlich, wie die Bezugnahme auf Bewusstseinsentitäten im relevanten Sinn neutral sein kann. Ich werde am Schluss dieses Kapitels aber noch einmal auf die Zitattheorie zurückkommen in Gestalt der mit ihr eng verwandten Auffassung, dass Bewusstseinsentitäten ihre eigenen Gegebenheitsweisen sind (Loar 1997, McLaughlin 2002).

Wenden wir uns nunmehr den Theorien zu, die Bewusstseinsbegriffe als indexikalische Ausdrücke auffassen, also etwa als so genannte Typendemonstrativa, oder als Rekognitionsbegriffe. Was damit gemeint sein soll, dass Bewusstseinsbegriffe Typendemonstrativa sind, kann man sich etwa so veranschaulichen: Man meint mit „Schmerz“ einen Zustand *diesen Typs*, wobei man sozusagen innerlich auf eine Schmerzinstantz deutet. Da eine solche ostensive Definition des Schmerztyps letztlich nur dann bewirken kann, dass der Schmerz begriff fest in der Sprache des Sprechers etabliert wird, wenn dieser auch die Fähigkeit besitzt, diesen Typ bei verschiedenen Gelegenheiten wiederzuerkennen, setzen Typendemonstrativtheorien auch die Fähigkeit zur Rekognition voraus. Ob dies auch umgekehrt gilt und die Trennung zwischen Typendemonstrativ- und Rekognitionstheorien damit letztlich künstlich ist,

braucht uns an dieser Stelle nicht zu interessieren. Das für unsere Zwecke Wesentliche und beiden Vorschläge Gemeinsame besteht darin, dass sie Bewusstseinsbegriffen jenes Merkmal zuschreiben, das Perry im Anschluss an Smart als ‚Neutralität‘ bezeichnete, das aber noch häufiger durch Wendungen der Art charakterisiert wird, dass ein Begriff nicht ‚beschreibend‘ sei, keinen ‚deskriptiven‘ Gehalt habe oder etwas ‚bloß benennt‘. In diesem Sinn schreibt etwa Brian McLaughlin:

Phenomenal concepts are of two sorts: nondescriptive name concepts and type-demonstrative concepts; as such phenomenal concepts lack any descriptive content. They thus do not conceptually reveal anything about the essential natures of phenomenal properties: they simply name or demonstrate them.  
(McLaughlin 2002: 324)

Dies ist geradezu eine kanonische Formulierung der Vorstellung der opaken Bezugnahme. Wir nehmen auf etwas Bezug, ohne dass sich uns die Natur des Bezugsobjekts eröffnet. Wir deuten nur darauf, benennen es bloß, beschreiben es aber nicht. John Perry erläutert, wie diese Vorstellung dazu dienen soll, das Problem informativer Identitätsaussagen zu lösen:

We can now, by way of review, see how Black’s dilemma is to be avoided. Let’s return to our imagined physicalist discovery, as thought by Mary, attending to her sensation of a red tomato:

This<sub>i</sub> sensation is brain state  $B_{52}$ .

This is an informative identity; it involves two modes of presentation. One is the scientifically expressed property of being  $B_{52}$ , with whatever structural, locational, compositional, and other scientific properties are encoded in the scientific term. This is not a neutral concept. The other is being a sensation that is attended to by Mary. This is a neutral concept; if the identity is true, it is a neutral concept of a physical property. (Perry: 205)

Die Idee ist bestechend einfach: der phänomenale Begriff greift das Objekt auf neutrale Weise heraus, er benennt oder demonstriert es nur; der physiologische Begriff dagegen bezieht sich darauf *als* etwas Physiologisches, also etwa *als* neuronale Aktivität eines bestimmten Typs. Und neuronale Aktivität zu sein (z.B.) ist nach dieser Auffassung das, was das Objekt an sich ist, was man als sein Wesen oder seine Natur

bezeichnen kann oder gar, wie McLaughlin mit einer pleonastischen Wendung, als seine „wesentliche Natur“ (*essential nature*; vgl. das Zitat oben). Wir nehmen auf dasselbe Objekt erst opak Bezug und dann transparent, auf eine Weise, die uns seine Natur oder sein Wesen eröffnet oder uns damit bekannt macht. So kann der Identitätssatz uns über die empirisch entdeckte Natur des neutral spezifizierten Objekts informieren.

Ich werde nun ganz allgemein gegen die Vorstellung einer opaken, ‚neutralen‘, oder ‚bloßen‘ Bezugnahme argumentieren. Ich möchte zeigen, dass diese Vorstellung sinnlos ist. Wir nehmen auf Objekte stets *als Objekte bestimmter Natur oder bestimmter Art* Bezug.<sup>43</sup> Die Vorstellung der ‚bloßen‘ Bezugnahme ist bestenfalls die abstrakte Vorstellung von Bezugnahme überhaupt, der deshalb höchstens ein Noumenon, gedacht durch die abstrakte Vorstellung eines Ding an sich, korrespondieren kann, nicht ein spezifisches Objekt wie ein Bewusstseins- oder Gehirnzustand. Dem naiven, philosophisch unverbildeten Denken ist dies offensichtlich. Ich meine *diesen Schmerz* als ein Objekt bestimmter und anderer Art als *meine gegenwärtige Stimmung* und *den Bildschirm vor mir*. Etwas wiedererkennen kann ich etwas nur als Gegenstand bestimmter Art. Indem ich mich auf meine Empfindung als einen Schmerz beziehe, erkenne ich sie als eine Empfindung bestimmter Art wieder, eben als Schmerz. Es ist offensichtlich, dass ich den Schmerz nicht als Gehirnzustand bestimmten Typs wiedererkenne, und wenn ich ihn dennoch als Zustand bestimmter Art wiedererkenne und mich damit auf ihn als Zustand dieser Art beziehe, ist klar, dass ich mich auf ihn als Zustand beziehe, der von anderer Art ist als der relevante Gehirnzustand.<sup>44</sup>

Warum kann es in philosophischen Debatten so schwer fallen, diese einfachen Punkte anzuerkennen? Die Antwort darauf ist sehr schwierig, und ich werde sie an dieser Stelle nur in Ansätzen geben können. Ein wesentlicher Grund liegt aber in der Verwirrung, die sich etwa in einer Bemerkung wie der von McLaughlin zeigt, dass Bewusstseinsbegriffe ihre Objekte nur benennen, sie nicht beschreiben. Womit, möchte man fragen, vergleicht McLaughlin Bewusstseinsbegriffe, wenn er dies sagt? Welche Begriffe beschreiben ihre Objekte? Beschreibt „Wärme“ Wärme, „Wasser“

---

<sup>43</sup> Ich setze also hiermit für die Zwecke dieser Diskussion den Begriff der Natur oder des Wesens mit dem der Art gleich. Daran hängt aber nichts Substanzielles. Wer damit nicht einverstanden ist, dem steht es frei, „Wesen“ und/oder „Natur“ so zu verwenden, dass sie nicht mit „Art“ gleichbedeutend sind, also etwa so, dass das Wesen oder die Natur einer Entität etwas sein kann, dass dem, der sich auf diese Entität bezieht, verborgen sein kann. Für meine Zwecke reicht es aus, wenn ich zeigen kann, dass wir uns auf Dinge stets als Dinge bestimmter Art beziehen.

<sup>44</sup> Dieses Argument wird später noch schematisch dargestellt und ausführlich diskutiert werden.



Wasser, „Neuron“ Neuronen, „Gehirnzustand B<sub>52</sub>“ Gehirnzustand B<sub>52</sub> oder „rot“ die Eigenschaft der Röte? Oder benennen auch diese Ausdrücke ihre Objekte ‚bloß‘? Und wenn sie ihre Objekte ‚bloß‘ benennen, sind sie dann auch ‚neutral‘, so dass wir uns mit der Aussage „Diese Empfindung = Gehirnzustand B<sub>52</sub>“ zweimal ‚neutral‘ auf dieselbe Entität beziehen?

Hier werden zwei Relationen miteinander vermengt, die fundamental verschieden sind. In einem bestimmten, grundlegenden Sinn ist die Beziehung zwischen einem sprachlichen Zeichen und seinem Referenzobjekt immer nur ein ‚bloßes‘ Benennen (oder Demonstrieren). Etwas Beschreiben dagegen heißt sich auf die Eigenschaften eines bereits identifizierten Objekts zu beziehen. Man greift ein Objekt heraus und charakterisiert es dann näher. In einer einfachen prädikativen Aussage wie „Dieser Baum ist alt“ etwa bezieht sich die Nominalgruppe „Dieser Baum“ auf ein Objekt, das dann durch Bezug auf eine seiner Eigenschaften beschrieben wird. Die Beschreibung eines Objektes setzt eine vorgängige Bezugnahme auf dieses Objekts also bereits voraus, und das Objekt, das beschrieben wird, ist ein anderes als das, auf das man sich bezieht um es zu beschreiben, also seine Eigenschaft. Eine Eigenschaft eines Baums ist zwar von diesem nicht in der derselben Weise verschieden wie etwa ein anderer Baum, aber sie sind dennoch nicht identisch, und ihre Verschiedenheit ist auch notwendig, um die Informativität der Aussage zu verstehen. Ich kann mich auf den Baum beziehen, ohne zu wissen, dass er alt ist – ein Objekt zu kennen heißt nicht, all seine Eigenschaften zu kennen – und deshalb kann der Bezug auf diese seine Eigenschaft informativ sein. Und dies zeigt auch, dass es keinen Konflikt zwischen der Tatsache gibt, dass ein Begriff seinen Gegenstand ‚bloß‘ benennt und der Tatsache, dass dieser Begriff in einem gegebenen Kontext eine beschreibende Funktion übernehmen kann. „Schmerz“ beschreibt nicht den Schmerz, aber ich kann meinen Bewusstseinszustand beschreiben, indem ich sage, dass ich Schmerzen habe.

Wenn McLaughlin also sagt, dass Bewusstseinsbegriffe nicht ‚beschreibend‘ sind, sondern ihren Gegenstand nur benennen – und viele Philosophen sagen Dinge dieser Art –, so ist seine Aussage in einer Interpretation eine Binsenweisheit, die auf alle Begriffe zutrifft und so Bewusstseinsbegriffe nicht von anderen Begriffen unterscheidet. Gleichzeitig suggeriert er aber, dass er damit eine besondere Eigenschaft von Bewusstseinsbegriffen erfasst hat. Wir können einstweilen nur konstatieren, dass dieser Behauptung bis jetzt kein klarer Sinn gegeben werden konnte. Nun gibt es allerdings eine Klasse von Ausdrücken, die nicht beschreibend verwendet werden

können, nämlich Demonstrativa. Man sagt normalerweise nicht so etwas wie „Peter ist dies“. Für diese Ausdrücke scheint die Vorstellung, dass sie einen Gegenstand ‚bloß‘ herausgreifen, also noch am ehesten plausibel, da sie tatsächlich keine Klassifikationsausdrücke und damit auch keine Begriffe im herkömmlichen Sinn sind. Es ist deshalb natürlich kein Zufall, dass die Identitätstheoretiker versuchen, Bewusstseinsbegriffe an Demonstrativa oder andere Typen indexikalischer Ausdrücke zu assimilieren. Aber erstens ist es sicherlich nicht richtig, dass Bewusstseinsbegriffe buchstäblich Demonstrativa sind. Bewusstseinsbegriffe können ja zweifelsohne beschreibende Funktion übernehmen. Selbst der Identitätstheoretiker gesteht dies zu, indem er sie als Typendemonstrativa bezeichnet. Auch mit „Ich hatte schon wieder eine Empfindung dieser Art“ kann ich meinen Bewusstseinszustand beschreiben – vorausgesetzt, man versteht, welche Art ich meine. (Bewusstseinsbegriffe sind ja nur Namen für all die verschiedenen Typen von Empfindungen und anderen Bewusstseinsentitäten, die uns helfen, all diese verschiedenen Typen von Entitäten auseinander zu halten.) Zweitens greifen selbst Demonstrativa, obwohl sie keine Klassifikationsausdrücke sind, in einem bestimmten und wesentlichen Sinn dennoch ihre Gegenstände als Gegenstände bestimmter Art heraus. Die These der bloßen Bezugnahme trifft also selbst auf die Fälle nicht zu, für die so noch am ehesten plausibel klingt. Dies will ich nun noch zeigen, um der Vorstellung der bloßen Bezugnahme jeden Boden zu entziehen.

Es muss dabei bei wenigen Bemerkungen bleiben, da die demonstrative Bezugnahme, wie oft bemerkt worden ist, das Bindeglied zwischen Wahrnehmung und Sprache ist und ihre Theorie daher sehr schwierige Fragen aufwirft. Der grundlegende Punkt sollte aber leicht einsichtig sein. Mit „dies“ z.B. – typischerweise verbunden mit einer Zeigengeste – beziehe ich mich auf ein Objekt *als in bestimmten räumlichen Beziehungen zu mir stehend*. Und da ich mich auf dieses Objekt in einer Wahrnehmungssituation beziehe, und die eigentliche Bezugnahme auf das Objekt durch die Wahrnehmung geleistet wird, beziehe ich mich auf dieses Objekt primär *als auf eine bestimmte räumlich ausgedehnte Gestalt*. Es kommt hier sicherlich noch mehr hinzu – Farbigkeit etwa –, und diese Aussage könnte noch in verschiedener Hinsicht interpretiert werden. Es scheint aber schwer zu bestreiten, dass sie in irgendeinem Sinn wahr ist, und in diesem Sinn wird sie unvereinbar mit der Vorstellung sein, dass ich mich hier in opaker Weise, ‚bloß‘ oder ‚neutral‘ auf die Welt be-

ziehe. Stattdessen habe ich intentionalen Zugang zu einer bestimmten räumlichen Gestalt *als solcher*.

Ich schließe somit, dass die Vorstellung der opaken direkten Bezugnahme sinnlos ist. Alle Versuche, ihr Sinn zu geben, sind gescheitert. Die Bezugnahme ist transparent. Wir meinen Entitäten als Entitäten bestimmter Art, und da wir uns z.B. auf Schmerzen nicht als Objekte einer Gehirnart beziehen, müssen wir sie als Objekte anderer Art meinen. Dieses Argument kann schematisch wie folgt dargestellt werden:

- (1) Alle Begriffe referieren auf ihre Bezugsobjekte als Objekte bestimmter Art.
- (2) Bewusstseinsbegriffe referieren auf ihre Bezugsobjekte nicht als Objekte einer physischen Art.

Daraus folgt (3):

- (3) Bewusstseinsbegriffe referieren auf ihre Bezugsobjekte als Objekte anderer Art als physische Objekte.

Zusammen mit der weiteren Prämisse (4):

- (4) Bewusstseinsbegriffe beziehen sich tatsächlich erfolgreich auf Objekte.

folgt daraus (5), also die Negation der Identitätstheorie:

- (5) Bewusstseinsbegriffe beziehen sich erfolgreich auf Objekte, die keine physischen Objekte sind.

Der Identitätstheoretiker ist auf (2) festgelegt, da die Wahrheit von (2) von seiner Erklärung dafür vorausgesetzt wird, wie psychophysische Identitätsaussagen informativ sein können. (Wenn (2) falsch wäre, würden wir uns von vornherein etwa auf Schmerzen als Gehirnzustände bestimmten Typs beziehen.) Auch (4) kann er nicht bestreiten, ohne dass seine Position in den Eliminativismus kollabiert. Der Schluss von (1)-(4) auf (5) scheint unangreifbar. Kann der Schluss auf (3) in Frage gestellt

werden? Ich sehe nicht wie. Und wenn der Identitätstheoretiker (1) verneint, legt er sich auf die sinnlose Vorstellung der opaken, ‚neutralen‘ Bezugnahme fest.

Besprechen wir abschließend einen letzten Versuch des Identitätstheoretikers, die Kritik an seiner Position abzuwehren. Wie oben schon erwähnt, operieren manche Versionen der Identitätstheorie mit der Vorstellung, die gemeinte Bewusstseinsentität könne als ihre eigene Gegebenheitsweise fungieren. Diese Auffassung findet sich etwa bei Loar (1997: 604f) und bei McLaughlin:

The mode of presentation associated with a phenomenal concept is the very phenomenal property to which the concept refers. The concept thus picks out its referent as it is in itself rather than picking it out “nakedly“ or as it stands in relation to something else that functions as a mode of presentation of it. (op. cit.: 326)

Die These, die phänomenale Eigenschaft sei ihre eigene Gegebenheitsweise, klingt von vornherein äußerst merkwürdig. Man kann aber nachvollziehen, welchen theoretischen Zweck sie erfüllen soll. Indem der Identitätstheoretiker den Begriff der Gegebenheitsweise ins Spiel bringt, will er sich gegen den Vorwurf schützen, seine Konzeption sei eine Konzeption opaker, ‚bloßer‘ Bezugnahme („picking it out “nakedly““). Gleichzeitig will er die – wie wir gesehen haben, unüberwindlichen – Schwierigkeiten vermeiden, die sich ergeben, wenn der Zugang über eine vom Bezugsobjekt verschiedene Eigenschaft läuft. Also erklärt er einfach die Eigenschaft zu ihrer eigenen Gegebenheitsweise! Dies gleicht allerdings dem Versuch, sich am eigenen Schopf aus dem Sumpf zu ziehen. Der Begriff der Gegebenheitsweise soll ja die Verschiedenheit des Zugangs zwischen Bewusstseins- und Gehirnbegriffen erklären und damit die kognitive Differenz zwischen dem Wissen, dass man z.B. Freude empfindet, und dem Wissen, dass man sich in dem damit angeblich identischen Gehirnzustand befindet, erklären. Es ist aber offensichtlich, dass das Bezugsobjekt dies nicht kann, da es ja angeblich bei beiden Begriffen dasselbe ist. Es gleichzeitig noch zu einem Präsentationsmodus zu erklären, ist eine leere Geste und ein Akt theoretischer Verzweiflung. Der Identitätstheoretiker kann nicht den Unterschied zwischen dem Fall, dass ich mir introspektiv der eigenen Freude bewusst bin und dem Fall, dass der Neurowissenschaftler auf einem Hirnscan das neuronale Korrelat der Freude bei mir erkennt, durch das seiner Meinung nach jeweils identische Bezugsobjekt erklären. Bevor wir uns im nächsten Kapitel einer der Hauptursachen dieser Verzweiflung zu, nämlich dem Problem der mentalen Verursachung, das allgemein als das

stärkste Argument für die Identitätstheorie angesehen wird, möchte ich nun noch kurz erklären, warum eine Konzeption, die mit dem Begriff der Komposition oder der Mikrostruktur arbeitet, die Probleme der Identitätstheorie vermeiden kann.

## 5.7 Komposition oder Mikrostruktur statt Identität

Warum ist es wichtig, wie es hier vorgeschlagen wird, die Mikro-Makrobeziehung als Kompositions- statt als Identitätsbeziehung zu verstehen? Es ist eine interessante Tatsache, dass viele Menschen keinen sehr klaren Unterschied zwischen der Kompositions- und der Identitätsbeziehung machen. In vielen Diskussionen sowohl mit Philosophen als auch mit Nichtphilosophen habe ich dies feststellen können. Wie schon erwähnt, findet sich auch in der Geschichte der Identitätstheorie ein Beispiel dafür, da U.T. Place in seinem Aufsatz "Is consciousness a brain process?" von 1956, der etwa von Smart (1959) als die erste Formulierung der Identitätstheorie angesehen wird, tatsächlich die Beziehung zwischen Bewusstsein und Gehirn als Kompositionsbeziehung charakterisierte. Ich vermute, dass sich die Tendenz, Identität und Komposition nicht klar zu trennen, darauf zurückführen lässt, dass unsere basalste Vorstellung von einem Objekt das einer bestimmten räumlichen Gestalt ist. Wenn man allein von dieser Vorstellung ausgeht, scheint es, dass wenn man die Bestandteile eines Objekts erfasst hat, man auch sozusagen dessen ganzes Wesen oder Sein erfasst hat. Dies ist sicherlich auch eine wichtige intuitive Motivation für den Physikalismus.

Der Grund dafür, die beiden Beziehungen dennoch scharf zu trennen, ergibt sich aus unserer Diskussion der Identitätstheorie. Er besteht darin, dass man ein Objekt erkennen und sich intentional darauf beziehen kann, ohne doch seine Bestandteile oder seine Mikrostruktur zu erkennen und sich auf diese zu beziehen. Offensichtlich kann man sich etwa auf einen Baum beziehen, ohne dessen Mikrostruktur zu kennen. Dies gilt aber auch umgekehrt. Man kann sich auch auf eine Mikrostruktur beziehen, ohne das Objekt zu erkennen, dessen Struktur sie ist. Wenn man beispielsweise ein ansonsten vertrautes Objekt wie ein Blatt oder einen Wassertropfen durch ein Mikroskop betrachtet, wird man es *als solches* nicht erkennen. Und ich hatte ja dafür argumentiert, dass der Sinn von Bezugnahme, in dem auf etwas intentional bezogen zu sein heißt, auf dieses als solches bezogen zu sein, der grundlegende und letztlich einzig wesentliche Sinn von intentionaler Bezugnahme ist. Wir sind auf Entitäten als

Instanzen bestimmter Typen von Entitäten bezogen, auf Wassertropfen als Instanzen des Wassertropfen-Typs, auf Moleküle als Instanzen des Molekül-Typs usw.. Indem wir Mikro-Makrobeziehungen als Kompositionsbeziehungen konzeptualisieren, können wir dem gerecht werden, denn die Bestandteile einer Entität instanziiieren andere Typen als diese selbst. Der Baum instanziiiert den Baum-Typ, die Moleküle, aus denen er besteht, andere Typen. So können wir die Aporie der Identitätstheorie vermeiden, die in der Sinnlosigkeit der Vorstellung bloß subjektiv verschiedener Zugänge bestand. Wir können dann ohne weiteres etwa sagen, dass Lavoisier, als er das H<sub>2</sub>O-Molekül und damit die Mikrostruktur des Wassers als erster erkannte, tatsächlich etwas Neues, bis dato Unbekanntes in der Welt entdeckte, obwohl er mit Wasser schon lange vertraut war.

Ich wende mich nun dem Problem der mentalen Verursachung zu, um zu zeigen, dass auch dieses gelöst oder aufgelöst werden kann, wenn wir die Mikro-Makrobeziehung und damit auch die Beziehung zwischen Bewusstseins- und Gehirnentitäten als Kompositions- oder Mikrostrukturbeziehung verstehen.

## 6 Das Bewusstsein wirkt

*To be real is to have causal powers.*

Alexanders Dictum

Das Problem der mentalen Verursachung wird oft als der Kern des Leib/Seele-Problems angesehen und als die Klippe, an der der Cartesianische Dualismus zerschellt. Entsprechend gilt es als das stärkste Argument für den ontologischen Physikalismus, dass dieser das Problem der mentalen Verursachung lösen kann. Sehr viele Philosophen sind der Meinung, dass das Problem der mentalen Verursachung uns zwingt, zwischen der Identitätstheorie und dem Epiphänomenalismus zu wählen, und man kann verstehen, wenn angesichts dieser Alternative die meisten sich für die Identitätstheorie entscheiden. Von dem in dieser Arbeit verteidigten Standpunkt aus erscheint das allerdings als eine Wahl wie die zwischen Szylla und Charybdis. Im letzten Kapitel habe ich dafür argumentiert, dass die Identitätstheorie letztlich unverständlich ist und dafür plädiert, die Beziehung zwischen dem Bewusstsein und seinen neuronalen Korrelaten als eine besondere Art von Kompositionsbeziehung aufzufassen. Deshalb ist es das zentrale Anliegen dieses Kapitels, plausibel zu machen, dass man nicht zwangsläufig zwischen Identitätstheorie und Epiphänomenalismus wählen muss.<sup>45</sup> Auch wenn man glaubt, dass die Beziehung zwischen Bewusstseinsentitäten und ihren neuronalen Korrelaten im Besonderen und zwischen Makro- und Mikroentitäten im Allgemeinen nicht Identität ist, sondern Komposition, ist es möglich, das Problems der mentalen Verursachung zu lösen bzw. aufzulösen. Bei genauerer Analyse dieses Problems zeigt sich nämlich, dass es selbst auf einer Verwirrung beruht und sich in Wohlgefallen auflöst, sobald diese Verwirrung klar diagnostiziert worden ist. Grob gesprochen ist die Konfusion die, dass die Kausalkräfte auf verschiedenen Ebenen sich sozusagen im Wettbewerb miteinander befinden, wobei dann die Kausalkräfte auf der untersten Ebene, wenn es denn eine solche überhaupt gibt, diesen Wettbewerb immer zu gewinnen scheinen. Ich werde versuchen zu zeigen, dass diese

---

<sup>45</sup> Für manche Autoren (z.B. Kim) besteht das Problem der mentalen Verursachung in der Schwierigkeit, eine antireduktionistische Deutung mentaler Verursachung zu geben. Diese Beschreibung krankt aber an der Unklarheit des Reduktionsbegriffes, insbesondere der Tendenz, in diesem Begriff die Vorstellung einer vollständigen Erklärbarkeit eines Makrophänomens durch Mikrophenomene (gemäß einem bestimmten Standard von Erklärbarkeit, z.B. der Ableitbarkeit durch Korrelationsgesetze) mit der Vorstellung einer Identität dieser Phänomene zu vermischen. Auf diese Weise wird die Möglichkeit einer Position der Art, wie sie in dieser Arbeit vertreten wird, die die Mikroerklärbarkeit mentaler Phänomene bejaht und also in diesem Sinne reduktionistisch ist, aber nicht in dem Sinn, in dem dieser Begriff Identität impliziert, systematisch verdeckt.

Vorstellung unsinnig ist und dass die kausalen Kräfte und Beziehungen auf den unteren Ebenen die auf den höheren *erklären* statt ihnen die kausale Relevanz gleichsam wegzunehmen.

Das Problem der mentalen Verursachung kann man als die Schwierigkeit beschreiben, die folgenden vier Aussagen miteinander in Einklang zu bringen. Jede dieser Aussagen ist für sich plausibel oder scheint gar unangreifbar; es hat aber leicht den Anschein, dass sie nicht alle gemeinsam wahr sein können.

- (1) Kausale Wirksamkeit des Bewusstseins im Bereich des Physischen/Physikalischen.

Diese Prämisse konstatiert nur das anscheinend offenkundige Faktum mentaler Verursachung. Es ist natürlich, wenn nicht gar unvermeidlich, viele Erklärungen tierischen und menschlichen Verhaltens, in denen auf Bewusstseinszustände Bezug genommen wird, als Kausalerklärungen zu verstehen: sie ging auf ihn zu, *weil* sie ihn kennenlernen *wollte*. Es ist natürlich, den Willen ganz buchstäblich als eine Kraft zu verstehen, die Menschen und Tiere antreibt. Man spricht denn ja auch von Willenskraft, -stärke und dergleichen. Die kausale Wirksamkeit des Mentalen ist zweifelsohne ein integraler Bestandteil unseres natürlichen Selbstverständnisses.

- (2) Kausale Geschlossenheit des Physischen/Physikalischen.

Um z.B. biologische oder geologische Gegebenheiten kausal zu erklären, ist es oft nötig, auf außerbiologische oder -geologische, vor allem physikalische, Phänomene Bezug zu nehmen. Einer gängigen Theorie zufolge ist etwa eine wesentliche Ursache für das Aussterben der Dinosaurier das kosmische Ereignis eines Meteoriteneinschlags gewesen. Dagegen muss man, um Physikalisches zu erklären, den Bereich des Physikalischen nie verlassen. Jedes physikalische Phänomen kann allein durch Bezug auf physikalische Entitäten vollständig kausal erklärt werden. Es gibt eine rein physikalische hinreichende Ursache für das Phänomen.



(3) Realität des Geistigen und Verschiedenheit vom Physischen/Physikalischen.

Es ist offensichtlich, dass wir gemeinhin geistige Phänomene als Teil der Realität behandeln und von physischen Entitäten unterscheiden. In Kapitel 5 habe ich außerdem dargelegt, dass auch der Versuch, von einem höheren philosophischen Standpunkt aus dennoch eine Identität zwischen geistigen und physischen oder physikalischen Phänomenen nachzuweisen, zum Scheitern verurteilt ist.

(4) Keine generelle kausale Überdeterminiertheit der Wirkungen mentaler Phänomene.

Wir gehen im Allgemeinen davon aus, dass Phänomene nicht zwei oder mehr voneinander *unabhängige* und für *sich hinreichende* Ursachen haben. Zwar mag eine solche kausale Überdeterminiertheit gelegentlich vorkommen, etwa wenn der Tod eines Menschen dadurch ausgelöst wird, dass ihn zwei für sich jeweils schon tödliche Kugeln treffen. Die Annahme, dass eine bestimmte Klasse von Phänomenen, etwa eben die physischen Wirkungen mentaler Entitäten, durchgängig überdeterminiert sind, scheint aber äußerst unplausibel.

Man kann nun aus jeder beliebigen Auswahl von drei der Prämissen die Falschheit der jeweils vierten folgern. Wenn das Bewusstsein im Bereich des Physischen/Physikalischen wirksam (1) und mit keinem Element dieses Bereichs identisch (3), dieser aber kausal geschlossen ist (2), müssen die Wirkungen von Bewusstseinsentitäten generell kausal überdeterminiert sein ( $\neg$  (4)). Man kann natürlich aber auch (1), (2) oder (3) aufgeben und die jeweils anderen Prämissen beibehalten. Wie eingangs schon erwähnt, ist es das wohl populärste Vorgehen gewesen, mit (1), (2) und (4) gegen (3) und damit für die Identitätstheorie oder den Eliminativismus zu argumentieren. Die Argumente aus Kapitel 4 und 5 dafür, dass die Negation von (3) unverständlich ist, sollen hier aber nicht noch einmal wiederholt werden, und deshalb wird im Folgenden (3) nicht noch einmal diskutiert. Wie schon gesagt, ist das Ziel dieses Kapitels ja zu zeigen, dass das Problem der mentalen Verursachung auch gelöst oder aufgelöst werden kann ohne (3) aufzugeben. Der Schlüssel zu dieser Auflösung liegt in einem adäquaten Verständnis von (2). Durch das Nebeneinanderstellen

der Begriffe des Physischen und des Physikalischen habe ich schon angedeutet, dass es entscheidend auf diese Begriffe und ihre Interpretation ankommt. Ich werde später eingehend dafür argumentieren, dass wenn wir den Begriff des Physischen klar von dem des Physikalischen trennen, wir (2) in einer Lesart akzeptieren können, in dem das Bewusstsein auf die physische, aber nicht auf die physikalische Welt wirkt.

Vorerst möchte ich mich aber den zwei anderen übrig bleibenden Reaktionen auf das Problem der mentalen Verursachung zuwenden, der Zurückweisung von (1) oder (4), also dem Epiphänomenalismus bzw. dem Versuch, dem Begriff der kausalen Überdeterminiertheit einen positiven Sinn zu geben. Durch eine eingehendere Betrachtung dieser vermeintlichen Optionen werden wir zu einem tieferen Verständnis der Problemlage kommen und ihre Diagnose und Auflösung vorbereiten.

## **6.1 Kritik des Epiphänomenalismus**

Das Problem der mentalen Verursachung wird oft mit dem Hinweis eingeführt, dass wir zwischen kausal relevanten und kausal irrelevanten Eigenschaften von Ereignissen oder anderen Entitäten unterscheiden müssen. Z.B. ist die Eigenschaft des von mir getrunkenen Whiskeys, Alkohol zu enthalten, die Ursache meiner Betrunkenheit, aber nicht sein Geschmack oder die Tatsache, dass ich ihn im Supermarkt gekauft habe (Bieri 1992). Nicht der Text des von der Sängerin gesungenen Liedes ist die Ursache dafür, dass das Glas zerspringt, sondern die Frequenz der von ihr hervorgebrachten Töne. Wenn man solche Beispiele betrachtet, kann die Idee entstehen, dass vielleicht manche Eigenschaften kausal völlig irrelevant, epiphänomenal im stärksten Sinne sind, dass sie also zwar vielleicht verursacht wurden, aber selbst nichts verursachen. Das Bewusstsein als ein Epiphänomen in diesem Sinn zu verstehen, ist die klassische epiphänomenalistische Position in der Philosophie des Geistes. Oft wird versucht, sie mit bestimmten suggestiven Bildern zu veranschaulichen: das Bewusstsein verhält sich zur mit ihm korrelierten Gehirnaktivität wie der Dampf, den der Antrieb einer Dampflokomotive erzeugt; Bewusstseinsereignisse sind „Fransen des Universums“. Wenn man diese Veranschaulichungen und die Beispiele für die kausale Irrelevanz bestimmter Eigenschaften genauer betrachtet, zeigt sich aber, dass sie die epiphänomenalistische These nicht begründen, plausibilisieren oder auch nur verständlich machen können. Der Epiphänomenalismus erweist sich als typisch phi-

losophische begriffliche Übersteigerung realer Verhältnisse, eine Ausgeburt philosophischer Phantasie, die jenseits jeder möglichen Erfahrung liegt.

Halten wir zuerst fest, dass in den oben angeführten Beispielen die kausale Relevanz oder Irrelevanz einer bestimmten Entität ja nur *bezüglich einer bestimmten Wirkung* konstatiert wird. Die Eigenschaft des Whiskeys, Alkohol zu enthalten, ist die Ursache der Betrunkenheit, aber nicht der Tatsache, dass die Waage 1 Kilo als Gewicht der gefüllten Flasche anzeigt. Dass der Whiskey ein Geschenk seiner Geliebten war, ist nicht die Ursache der Betrunkenheit, wohl aber der Tatsache, dass der Beschenkte sich über die überreichte Flasche sehr gefreut hat. Die Beispiele illustrieren die banale Beobachtung, dass ein gegebenes Etwas die Ursache von manchen Dingen ist, aber nicht von allen. Sie können aber nicht die Vorstellung eines gar nichts verursachenden Etwasses begründen.

Der Philosoph, der sich durch das Problem der mentalen Verursachung zum Epiphänomenalismus gedrängt fühlt, eliminiert den Bezug auf eine bestimmte relevante Wirkung und führt so die Vorstellung einer kausalen Irrelevanz *simpliciter*, in Bezug auf alle Entitäten überhaupt, ein. Es ist die Vorstellung eines Etwasses, dessen Existenz oder Nichtexistenz gar keinen Einfluss auf das Geschehen in der Welt hat. Es ist die Vorstellung einer Entität, die gar nicht mit anderen Entitäten interagieren kann, denn was sollte Interaktion anders sein als kausale Interaktion? Machen wir uns zuerst klar, dass die klassischen Veranschaulichungen der epiphänomenalistischen These diese tatsächlich nicht stützen können. Der der Lokomotive entweichende Dampf hat zwar keine aktive kausale Rolle bezüglich des weiteren Funktionierens der Dampfmaschine, oder jedenfalls höchstens eine vernachlässigbare. Er ist eben eine Wirkung dieses Funktionierens, ihr Abfallprodukt. Aber das heißt natürlich nicht, dass er selbst vollkommen wirkungslos ist. Ein Abfallprodukt ist genauso wenig wirkungslos wie seine Ursache und kann im Übrigen auch wie z.B. bei einem Abgasturbolader dazu benutzt werden, die Leistung der Maschine zu steigern, deren Abfallprodukt es ist. Niemand würde wohl ernsthaft annehmen, dass Wasserdampf oder andere Abfallprodukte Epiphänomene in dem starken, philosophischen Sinn sind. Also selbst wenn es Sinn machen würde, das Bewusstsein als Abfallprodukt physiologischer Gehirnvorgänge anzusehen – und dies entspricht zumindest nicht unserer gewöhnlichen Konzeption der kausalen Rolle des Bewusstseins – wäre damit kein hinreichender Grund gegeben, ihm jedwede Wirkung, oder selbst jedwede Wirkung auf physische/physikalische Entitäten abzusprechen. Denn davon, dass etwas

sich vollkommen im Nichts auflöst oder verliert, ohne jede Wirkung, also ohne eine Spur irgendwelcher Art zu hinterlassen, haben wir sicherlich keinen klaren Begriff, erst recht keinen empirisch gestützten. Und da das Bewusstsein rundherum von physischen Entitäten umgeben ist, könnte es, wenn es überhaupt Wirkungen hat, wohl kaum umhin, auch auf physische Vorgänge zu wirken.

Gegen den Epiphänomenalismus in Bezug auf geistige Phänomene spricht nun nicht nur, dass er unserem Selbstverständnis als handelnde, aufgrund bewusster Entscheidungen Veränderungen in der Welt verursachende Wesen, zuwiderläuft – obwohl dies schon ein mehr als ausreichender Grund ist, um diese Vorstellung nur dann zu akzeptieren, wenn man zu ihr absolut gezwungen zu sein scheint. Ein weiteres schlagendes Argument gegen den Epiphänomenalismus ergibt sich daraus, dass er nicht verständlich machen kann, wie epiphänomenale Entitäten überhaupt erkannt werden können. Es ist natürlich, Erkenntnisvorgänge als kausale Vorgänge zu verstehen.<sup>46</sup> Diese Auffassung sitzt so tief, dass sie sich in vielen Ausdrücken natürlicher Sprachen niedergeschlagen hat. Man spricht etwa von den sinnlichen oder nicht-sinnlichen *Eindrücken* oder *Impressionen*, die man von der Umwelt empfangen hat, von *Spuren* der Erinnerung, die bestimmte Ereignisse hinterlassen haben und dergleichen mehr. Es gibt keinen Grund, warum die Erkenntnis eigener oder fremder Bewusstseinszustände dabei eine Ausnahme sein sollte. Das Zischen der Silvesterraketen verursacht unter normalen Umständen eine entsprechende Wahrnehmung; warum sollte die Schmerzempfindung nicht eine, oder sogar die Ursache nicht nur des Schmerzensschreis, sondern auch der (mehr oder minder nüchternen) Feststellung sein, dass man Schmerzen empfindet? Der Epiphänomenalist behauptet, dass mein Denken, Wollen und Empfinden nie die Ursache oder Teil der Ursache davon ist, dass ich mir bewusst werde, dass ich denke, will und empfinde. Wie aber soll ich etwa wissen, dass ich Schmerzempfindungen habe, wenn diese Empfindungen gar nicht die Kraft haben, entsprechende Gedanken auszulösen und diese wiederum entsprechende sprachliche Äußerungen?

Eine mögliche Antwort ist, dass es sich hier um rationale Beziehungen handelt, dass wir uns im „Reich der Gründe“, nicht der Ursachen, befinden. Demnach wäre also etwa die Tatsache, dass ich Schmerzempfindungen habe, zwar ein Grund für die Überzeugung, dass ich Schmerzempfindungen habe, aber nicht die Ursache

---

<sup>46</sup> Dies bedeutet nicht, dass man alle Thesen der sogenannten kausalen Erkenntnistheorie akzeptieren muss.

dieser Überzeugung.<sup>47</sup> Diese Art der Opposition zwischen Gründen und Ursachen hat zwar in der Philosophie eine lange Tradition, ist aber meiner Meinung nach sachlich nicht gerechtfertigt. Das soll aber nicht heißen, dass es nicht durchaus angebracht und wichtig ist, Gründe und Ursachen voneinander zu unterscheiden.<sup>48</sup> Z.B. ist die Tatsache, dass x sich in einem bestimmten Gehirnzustand G befindet, für mich, wenn ich weiß, dass Gs neuronale Korrelate des Bewusstseins sind, ein Grund zu glauben, dass x bei Bewusstsein ist, obwohl G nicht die Ursache von Bewusstsein ist – jedenfalls wenn die Thesen dieser Arbeit über das Verhältnis zwischen Gehirn- und Bewusstseinszuständen richtig sind. Oder, um ein weniger kontroverses Beispiel zu nehmen: dass mein Gegenüber häufig hustet und eine vom Schnupfen gerötete Nase hat, ist ein (wenn auch nicht zwingender) Grund zu der Annahme, dass er sich erkältet hat. Schnupfen und Husten sind aber Symptome einer Erkältung, nicht ihre Ursachen. Es folgt aber nicht, dass Gründe nicht stets auch Ursachen sind. Und zwar gilt dies unabhängig davon, ob wir unter Gründen nun, wie es in jüngerer Zeit verschiedentlich propagiert worden ist (z.B. McDowell 1994) die jeweiligen „Tatsachen in der Welt“ verstehen, oder die entsprechenden mentalen Einstellungen, durch deren Existenz diese Tatsachen allein rationale Bedeutsamkeit bekommen können. Sicherlich ist die Tatsache, dass mein Gegenüber Husten hat, kausal relevant. Unter anderem ist sie ursächlich für meine Überzeugung, dass er Husten hat. Aber auch der rational motivierte Übergang von dieser Überzeugung zu der, dass er sich eine Erkältung eingefangen hat, dieses Schließen von dem einen Sachverhalt auf den anderen, ist ein bestimmter Typ eines kausalen Prozesses. Dass eine Beziehung kausal ist, impliziert nicht, dass sie nicht rational oder intentional ist (Searle 1983, Seebaß 1993).

Wenn Gründe keine kausale Relevanz hätten, hätte allem Anschein nach die Rationalität von Überzeugungen oder anderen geistigen Einstellungen keinen Einfluss auf das Weltgeschehen. Ich glaube nicht, dass es vernünftige Gründe geben kann, die uns dazu zwingen (!) würden, Prämissen zu akzeptieren, die zu einer solchen Konklusion führen. Eine letzte Bemerkung noch zu der These, dass Gründe keine Ursachen sind. Eine wesentliche Motivation für diese Auffassung ist sicherlich das Problem der mentalen Verursachung gewesen bzw. die (meist stillschweigende) Annahme, dass dieses Problem nicht auf befriedigende Weise gelöst werden kann.

---

<sup>47</sup> Hier und im Folgenden meine ich, wenn ich sage, dass etwas ursächlich für etwas ist, dass es entweder *die* Ursache ist oder Teil einer komplexen Ursache.

<sup>48</sup> Vgl. z.B. Wittgensteins Blaues Buch als einen *locus classicus* dieser Unterscheidung.

Vor diesem Hintergrund scheint ein starker Antagonismus von Ursachen und radikal nicht-kausalen Gründen die Möglichkeit zu eröffnen, eine sich als die Erforschung der Gründe von Handlungen verstehende Handlungstheorie zu betreiben, ohne mit dem Prinzip der kausalen Geschlossenheit des Physikalischen in Konflikt zu kommen. Wenn aber dieser Konflikt, wie ja hier gezeigt werden soll, ohnehin nur ein scheinbarer ist, erledigt sich damit eine, wenn nicht die, wesentliche Motivation für eine nicht-kausale Handlungstheorie. Und diese Unterhöhlung der Motivation für bestimmte Theorien scheint mir auch allgemein philosophisch wichtiger und interessanter zu sein als ihre Widerlegung. Den Epiphänomenalismus zu widerlegen ist z.B. einerseits sehr leicht, weil er unserem Selbstverständnis in offenkundigen und auch, wie wir gesehen haben, etwas weniger offenkundigen Hinsichten widerspricht. Gleichzeitig ist es, wenn sich jemand genötigt sieht, eine Auffassung anzunehmen, die so offensichtlich unserer natürlichen Selbstkonzeption zuwiderläuft, schwer, den *common ground* zu finden, den man ja zum Argumentieren unbedingt braucht. Und schließlich beseitigt die Widerlegung, selbst wenn sie gelingt, selbst wenn sie auch den Widerlegten überzeugt und nicht nur den Widerleger, ja nicht die Schwierigkeiten, die ihn zu der philosophischen Verzweiflungstat gebracht haben, einen Ausweg z.B. im Epiphänomenalismus (Skeptizismus, Antirealismus etc.) suchen zu wollen. Deshalb ist die vordringlichste philosophische Aufgabe die, zu zeigen, dass es keinen Zwang gibt, zwischen solchen gleichermaßen unattraktiven, ja widersinnigen Alternativen wie etwa dem Epiphänomenalismus und der Identitätstheorie wählen zu müssen.

In den Debatten über mentale Verursachung wird oft angenommen, dass die Kausalität vom Geistigen zum Physischen (Physikalischen) problematisch ist, die Kausalität vom Physischen zum Geistigen und die vom Geistigen zum Geistigen aber nicht oder jedenfalls nicht in gleichem Maße. Entsprechend versteht die Standardvariante des Epiphänomenalismus geistige Phänomene als von physischen Phänomenen verursacht, aber selbst nichts Physisches verursachend. Oder man könnte auch vorschlagen, dass geistige Phänomene zwar nichts Physisches, aber andere geistige Phänomene verursachen.<sup>49</sup> Mir scheinen aber beide Züge sehr fragwürdig zu sein. Dass Bewusstseinszustände zwar keine aktive, wohl aber eine passive kausale Rolle haben können, ist nur vordergründig plausibel. Gegeben eine bestimmte natürliche Annahme über Kausalität, erweist sich diese These aber als unhaltbar. Ich meine die An-

---

<sup>49</sup> Diesen Vorschlag hat mir Bernhard Thöle gemacht.

nahme, dass Kausalität (zumindest Ereigniskausalität) einen Transfer von Energie zwischen Ursache und Verursachtem involviert.<sup>50</sup> Wenn diese Annahme richtig ist, und zumindest für die prototypischen Fällen von Kausalität (ein Billardball trifft den nächsten usw.) trifft sie zweifelsohne zu, ist die Kausalität vom Physischen zum Geistigen genauso problematisch wie die in umgekehrter Richtung. Denn dann müsste die Verursachung des Bewusstseinsereignisses Energie erfordern, die demnach auf der physikalischen Ebene ‚fehlt‘ und zu einer Verletzung des Energieerhaltungssatz und damit einem Bruch der kausalen Geschlossenheit des Physikalischen führen muss. (Was physikalisch geschieht, würde davon abhängen, ob ein physikalisches Ereignis ein Bewusstseinsereignis verursacht oder nicht.)

Mit dem gleichen Argument kann auch gezeigt werden, dass die Kausalität zwischen geistigen Phänomenen genauso problematisch ist wie die zwischen geistigen und physischen. Wenn Kausalität generell mit dem Transfer von Energie verbunden ist, konfliktiert eine Kausalität zwischen geistigen Ereignissen genau so mit dem Energieerhaltungssatz und der kausalen Geschlossenheit des Physikalischen wie die Kausalität vom Physischen zum Geistigen. Wenn etwas Energie ‚kostet‘, muss sie auf der physikalischen Ebene fehlen. Und es entspricht durchaus unserem gewöhnlichen Verständnis von Bewusstseinsphänomenen, dass innermentale Kausalität Energie ‚kostet‘. Z.B. hat es mich viel *Willenskraft* gekostet, diese Arbeit abzuschließen. Ich musste mitunter meinen Geist gegen seine natürliche Richtung dazu zwingen, bestimmte Überlegungen anzustellen und zu einem Resultat zu kommen. Der Wille, zu einem Ende zu kommen, war also die Ursache vieler Überlegungen, die ich angestellt habe.

Die letzten Argumentationsschritte zusammenfassend lässt sich sagen, dass weder die Auffassung, dass die Beziehung zwischen einem Bewusstseinszustand und der Erkenntnis, dass man selbst oder jemand anders sich in diesem Bewusstseinszustand befindet, gar nicht kausal – sondern rein rational oder intentional – ist, plausibel zu sein scheint, noch die, dass eine Kausalität zwischen mentalen Entitäten, oder auch nur von physischen zu mentalen Entitäten, weniger problematisch ist als die von mentalen zu physischen. Wenn letztere mit der kausalen Geschlossenheit des Physikalischen unvereinbar ist, so ist es erstere auch. Wenn es uns aber gelingt, das

---

<sup>50</sup> Dies muss von der Kausalitätstheorie unterschieden werden, die Kausalität mit Energietransfer identifiziert. Die These, dass Kausalität stets mit Energietransfer *verbunden* ist, ist unabhängig davon plausibel, ob man eine solche reduktive Definition von Kausalität akzeptiert.

Problem der mentalen Verursachung aufzulösen, werden all diese Formen der Kausalität gleichermaßen unproblematisch sein.

Der Epiphänomenalismus kann also die Erkenntnis von Bewusstseinszuständen nicht verständlich machen. Epiphänomene in dem Sinn von Entitäten, die nichts verursachen, kann es nur geben, wenn sie entweder gar nicht erkannt werden oder es Erkenntnisprozesse oder -beziehungen gibt, die radikal nichtkausal sind. Ob die Vorstellung von Phänomenen, die wir gar nicht erkennen können, sinnvoll ist, braucht hier nicht erörtert zu werden, da es außer Zweifel steht, dass wir Bewusstseinsentitäten erkennen. Dass Erkenntnisprozesse kausal sind, scheint kaum weniger gewiss. Ansonsten wäre es ein Wunder, dass der Geist seine Gegenstände normalerweise richtig erfasst. Man müsste sich auf eine „prästabilisierte Harmonie“ zwischen der Erkenntnis und ihren Objekten à la Leibniz berufen, und dabei handelt es sich ja um wenig mehr als einen elaborierten Terminus für das mysteriöse Faktum einer Übereinstimmung ohne kausale Beziehung.<sup>51</sup> Mir scheint somit, dass der Epiphänomenalismus genau wie die Identitätstheorie nur als ein theoretischer Verzweiflungsakt angesichts eines unlösbar scheinenden Problems der mentalen Verursachung zu verstehen ist. Ich möchte nun zeigen, dass dies auch für die Vorstellung gilt, die Wirkungen von Bewusstseinsentitäten seien generell kausal überdeterminiert.

## 6.2 Kritik der Vorstellung kausaler Überdeterminiertheit

Warum kann man das Problem der mentalen Verursachung eigentlich nicht einfach lösen, indem man sagt, dass alle Wirkungen mentaler Entitäten sowohl von diesen als auch von physikalischen Phänomenen verursacht werden? Damit behielte das Bewusstsein eine kausale Rolle, und gleichzeitig bliebe die kausale Geschlossenheit des Physikalischen gewahrt. Eine mögliche Reaktion auf diesen Zug ist die Behauptung, dass diese so genannte kausale Überdeterminiertheit schon durch das Prinzip der kausalen Geschlossenheit des Physikalischen ausgeschlossen ist. Dieses Prinzip kann man in der Tat so verstehen, dass es besagt, dass jedes physikalische Ereignis

---

<sup>51</sup> Man könnte meinen, dass sich daraus auch ein Problem für die schon angedeutete und später noch weiter zu entwickelnde Auffassung ergibt, dass die Beziehung zwischen dem Bewusstsein und seinen neuronalen Korrelaten nicht kausal ist. Muss man dann nicht sagen, dass auch diese Form von „Harmonie“ oder „Übereinstimmung“ mysteriös ist? Behauptet wird hier aber ja nicht die (absurde) These, dass es gar keine nichtkausalen Zusammenhänge zwischen Entitäten gibt. Anstößig ist nicht *per se*, dass der Epiphänomenalismus einen nichtkausalen Zusammenhang zwischen bestimmten Bewusstseinsentitäten und auf sie gerichteten Zuständen der Erkenntnis postulieren muss. Merkwürdig ist nur, dass diese Beziehung nach den üblichen Kriterien ein Kandidat für eine Kausalbeziehung ist



nicht nur eine hinreichende physikalische Ursache hat, sondern dass diese auch die vollständige und alleinige Ursache ist. Man könnte weiter argumentieren, dass dies sogar bereits analytisch im Begriff der hinreichenden Ursache enthalten ist. Alternativ könnte man den Ausschluss anderer als physikalischer Ursachen als ein eigenes Exklusionsprinzip formulieren. Dieses würde aber im Wesentlichen auf dasselbe hinauslaufen wie unsere Prämisse (4). Für unsere Zwecke ist aber nicht wichtig, wie man sich an dieser Stelle verhält. Ich habe hier die Leugnung der kausalen Überdeterminiertheit aus rein expositorischen Gründen als eine eigene Prämisse aufgeführt.<sup>52</sup> Der Sache nach muss so oder so entschieden werden, ob Überdeterminiertheit eine vernünftige Option ist oder nicht. Dabei ist es nebensächlich, ob man die Negation der Überdeterminiertheit als im Prinzip der kausalen Geschlossenheit enthalten ansieht oder als eigene Prämisse abtrennt.

Was also macht die Überdeterminiertheitshypothese so anstößig? Viele Autoren sind doch anscheinend überzeugt, dass Fälle kausaler Überdeterminiertheit tatsächlich vorkommen. Oft werden Beispiele der folgenden Art gegeben. Ein Mensch wird von zwei Kugeln in Herz und Gehirn getroffen. Jede verursacht Verletzungen, die für sich genommen ausreichen würden, ihn vom Leben zum Tod zu befördern. Ist also sein Tod nicht kausal überdeterminiert? Und zeigt dies nicht, dass kausale Überdetermination eine eigentlich – jedenfalls metaphysisch gesehen – ganz harmlose Sache ist? Besonders am Fall des Bewusstseins wäre demnach nur, dass seine Wirkungen stets überdeterminiert sind! Dies wäre zwar merkwürdig, aber reicht dies wirklich aus, um die Überdeterminiertheitshypothese zurückzuweisen? Immerhin würde sie doch das Problem der mentalen Verursachung lösen!

Wir müssen aber hier nicht die Frage entscheiden, ob eine generelle Überdeterminiertheit der Wirkungen mentaler Entitäten – im Gegensatz zu einer gelegentlichen kausalen Überdeterminiertheit anderer Phänomene – eine akzeptable Vorstellung ist oder nicht. Wenn man die angeblichen Beispiele kausaler Überdeterminiertheit näher betrachtet, zeigt sich nämlich bald, dass hier in Wirklichkeit keine Überdeterminiertheit vorliegt. D.h. es ist nicht der Fall, dass hier zwei Ursachen die selbe und damit doppelt verursachte Wirkung hervorbringen. Analysieren wir diese Situationen also etwas genauer. Dadurch werden wir auch ein besseres Verständnis der

---

<sup>52</sup> Meine Darstellung ist von der von Scott Sturgeon (2000) beeinflusst. Wenn man die Leugnung der Überdeterminiertheit als in der Prämisse der kausalen Geschlossenheit des Physikalischen enthalten betrachtet, kann man wie z.B. Bieri (1981) das Problem der mentalen Verursachung auch mit nur 3 Prämissen entwickeln.

Umstände gewinnen, unter denen zwei oder mehrere vermeintliche Ursachen eines Phänomens tatsächlich miteinander ‚konkurrieren‘, d.h. in denen man sinnvoll fragen, welches denn nun die Ursache dieses Phänomens ist. Dies wird ein wesentlicher Schritt auf dem Weg zur Lösung des Problems der mentalen Verursachung sein. Denn wir werden sehen, dass die Beziehung zwischen Bewusstseinsentitäten und ihren neuronalen Korrelaten im Besonderen und die zwischen Makroentitäten und den Mikroentitäten, aus denen sie bestehen, nicht von dieser Art ist. Eine Entität und ihre Bestandteile oder ihre Mikrostruktur sind keine ‚konkurrierenden‘ Ursachen.

Kehren wir also zu unserem Kugelbeispiel zurück. Zunächst einmal ist ja klar, dass die zwei Kugeln raumzeitlich und auch qualitativ deutlich voneinander unterschiedene Wirkungen haben, die eine im Gehirn, die andere im Herzen. Wie verhält sich dies aber zur Verursachung des Todes? Dass die beiden Ereignisse verschiedene Wirkungen haben, schließt ja nicht *per se* aus, dass sie auch mindestens eine gemeinsame Wirkung haben. Hier kann man sich verschiedene Fälle denken. In manchen Fällen würde man sagen, dass die von den beiden Kugeln in Herz und Gehirn ausgelösten Ereignisse *gemeinsam* den jeweiligen Tod verursacht haben. Dies wäre etwa dann der Fall, wenn ein durch den Herzschuss verursachter Blutverlust und eine durch den Kopfschuss ausgelöste Störung der kreislaufregulierenden Funktionen des Gehirns zusammen einen irreversiblen Kreislaufkollaps bewirken würden. Dass aber die beiden Ereignisse zusammen den Tod verursachen, würde man nur dann sagen, wenn keines von beiden für sich hinreichend gewesen wäre, um den Tod herbeizuführen – oder jedenfalls um *diesen Tod* herbeizuführen. D.h. entweder hätte jedes der beiden Ereignisse gar nicht den Tod des Schussverletzten herbeigeführt, oder es hätte ihn z.B. erst wesentlich später und auf andere Weise herbeigeführt. Damit ist klar, dass es sich hier nicht um Überdetermination in dem relevanten Sinne handelt, denn eine solche läge nur dann vor, wenn jedes der beiden Ereignisse für sich genommen hinreichend gewesen wäre, den Tod herbeizuführen. Dies ist aber nicht der Fall, wenn sie Teil einer gemeinsamen Ursache sind.

Was ist, wenn eins der Ereignisse nur für eine minimale Veränderung verantwortlich ist, also z.B. dafür, dass der Tod eine Millisekunde früher eintritt? In einem solchen Fall würden wir seinen Einfluss normalerweise für vernachlässigbar halten und es nicht als Todesursache ansehen. Stellen wir uns etwa vor, dass die zwei Schüsse von zwei Personen unabhängig voneinander abgegeben werden. Der erste Schütze will seinem Opfer nur einen Denkart verpassen. Er schießt ihm in den

Fuß und ruft nur vergleichsweise geringfügige Verletzungen hervor. Der zweite Schuss trifft mitten ins Herz und führt rasch den Tod herbei, der wegen des durch den ersten Schuss bereits erlittenen Blutverlusts noch etwas schneller eintritt als dies sonst der Fall gewesen wäre. In so einem Fall würden wir normalerweise nicht sagen, dass der erste Schuss die Ursache des Todes war. Aber was wäre, wenn das Opfer noch gerettet werden könnte, Hilfe auch unterwegs ist, aber wegen der Beschleunigung des Sterbeprozesses durch den ersten Schuss etwas zu spät kommt?

Ein weiterer Grund für den Eindruck, dass das gleiche Ereignis zwei Ursachen haben kann, ist oft mangelnde Klarheit in Bezug auf die nähere Abgrenzung dieses Ereignisses. Z.B. kann man sowohl das permanente Ende des Herzschlagens als auch das der Gehirntätigkeit als den Eintritt des Todes auffassen. Dies sind verschiedene Todesbegriffe, verschiedene Vorschläge, den Todesbegriff zu bestimmten, theoretischen oder praktischen Zwecken, zu präzisieren. (Und was ist der ‚richtige‘ Todesbegriff? Es gibt nicht den richtigen Todesbegriff, verschiedene Todesbegriffe können für bestimmte Zwecke und Kontexte die jeweils richtigen, angemessenen sein.) Um unseren Fall angemessen zu beschreiben, muss dann geklärt werden, ob man unter dem Tod des Individuums seinen Herztod oder den Gehirntod oder ggfs. auch beides in Kombination verstehen will. Wenn dies geklärt ist, lässt sich der Verdacht auf Überdetermination wiederum entkräften. Entweder ist das jeweilige Ereignis jeweils nur für den Tod im Sinne von Herztod oder Gehirntod verantwortlich, oder, wenn man den Todesbegriff so verstehen will, dass beides in Kombination erforderlich ist, sind die beiden Schüsse wiederum Teil einer gemeinsamen Ursache.

Betrachten wir nun den Fall, dass beide Kugeln so schwere Verletzungen auslösen, dass jede für sich ausreichend wäre, den Tod herbeizuführen. Ist das nicht ein klarer Fall von Überdetermination? Ich glaube nicht, denn die Tatsache, dass beide Kugeln den Tod herbeigeführt *hätten*, bedeutet nicht, dass sie ihn auch tatsächlich beide herbeigeführt haben. Dass eine Kugel ihn herbeigeführt hätte, heißt ja, dass dieser durch die Wirkung der Kugel eingetreten wäre, wenn nichts Relevantes dazwischen gekommen wäre. Aber dies kann natürlich geschehen und geschieht mitunter auch. Das Eingreifen eines Arztes etwa kann den Tod verhindern, oder ein anderes Ereignis kann dem Schuss als Todesursache zuvorkommen. Dieser letzte Fall könnte etwa auch in unserem Beispiel eintreten. Wenn der zweite Schuss dem armen Opfer das Lebenslicht ausbläst, noch bevor der erste sein tödliches Werk vollendet hat, würden wir sagen, dass er die Ursache des Todes war, oder, unter manchen Um-

ständen, entsprechend dem oben beschriebenen Muster, dass beide Schüsse gemeinsam den Tod herbeigeführt haben, wobei ihr Beitrag natürlich unterschiedlich groß sein kann. Es gibt aber keinen Fall, der plausibel als einer beschrieben werden kann, in dem es zwei unabhängige und für sich hinreichende und vollständige Ursachen eines Ereignisses gibt.

Das Fazit dieser Überlegungen ist, dass kein Szenario des besprochenen Typs plausibel als eines beschrieben werden kann, in dem es zwei unabhängige und dennoch für sich hinreichende und vollständige Ursachen desselben Ereignisses (Zustands, Tatsache etc.) gibt. Ich denke, es ist auch klar, dass dies so sein muss. Wenn wir den Weg der Kugeln durch den Körper und ihr Zerstörungswerk nachvollziehen, ist klar, dass wir hier zwei verschiedenen Kausalpfaden oder -ketten folgen. Es wird normalerweise etwa keine Frage sein, welche Kugel die Herzwand perforiert hat. Schwieriger wird es natürlich, wenn wir den Kausalpfaden weiter folgen wollen bis hin zu einem vergleichsweise so diffusen und vage definierten Ereignis wie dem Eintritt des Todes. Hier muss, um zu einem klaren Bild der Situation zu kommen, die Definition des Todes oder eines anderen Wirkungsereignisses ggfs. geschärft werden. Wenn man dies tut, zeigt sich aber, dass das Wirkungsereignis entweder einer der beiden möglichen Ursachen zurechenbar ist oder dass diese Teile einer gemeinsamen Ursache sind. Zwischen diesen beiden Typen von Fällen eine scharfe Trennlinie zu ziehen ist sicherlich oft sehr schwierig und mit einer gewissen Willkür verbunden. Es ist bekannt nicht leicht, systematisch zwischen Ursachen und notwendigen Hintergrund- oder Randbedingungen zu unterscheiden. Dies gilt besonders für ein so außerordentlich komplexes System wie den menschlichen Organismus, dessen Funktionieren von einer Vielzahl verschiedener Faktoren abhängt. (Es ist aber in der Praxis auch meist nicht nötig, eine Entscheidung in diesen Arten von Fällen herbeizuführen, es sei denn, es geht etwa um solche Fragen wie die, wer für den Tod eines Menschen moralisch und/oder juristisch verantwortlich ist.) In keinem Fall aber werden wir Grund haben, zwei unabhängige Kausalketten als jeweils für sich hinreichende Ursachen ein und desselben Phänomens anzuerkennen.

Wichtig für die folgenden Erörterungen ist aber, dass wir es in dem diskutierten Typ von Szenario durchaus mit Fällen zu tun haben, in denen zwei oder mehr potentiell *miteinander konkurrierende* Ursachen vorliegen. Ich glaube, dies liegt daran, dass die folgenden Bedingungen erfüllt sind:

1. Es liegen Kausalbeziehungen respektive -erklärungen der gleichen Art vor.
2. Die Ursachen sind derart, dass sie potentiell die gleichen Wirkungen erklären können.

Damit, dass die Kausalerklärungen Kausalerklärungen gleicher Art sind, ist vor allem gemeint, dass es sich in beiden Fällen um Kausalerklärungen handelt, die eine Kausalkette zeitlich aufeinander folgender Ereignisse nachzeichnen. Man kann diese Erklärungen als mechanistische Erklärungen bezeichnen. Man sucht nach einem Mechanismus, einer Abfolge von Ereignissen, die einen kausalen Pfad zu dem zu erklärenden Ereignis bildet. Und beide Schüsse sind offensichtlich Kandidaten für die Rolle des Auslösers der entscheidenden Kausalkette und damit für den Titel der Ursache dieses Ereignisses – in einem der vielen Bedeutungen dieses Begriffs. Deshalb lässt sich sinnvoll die Frage stellen, welches der Ereignisse denn nun die Ursache ist, oder, wenn es angemessener erscheint, beide Ereignisse als Teil einer Gesamtursache zu behandeln, wie groß ihr Anteil an dieser jeweils ist. Das ist damit gemeint, dass es sich um potentiell konkurrierende Ursachen handelt. Um das Problem der mentalen Verursachung zu lösen oder aufzulösen, müssen wir uns jetzt genauer klarmachen, warum es eine entsprechende Konkurrenzsituation zwischen Bewusstseinsereignissen und ihren neuronalen Korrelaten nicht gibt und wie der Eindruck entstehen kann, dass es sie doch gibt.

### **6.3 (Auf)lösung des Problems der mentalen Verursachung**

Ich möchte nun zeigen, dass sich das Problem der mentalen Verursachung lösen oder auflösen lässt, wenn man die hier entwickelte Auffassung der Bewusstseins-Gehirnbeziehung als Kompositions- oder Mikrostrukturbeziehung akzeptiert. (Zur Frage, inwiefern es sich bei der in dieser Arbeit vorgeschlagenen Reaktion auf das Leib-Seele-Problem um eine Lösung oder eine Auflösung desselben handelt, komme ich in der Schlussbetrachtung.) Der erste Schritt besteht darin zu sehen, dass die Vorstellung einer Konkurrenz um Kausalkräfte zwischen Entität und ihrer Struktur, oder dem, woraus sie besteht, unsinnig ist. Dies scheint mir ziemlich offensichtlich zu sein. Die Struktur gehört ja zu dieser Entität, auch wenn sie nicht mit ihr identisch ist. Die Entität *hat* diese Struktur, und deshalb können wir mit Bezug auf diese Struk-

tur die Kausalität dieser Entität erklären, ohne dass deshalb das Explanans dem Explanandum die Kausalkräfte sozusagen wegnimmt.

Der zweite Schritt besteht darin, sich klar zu machen, dass die Erklärung des Auftretens eines Bewusstseinsphänomens durch ein Korrelationsgesetz, das dieses Bewusstseinsphänomen mit seiner Mikrostruktur oder bestimmten Eigenschaften dieser Struktur verbindet, von wesentlich anderer Art ist als die Erklärungen, die wir oben als mechanistische bezeichnet hatten, also Erklärungen, die ein Ereignis unter Bezug auf eine Kausalkette zeitlich vorausgehender Ereignisse erklären. Mechanistische Erklärungen laufen im Wesentlichen *innerhalb* der verschiedenen Mikro- und Makroebenen ab. (Auf den Fall, dass wir innerhalb einer Erklärung die Ebenen wechseln, komme ich später noch zu sprechen.) D.h. wir erklären physikalische Ereignisse durch ihre physikalischen Antezedentien, neuronale durch ihre neuronalen Antezedentien usw.. Die Bewusstseinssebene ist dabei wesentlich Teil einer *common sense*-Makroebene, auf der wir sowohl Bewusstseinsphänomene als auch die Phänomene der manifesten physischen Welt erklären. So erklären wir dann etwa die Tatsache, dass ein Mann freudig erregt ist damit, dass ihm eine schöne Frau auf der Strasse zugelächelt hat, oder die Tatsache, dass er nun beschwingt davon schreitet oder sie anspricht, durch diese freudige Erregung. Die Erklärung seiner Freude durch die Tatsache, dass sein Gehirn sich in einem Zustand befindet, der gesetzesmäßig mit Freude verknüpft ist, ist ein ganz anderer Typ von Erklärung. Diese Erklärungen sind keine Kausalerklärungen, jedenfalls nicht im Sinn einer mechanistischen Erklärung. Dies zeigt sich daran, dass das Verhältnis zwischen dem Bewusstseinszustand und seinem neuronalen Korrelat ein synchrones Verhältnis ist, kein diachrones wie bei mechanistischen Erklärungen und daran, dass die Erklärung in beiden Richtungen möglich ist. Weil Korrelationsgesetze bikonditionale Gesetze sind, kann man auch den Gehirnzustand durch die Freude erklären. Die Form der Erklärung ist in beiden Fällen einfach die Subsumption unter ein Gesetz.

Der dritte Schritt zur Auflösung des Problems ist schließlich die Einsicht, dass der Anschein eines Konflikts zwischen der Kausalität des Bewusstseins und dem Prinzip der kausalen Geschlossenheit des Physikalischen auf einer Unklarheit oder Äquivokation bezüglich des Begriffs des Physischen beruht (vgl. Sturgeon 2000). Ich hatte dies schon eingangs angedeutet und bei der Formulierung dieses Prinzips und an anderen Stellen sowohl den Begriff des Physischen als auch den des Physikalischen benutzt. Um die Illusion eines Konflikts zu beheben, müssen wir nur

den Bereich des Physikalischen, insbesondere den Bereich des Mikrophysikalischen, klar von den Bereichen der Chemie, Geologie, Biologie und anderer so genannter *special sciences*, sowie von der Alltagsebene manifester physischer Objekte und Eigenschaften abzugrenzen. Praktischerweise gibt uns das Deutsche mit seiner Unterscheidung des Physikalischen vom Physischen, die es ja etwa im Englischen gar nicht gibt, auch bereits die Terminologie an die Hand, um diese Unterscheidung zu markieren. Ich werde also im Folgenden als „physikalisch“ nur die Entitäten bezeichnen, die vom theoretischen Spezialvokabular der Physik denotiert werden, während „physisch“ als Oberbegriff für die Entitäten verstanden werden kann, die zu den anderen genannten Ebenen gehören. Dabei ist es an sich nicht wesentlich, einen Begriff zu haben, der so heterogene Entitäten wie z.B. Gold und Wasserstoff, Kontinentalplatten, Wirbelstürme, Buchen und Beine umfasst und sie von den Entitäten der Physik auf der einen und psychologischen, soziologischen und kulturellen Phänomenen auf der anderen Seite abgrenzt. Tatsächlich wäre es vermutlich leichter, die eine oder andere Verwirrung zu vermeiden, wenn wir einen solchen Begriff nicht hätten. Aber da er nun einmal existiert, scheint mir die vorgeschlagene Deutung noch die beste zu sein. Wesentlich ist nur, dass man den Bereich des Physikalischen klar von den verschiedenen Bereichen des Physischen unterscheidet – egal ob man diese nun wiederum unter dem Begriff des Physischen zusammenfasst oder nicht.

Dies ist deshalb entscheidend, weil das Prinzip der kausalen Geschlossenheit nur auf der physikalischen Ebene gilt, nicht allgemein auf der physischen. D.h. nur im Bereich des Physikalischen kann man eine vollständige Kausalerklärung angeben (wenigstens ‚im Prinzip‘), ohne diesen Bereich zu verlassen. Wie eingangs schon erwähnt wurde, kann eine vollständige Kausalerklärung z.B. der biologischen Evolution dagegen nicht im Bereich des Biologischen verbleiben. Der Bereich des Biologischen ist kein geschlossenes System, sondern vielfältigen äußeren, z.B. kosmologischen Einflüssen unterworfen. Sogar die bloße Existenz des Lebens auf der Erde hängt kausal von der Sonne ab. Offensichtlich kann der Bereich des Physikalischen nur deshalb ein kausal geschlossenes System sein, weil er räumlich und zeitlich so umfassend ist. Alle raumzeitlichen Objekte bestehen aus physikalischen Mikroentitäten. Deshalb gibt es auf der physikalischen Ebene buchstäblich keine *äußeren* Einflüsse. Nichts liegt außerhalb des Bereichs des Physikalischen. Die Physik ist in der Lage, die Welt raumzeitlich sozusagen vollständig zu durchdringen, weil alle makroskopischen Entitäten letztlich aus physikalischen Mikroentitäten bestehen. Die

Physik legt ein begriffliches Netz aus, das bis in den letzten Winkel des Universums reicht. Die Umfassendheit der Physik bedeutet auch, dass es kein Kausalgeschehen auf der Makroebene gibt, das nicht mit physikalischer Kausalität *einhergeht*. Es gibt keine Veränderungen auf irgendeiner der Makroebenen ohne Veränderungen auf den Ebenen, die relativ zu dieser Ebene Mikroebenen sind, und die mikrophysikalische Ebene ist die unterste Ebene überhaupt. (Der Bereich der Chemie z.B. ist eine Mikroebene relativ zur biologischen und eine Makroebene relativ zur physikalischen Ebene.)

Der Bereich des Physischen ist nicht kausal geschlossen, weil es mentale Verursachung gibt. Wenn wir aber nun das Physische mit dem Physikalischen gleichsetzen, scheint es, dass die Kausalität des Bewusstseins unmittelbar mit dem Prinzip der kausalen Geschlossenheit des Physikalischen konfligiert. Denn wie wir schon herausgearbeitet haben, sind die Erklärungen auf der *common sense*-Makroebene des Bewusstseins und der manifesten physischen Welt mechanistische Erklärungen. Und mechanistische Erklärungen physikalischer Phänomene durch Bewusstseinsphänomene würden tatsächlich mit dem Prinzip der kausalen Geschlossenheit des Physikalischen im Widerspruch stehen. Wenn wir dem Bewusstsein die Fähigkeit zuschreiben würden, Veränderungen auf der physikalischen Ebene zu *verursachen*, könnten die physikalischen Erklärungen des physikalischen Geschehens nicht vollständig sein. Das Bewusstsein verursacht aber keine Veränderungen auf der physikalischen, sondern auf der physischen Ebene. Diese Veränderungen gehen mit physikalischen Veränderungen einher, weil sowohl die mentalen als auch die physischen Entitäten eine physikalische Mikrostruktur haben und nichts sich verändern kann, ohne dass seine Mikrostruktur sich verändert. Aber weder die mentalen noch die physischen Entitäten verursachen Veränderungen auf der physikalischen Ebene. Deshalb ist es wichtig, dass man die Kausalität des Bewusstseins so versteht, dass sie sich auf physische, nicht auf physikalische Phänomene bezieht. Wenn man „physisch“ so versteht, dass es mit „physikalisch“ synonym ist, ist unsere Prämisse (1) der kausalen Wirksamkeit des Bewusstseins tatsächlich falsch. Aber daraus folgt nicht, dass sie falsch ist in dem Sinn, in dem sich „physisch“ auf die manifeste physische Welt bezieht. Und nur auf die in diesem Sinn physische Welt beziehen wir uns ja, wenn wir etwa die Handlungen eines Menschen durch seine Absichten erklären.

So kann also eine Äquivokation bezüglich des Begriffs des Physischen zu dem Eindruck führen, dass die Kausalität des Bewusstseins mit der kausalen Ge-



geschlossenheit des Physikalischen konfligiert. Diese Äquivokation kommt aber natürlich nicht von ungefähr. Sie ist Ausdruck der Tendenz, die manifeste physische Welt zugunsten der physikalischen zu eliminieren. Wenn man dies tut, ist das Problem der mentalen Verursachung tatsächlich unlösbar. Denn dann muss die Kausalität des Bewusstseins sich unmittelbar auf die physikalische Ebene richten, weil es keine manifesten Entitäten mehr gibt, auf die es wirken kann. Wir haben gezeigt, wie dies vermieden werden kann.

Um letzte Zweifel zu beseitigen und den Einwand zu beantworten, dass auch diese Konzeption zu kausaler Überdetermination führt, wollen wir ein Beispiel mentaler Verursachung mitsamt den dazu gehörigen Verhältnissen auf der Mikroebene betrachten noch eingehender betrachten. Ein mentales Ereignis  $m$  (Erkennen des Bekannten) löst eine Handlung  $h$  aus (Grüßen durch Armheben). Ursache und Wirkung bestehen wiederum in einer Vielzahl von physiologischen Mikroereignissen, diese wiederum in Ereignissen auf der molekularen Ebene usw., bis hin zur untersten Mikroebene (wenn es eine unterste Ebene gibt). Bezeichnen wir die physiologischen Korrelate des mentalen Ereignisses als  $Km_1, Km_2, \dots Km_n$  und entsprechend die physiologischen Korrelate der Handlung (die selbst wiederum natürlich schon auf der Makroebene in verschiedene Komponenten zerlegt werden kann) als  $Kh_1, Kh_2, \dots Kh_n$ .

Der Hauptunterschied zwischen dieser und vielen anderen Darstellungen des gleichen Typs besteht darin, dass hier die Tatsache, dass es sich um Mikro-Makrobeziehungen handelt, stark hervorgehoben wird und jeder Makroentität eine Vielzahl von Mikroentitäten zugeordnet wird, nicht jeweils bloß eine. Dies ergibt sich aus der Natur der Mikro-Makrobeziehung, und deshalb ist auch die Kompositionsbegrifflichkeit angemessen. Diese wiederum beinhaltet auch, dass wir die jeweilige Gesamtheit der mit der Makroentität korrelierten Mikroereignisse, also das entsprechende Aggregat, als eine Einheit betrachten können. Dabei sollten wir uns aber darüber im Klaren sein, dass diese Einheit in gewissem Sinne eine Projektion von der Makro- auf die Mikroebene darstellt. D.h. wenn wir die Mikroebene ausschließlich durch die Brille der ihr eigentlich angemessenen (also z.B. physiologischen oder physikalischen) Begrifflichkeit betrachteten, würde das Korrelat nicht unbedingt als Einheit sichtbar werden. Entsprechend sind die primären Kausalbeziehungen auf der Mikroebene erst einmal die zwischen den einzelnen Mikroereignissen.

Konzentrieren wir uns nun auf die Handlung  $h$ . Wir nehmen an, dass sie von  $m$  verursacht wird und wir können auch annehmen, obwohl diese Annahme vielleicht nicht unbedingt notwendig ist, dass es ein entsprechendes Kausalgesetz gibt, mittels dessen das Auftreten von  $h$  aus dem Auftreten von  $m$  abgeleitet werden kann. Weiterhin nehmen wir an, dass  $h$  durch ein Korrelationsgesetz mit seinen Mikrobestandteilen verbunden ist. Entsprechend kann das Auftreten von  $h$  also aus dem Vorliegen von  $Kh_1 \dots Kh_n$  abgeleitet werden. Ist  $h$  damit kausal überdeterminiert, so dass, gegeben unsere Prämisse (4), wieder fraglich wird, ob  $k$  tatsächlich durch  $m$  verursacht wird? Die Antwort ist ein klares „nein“, denn  $h$  lässt sich zwar aus dem Korrelationsgesetz ableiten, aber wie wir schon herausgearbeitet haben, ist dieses kein Kausalgesetz, da die Beziehung zwischen  $h$  und  $Kh_1 \dots Kh_n$  nicht kausal ist, und deshalb liegt auch keine kausale Überdeterminiertheit vor.

Man könnte nun behaupten, dass zwar keine kausale Überdetermination vorliegt, aber doch irgendwie Überdetermination, da das gleiche Explanandum auf zweierlei Weise abgeleitet und erklärt werden kann. Wer dies behauptet, muss aber die Beweislast für ein Prinzip schultern, das besagt, dass es nicht zwei Erklärungen verschiedenen Typs für ein Phänomen geben kann. Wir haben oben gesehen, dass die Vorbehalte gegenüber kausaler Überdetermination zu Recht bestehen. Die Vorstellung, dass es zwei unabhängige und doch jeweils für sich hinreichende Ursachen für ein und dasselbe Phänomen geben soll, macht tatsächlich keinen Sinn. Die Behauptung, dass es überhaupt nur eine Erklärung, gleich welcher Art, für ein Phänomen geben kann, scheint aber ziemlich willkürlich zu sein. Ich sehe keinen Grund sie zu akzeptieren. Man könnte schließlich an dieser Stelle auch mit Recht darauf insistieren, dass das Faktum der Bewusstseinskausalität viel gewisser ist als jedes Prinzip dieser Art, so dass die Tatsache, dass es diese Kausalität in Frage stellt, schon ein hinreichender Grund wäre, es zurückzuweisen.

Aber können, müssen wir nicht sagen, dass auch die neuronalen Korrelate von  $m$  die Handlung  $h$  verursachen? Sind sie nicht allein kausal hinreichend dafür, dass der Arm sich zum Gruß hebt? Dies ist der letzte Pfeil im Köcher des Gegners, der behauptet, dass mentale Verursachung zu kausaler Überdetermination führt. Er verweist auf die (vermeintliche) Kausalbeziehung zwischen  $Km_1, Km_2, \dots Km_n$  und  $h$ . Diese Beziehung ist intuitiv gesehen ein besserer Kandidat für eine Kausalbeziehung als die Beziehungen zwischen  $m$  und  $h$  und ihren jeweiligen Korrelaten, weil  $Km_1, Km_2, \dots Km_n$  und  $h$  nicht in einer Kompositionsbeziehung stehen und entsprechend

zeitlich und räumlich voneinander getrennt sind. Ich möchte denn auch nicht abstreiten, dass es sich um eine Kausalbeziehung handelt und werde diese Art von Kausalität im Folgenden als „diagonale“ Kausalität bezeichnen. (Dies ist aber eine rein terminologische Frage.) Dabei kann es sich einmal um eine Beziehung zwischen den Korrelaten des Bewusstseinszustands und der Handlung und damit um „Aufwärtsverursachung“ von der Mikro- zur Makroebene handeln. Der Kritiker kann aber auch mittels der Beziehung zwischen  $m$  und  $Kh_1, Kh_2, \dots Kh_n$  argumentieren und damit mit „Abwärtsverursachung“. Auf diese Weise würde er zu zeigen versuchen, dass  $Kh_1, Kh_2, \dots Kh_n$  überdeterminiert ist. Oder er könnte aufgrund der Tatsache, dass man von  $Kh_1, Kh_2, \dots Kh_n$  über weitere Kompositionsbeziehungen schnell bis zur mikrophysikalischen Ebene kommt, behaupten, dass diese Abwärtsverursachung den Lauf der Dinge auf der physikalischen Ebene durcheinander bringt und damit direkt mit der kausalen Geschlossenheit des Physikalischen konfligiert.

Egal ob nun mit Aufwärts- oder Abwärtsverursachung argumentiert wird, die Antwort ist in beiden Fällen folgende. Zwar ist im Prinzip nichts gegen die Rede von diagonalen Kausalbeziehungen einzuwenden. Wir gebrauchen auch tatsächlich oft Erklärungen, die sich aufwärts oder abwärts zwischen den verschiedenen Mikro- und Makroebenen bewegen. Es trifft aber nicht zu, dass damit weitere *unabhängige* Ursachen und Kausalketten ins Spiel gebracht werden. Die Beziehungen zwischen  $m$  und  $Kh_1, Kh_2, \dots Kh_n$  und zwischen  $h$  und  $Km_1, Km_2, \dots Km_n$  sind von den bereits diskutierten Kausalbeziehungen innerhalb der Ebenen und von den Kompositions- und Korrelationsbeziehungen zwischen den Ebenen nicht unabhängig. Dass die Aufwärts- und Abwärtsbeziehungen bestehen, folgt unmittelbar aus dem Bestehen der anderen Beziehungen. Dass  $Km_1, Km_2, \dots Km_n$  die Handlung  $h$  diagonal verursacht, bedeutet nur, dass  $Km_1, Km_2, \dots Km_n$  die Handlungskorrelate  $Kh_1, Kh_2, \dots Kh_n$  verursacht und diese eben durch ein Korrelationsgesetz mit der Handlung verbunden sind. Hier liegen keine weiteren unabhängigen Kausalketten oder -kräfte vor. Von Überdetermination könnte aber sinnvoller Weise nur die Rede sein, wenn es um zwei voneinander unabhängige Ursachen des gleichen Phänomens ginge. Dies ist hier nicht der Fall. Ich komme deshalb zu dem Schluss, dass das Überdeterminiertheitsargument scheitert und mentale Verursachung in dem hier skizzierten Rahmen verständlich gemacht werden kann. Wir müssen nicht aus theoretischer Verzweiflung versuchen, entweder der Identitätstheorie, dem Epiphänomenalismus oder der Überdeterminiertheitsthese Sinn abzugewinnen.

## 7 Schlussbetrachtung: Lösung oder Auflösung?

*Was mich an den Philosophen am meisten abstößt, ist der Entleerungsprozess ihres Denkens. Je häufiger und geschickter sie ihre Grundworte anwenden, umso weniger bleibt von der Welt um sie übrig. Sie sind wie Barbaren in einem hohen, geräumigen Haus voll von wunderbaren Werken. Sie stehen in Hemdsärmeln da und werfen methodisch und unbeirrbar alles zum Fenster hinaus, Sessel, Bilder, Teller, Kinder, Tiere, bis nichts übrig ist als ganz leere Räume. Manchmal kommen zuletzt die Türen und Fenster nachgeflogen. Das nackte Haus bleibt stehen. Sie bilden sich ein, dass es um diese Verwüstungen besser steht.*

– Elias Canetti, *Alle vergeudete Verehrung*

*Die Philosophie läßt alles, wie es ist.*

– Ludwig Wittgenstein, *Philosophische Untersuchungen*

Die kausale Geschlossenheit der physikalischen Welt zwingt uns nicht zu einer der fundamentalistischen Positionen. Sie gibt uns vielmehr einen guten empirischen Grund, das Bewusstsein, so wie ich es vorgeschlagen habe, als ein *Makrophänomen* unter anderen, eine raumzeitliche Entität mit einer physiologischen und letztlich auch physikalischen Mikrostruktur, zu rekonzeptualisieren und damit den Bewusstseins- und den Physikfundamentalismus gleichermaßen hinter uns zu lassen. Ich möchte nun noch einmal zusammenfassend darlegen, warum mir dies die vernünftigste Reaktion auf das Leib-Seele-Problem zu sein scheint und werde dann abschließend kurz sagen, warum wir es hier eher mit einer Auflösung des Leib-Seele-Problems als einer Lösung im gewöhnlichen Sinne zu tun haben.

Erinnern wir uns zuerst an das in Kapitel diskutierte einfache Beispiel einer empirisch motivierten Rekonzeptualisierung oder Redefinition, das Hesperus/Phosphorus-Beispiel. Ich habe versucht herauszuarbeiten, dass Hesperus und Phosphorus, anders als es die gesamte philosophische Debatte unterstellt, in unserer tatsächlichen sprachlichen Praxis gar nicht identifiziert wurden. Stattdessen wurde der Name „Venus“ für den Planeten eingeführt und Hesperus und Phosphorus als Planetenstadien rekonzeptualisiert und damit – wenn man so will – in ihrem ontologischen Status herabgestuft. Dies erschien auch als zweckmäßig und vernünftig, weil es witzlos ist, zwei Namen für denselben Gegenstand zu haben und weil es so möglich war, die Übereinstimmung mit unseren früheren Selbsten zu maximieren.

Gleichzeitig aber erlitt dieses Vokabular damit einen Bedeutungsverlust und verschwand in einige Randbereiche der Sprache.

Die so genannten wissenschaftlichen Identifikationen von manifesten physischen Phänomenen wie Wasser, Wärme und Licht können in ähnlicher Weise gedeutet werden. Die philosophische Tradition auf ihrer Suche nach den letzten, einzig realen, fundamentalen Elementen der Realität, ihrem Wesen, interpretierte die wissenschaftlichen Resultate als Entdeckungen eines solch metaphysischen Wesens und eliminierte die manifeste physische Welt. Aber auch hier erscheint eine konservative Reaktion als die vernünftige. Naturwissenschaftliche Erkenntnisse haben gezeigt, dass die manifeste Makrowelt eine der naiven Weltansicht verborgene Struktur hat, dass Wärme keine Substanz ist und Wasser, Feuer, Erde und Luft keine Elemente sind. Aber dies bedeutet nicht, dass der naive Wahrnehmungsrealismus über Bord geworfen werden muss. Dieser erfordert nicht, dass diese Phänomene irreduzibel einfache Substanze oder Elemente sind, sondern nur, dass die Wahrnehmung uns einen Zugang zu diesen Phänomenen eröffnet. Wir sind also frei, sie als physische Makrophänomene mit verborgener physikalischer Struktur zu rekonzeptualisieren. Auch in diesem Fall ist das mit einem metaphysischen Statusverlust verbunden – einer Herabstufung von Elementen oder Substanzen zu strukturierten Makrophänomenen.

Entsprechend ist die metaphysische Herabstufung des Gegensatzes zwischen Körper und Geist die angemessene Reaktion auf die wissenschaftlichen Erkenntnisse, die im Prinzip der kausalen Geschlossenheit des Physikalischen ihren Ausdruck finden. Dieses besagt letztlich nur, dass die gesamte raumzeitliche Welt eine physikalische Mikrostruktur hat. Deshalb gibt es hier buchstäblich keine äußeren Einflüsse auf die physikalische Ebene. Dabei ist die Aussage, dass die raumzeitliche Welt eine mikrophysikalische Struktur hat, durchaus als eine *physikalische* Aussage anzusehen. Dies ist etwas, was Physiker kraft der Autorität ihres Fachs behaupten können und behaupten und, wie ich herauszuarbeiten versucht habe, scharf zu trennen von der metaphysischen Behauptung des materialistischen Philosophen, die Welt sei mit der physikalischen Welt identisch. Die physikalische Aussage ist meiner Ansicht nach gleichermaßen mit einem naiven Realismus bezüglich der manifesten physischen Welt als auch mit einem entschiedenen Realismus in Bezug auf das Bewusstsein vereinbar, während die Identitätsbehauptung, wie ich ausführlich zu zeigen versucht habe, entweder bloß ein verkappter Eliminativismus ist, oder zu der unverständlichen

Konzeption bloß subjektiv, aber dennoch kognitiv verschiedener Zugänge zur selben Entität führt bzw. zu einem ebenfalls unverständlichen Konzept opaker Bezugnahme. Die physikalische Aussage macht es aber vernünftig, die hier vertretene Konzeption von Bewusstsein als Makrophänomen mit physiologischer und letztlich auch physikalischer Mikrostruktur zu akzeptieren. Wenn das Bewusstsein Teil der raumzeitlichen Welt ist, muss es aus heutiger wissenschaftlicher Sicht heraus eine physikalische Mikrostruktur haben.

Ist es wirklich eine Alternative, Bewusstsein als ein radikal nichträumliches Epiphänomen zu konzipieren? Der Epiphänomenalismus setzt sich fundamental in Widerspruch zu unserem natürlichen Selbstverständnis als handelnde und in den Lauf der Welt eingreifende Wesen und kann den Überlegungen des letzten Kapitels zufolge auch die innermentale Kausalität nicht verständlich machen. Und wie im dritten Kapitel argumentiert wurde, kann die radikale Skepsis bezüglich der Räumlichkeit des Bewusstseins den Zusammenhang zwischen der Person oder ihrem Körper und dem Bewusstsein nicht mehr verständlich machen. Wie kann etwas radikal Nichträumliches der Zustand von etwas wesentlich Räumlichem sein? Kann man sich so die Zuschreibung von Bewusstseinsphänomenen an bestimmte menschliche und tierische Individuen denken? Der dualistische Epiphänomenalismus verfängt sich also in metaphysischen Aporien und dies gilt auch, wie schon in der Einleitung gezeigt wurde, für andere bewusstseinsfundamentalistische Positionen wie etwa den so genannten Russellschen Monismus, der zum Panpsychismus führt.

Schließlich sollte man sich fragen, warum man überhaupt versuchen soll, solchen metaphysischen Extrempositionen Sinn abzugewinnen, wenn die Tatsache, dass die Welt eine mikrophysikalische Struktur hat, doch mit der Realität des Bewusstseins gar nicht im Widerspruch steht und die Mikrostrukturkonzeption der Bewusstsein-Gehirnbeziehung, wie ich im letzten Kapitel versucht habe zu zeigen, auch das Problem der mentalen Verursachung lösen bzw. auflösen kann, indem sie klarstellt, dass die Vorstellung, dass eine Entität und ihre Bestandteile um kausale Relevanz konkurrieren, sinnlos ist?

Ein Grund, warum es dennoch schwer sein mag, dies zu akzeptieren, ist der, dass es sich eben tatsächlich um eine Auflösung des Problems handelt und nicht um eine Lösung im konventionellen Sinn. Man mag das Gefühl haben, dass der hier gemachte Vorschlag sozusagen gegen die Regeln des Problems verstößt. Und dies ist auch tatsächlich in gewissem Sinne der Fall.

Ausgangspunkt dieser Arbeit war die Überlegung, dass die Elimination der manifesten Welt notwendigerweise zu dem Dilemma führen muss, das uns als Leib-Seele-Problem bekannt ist. Denn die Lokalisierung des verbleibenden ‚qualitativen Überschusses‘ im Bewusstsein führt zu einem metaphysischen Sonderstatus desselben, der sich wissenschaftlich nicht rechtfertigen lässt. Vor diesem Hintergrund erscheint das Bewusstseinsrätsel. Wie es ist möglich, dass ausgerechnet die weiche Materie des Gehirns mit einem solchen qualitativen Reichtum verbunden ist? Woher kommt es, dass wir anscheinend die übrige Welt so vollkommen transparent erklären können, uns dies beim Bewusstsein aber nicht gelingen will? Im zweiten Kapitel habe ich ausführlich zu zeigen versucht, dass sich dieses Staunen nicht in eine beantwortbare Frage transformieren lässt. Es kann nur zu höchst spekulativen Theorien führen, die jeder empirischen Grundlage entbehren und dem Anspruch, den sie an eine Erklärung des Bewusstseins stellen, selbst nicht genügen können. Denn das Staunen über den qualitativen Reichtum des Bewusstseins und die vollkommen transparente Erklärung der physischen Welt, die uns scheinbar gelungen ist, ist nur ein Artefakt der tatsächlichen Elimination dieser Welt. Diese kreiert eine fundamentale metaphysische Asymmetrie zwischen dem Bewusstsein und der manifesten physischen Welt, die sich wissenschaftlich nicht rechtfertigen lässt und nur zu verschiedenen aporetischen metaphysischen Theoriebildungen führt. Weil dies so ist, ist es vernünftig, die stillschweigend akzeptierte Grundregel des Problems, die darin besteht, die Elimination der manifesten Welt zu akzeptieren, als solche kenntlich zu machen, sie zu kritisieren und letztlich zu verabschieden. Wenn man das tut, verschwindet das Problem. Es löst sich auf.

Wie Austen Clark bemerkt hat (vgl. Zitat S. 4), erleichterte die Elimination der manifesten Welt zu einem gewissen Zeitpunkt den Physikern ihre Forschungsarbeit, da sie sie von Phänomenen befreite, mit denen sie in ihrer Theoriebildung nichts anfangen konnten. In ähnlicher Weise hat wohl das *age of avoidance* eine Zeit lang den Psychologen, Kognitions- und Neurowissenschaftlern die Arbeit erleichtert, in dem es ihnen die störenden Bewusstseinsphänomene vom Hals schaffte, mit denen sie in diesem Stadium ihrer Forschung nichts anfangen konnten. Jetzt, da dieses Zeitalter endlich im Wesentlichen sein Ende erreicht zu haben scheint, haben beide metaphysische Konstruktionen jedwede Nützlichkeit verloren und ich schlage deshalb vor, sie beide gleichermaßen und damit auch das Leib-Seele-Problem in seiner traditionellen Form hinter uns zu lassen.

## 8 Literatur

- Beckermann, A., Flohr H. & Kim, J. (Hrsg.) (1992) *Emergence or Reduction Essays on the Prospects of Nonreductive Materialism*, Berlin, New York.
- Bieri, Peter (1981) „Generelle Einführung“, in Bieri (Hrsg.) (1981).
- Bieri, Peter (Hrsg.) (1981) *Analytische Philosophie des Geistes*, Königstein.
- Bieri, Peter (1992) „Trying out Epiphenomenalism“, *Erkenntnis*.
- Bieri, Peter (1996) „Was macht Bewußtsein zu einem Rätsel?“, in Metzinger (Hrsg.) (1997)
- Block, N. (Hrsg.) (1980) *Readings in the Philosophy of Psychology, Volumes 1 and 2*, Cambridge, MA.
- Block, N., Flanagan, O. & Güzeldere, G. (Hrsg.) (1997) *The Nature of Consciousness*, Cambridge MA.
- Block, N. (im Erscheinen) “Max Black's objection to mind-body identity”, *Oxford Review of Metaphysics* 3. Online zugänglich unter:  
<http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/MaxBlack2clean.pf>
- Borst, C.V. (Hrsg.) (1970) *The Mind-Brain Identity Theory*, London.
- Burge, Tyler (1993) „Mind-Body Causation and Explanatory Practice“, in John Heil & Alfred Mele (Hrsg.) (1993) *Mental Causation*, Oxford.
- Burt, Edwin Arthur (1954) *The Metaphysical Foundations of Modern Science*, New York.
- Campbell, John (1997) “A Simple View of Colour”, in Alex Byrne & David R.Hilbert (Hrsg.) *Readings on Color. Volume 1: The Philosophy of Color*, Cambridge MA.
- Canetti, Elias (1970) *Alle vergeudete Verehrung. Aufzeichnungen 1949-1960*, München.
- Carruthers, P. (2003) “Phenomenal concepts and higher-order experiences”, *Philosophy and Phenomenological Research*.
- Chalmers, David (1996) *The Conscious Mind*, Oxford.
- Chalmers, David (2002) “Consciousness and its Place in Nature”, in D. Chalmers (Hrsg.) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford.
- Chalmers, David (im Erscheinen) “Phenomenal Concepts and the Explanatory Gap”, in *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Torin Alter and Sven Walter (Hrsg.), Cambridge MA. Online zugänglich unter <http://consc.net/papers/pceg.html>



- Clark, Austen (1993) *Sensory Qualities*, Oxford.
- Crane, T. & Mellor, D.H. (1990) "There is no question of physicalism", *Mind* 99, 185-206.
- Crick, Francis (1994) *The Astonishing Hypothesis: The Scientific Search for the Soul*, New York.
- Crick, Francis & Koch, Christoph (1997) "Towards a Neurobiological Theory of Consciousness", in Block et. al. (Hrsg.) (1997): 277-292.
- Crick, Francis & Koch, Christoph (2003) "A framework for consciousness", *Nature Neuroscience* 6: 119-26.
- Damasio, Antonio R. (1997) *Descartes' Irrtum*, München.
- Davidson, Donald (1980) „Mental Events“, in *Essays on Actions and Events*, Oxford: 207-25.
- Dretske, Fred (1995) *Naturalizing the Mind*, Cambridge MA.
- Dubois-Reymond, Emil (1916) *Über die Grenzen der Naturerkenntnis*, Leipzig.
- Dummett, Michael (1973) *Frege: The Philosophy of Language*, London.
- Edelman, Gerald M. & Tononi, Giulio A *Universe of Consciousness*, New York.
- Field, H. (1980), "Mental Representation", in Block (Hrsg.) (1980).
- Feigl, Herbert (1958) *The "Mental" and the "Physical"*, Minneapolis.
- Feyerabend, Paul (1963) "Comment: 'Mental events and the brain'", in Borst (Hrsg.) (1970): 140-41.
- Flanagan, Owen (1992) *Consciousness Reconsidered*, Cambridge MA.
- Flohr, Hans (1992) *Qualia and Brain Processes*, in Beckermann et. al. (Hrsg.) (1992).
- Frege, Gottlob (1986) „Über Sinn und Bedeutung“, in Günther Patzig (Hrsg.) *Funktion, Begriff, Bedeutung*, Göttingen: 40-65.
- Freud, Sigmund (1940) „Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse“, in *Werke*, Bd. 15, S.76f, London und Frankfurt 1940ff.
- Gillett, C. & Loewer, B. (Hrsg.) (2001) *Physicalism and its Discontents*, Cambridge.
- Güzeldere, Güven (1997) "Introduction", in Block et al. (Hrsg.) (1997): 1-67.
- Goldmann, Alvin (1997) *Consciousness, Folk Psychology and Cognitive Science*, in Block et al. (Hrsg.) (1997): 111-26.
- Goldstrasz, Thomas (2002) *Das Problem der universellen Realisierbarkeit bei Putnam und Searle*, Magisterarbeit Humboldt-Universität Berlin.
- Haggard, Patrick (2005) "Conscious Intention and Motor Cognition", *Trends in Cognitive Science*, Vol. 9, No. 6.

- Hill, C.S. (1997) "Imaginability, conceivability, possibility, and the mind-body problem", *Philosophical Studies* 87: 61-85.
- Hill, C. S. & McLaughlin, B. P. (1998) "There are fewer things in reality than are dreamt of in Chalmers' philosophy", *Philosophy and Phenomenological Research*.
- Hintikka, J. (1956) "Identity, Variables and Impredicative Definitions", *Journal of Symbolic Logic*, 21: 225-45.
- Huxley, T.H. (1866) *Lessons in Elementary Physiology*, London.
- Ismael, J. (1999) "Science and the phenomenal", *Philosophy of Science* 66: 351-69.
- Jackson, Frank (1982) "Epiphenomenal Qualia", *Philosophical Quarterly*, XXXII: 5, 127-136.
- Jackson, Frank (1997) "Finding the Mind in the Natural World", in Block et al. (Hrsg.) (1997): 483-492.
- Jackson, Frank (1998) *From Metaphysics to Ethics*, Cambridge.
- Jaynes, J. (1976) *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Boston.
- Kim, Jaegwon (1997) "Does the Problem of Mental Causation Generalize?", *Proceedings of the Aristotelian Society*, 281-96.
- Kim, Jaegwon (1998) *Mind in a Physical World*, Cambridge MA.
- Kripke, Saul (1980) *Naming and Necessity*, Cambridge MA.
- Lakoff, George (1987) *Women, Fire and Dangerous Things*, Chicago.
- Leibniz, G.W. (1720) *Monadologie*, Hamburg: Meiner.
- Levine, Joseph (1997) "On leaving out what it's like", in Block et al. (Hrsg.) (1997): 543-55.
- Levine, Joseph (1983) "Materialism and qualia: The explanatory gap", *Pacific Philosophical Quarterly*, 64, 354-61.
- Loar, Brian (1997) "Phenomenal States", in Block et al. (Hrsg.) (1997).
- Lockwood, Michael (1989) *Mind, Brain, and the Quantum*, Oxford.
- Lycan, William G. (1996) *Consciousness and Experience*, Cambridge MA.
- Malcom, Norman (1970) *Scientific Materialism and the Identity Theory*, in Borst (Hrsg.) (1970): 171-80.
- Marsh, Robert Charles (Hrsg.) (1964) *Logic and Knowledge*, London.
- McLaughlin, B. P. (2001) "In defense of new wave materialism: A response to Horgan and Tienson", in Gillet & Loewer (Hrsg.) (2001).
- McLaughlin, B. P. (2002) "Color, consciousness, and color consciousness", in Q. Smith &

- A. Jokic (Hrsg.) *Consciousness: New Philosophical Perspectives*, Oxford.
- McGinn, Colin (1995) *Consciousness and Space*, in Metzinger (Hrsg.) (1995).
- McGinn, Colin (1991) *The Problem of Consciousness: Essays Toward a Resolution*, Oxford.
- McGinn, Colin (2001) "How Not to Solve the Mind-Body Problem", in Gillet & Loewer (Hrsg.) (2001).
- Metzinger, Thomas (Hrsg.) (1995) *Conscious Experience*, Paderborn.
- Metzinger, Thomas (Hrsg.) (1997) *Bewußtsein. Beiträge aus der Gegenwartsphilosophie*, 2. Auflage, Paderborn.
- Metzinger, Thomas (2000) „Introduction: Consciousness Research at the End of the Twentieth Century”, in ders. (Hrsg.) *Neural Correlates of Consciousness*, Cambridge MA.
- Nagel, Thomas (1970) "Physicalism", in Borst (Hrsg.) (1970): 214-230.
- Nagel, Thomas (1986) *The View From Nowhere*, Oxford.
- Nagel, Thomas (2004) *What Does It All Mean?*, Oxford
- O'Dea, J. (2002) "The indexical nature of sensory concepts", *Philosophical Papers* 32.
- Pauen, Michael (1999) *Das Rätsel des Bewusstseins*, Paderborn.
- Papineau, David (2001) *The Rise of Physicalism*, in Gillet, Loewer (Hrsg.) (2001).
- Papineau, David (2002) *Thinking about Consciousness*, Oxford.
- Perry, John (2001) *Knowledge, Possibility, and Consciousness*, Cambridge MA.
- Place, U.T. (1956) "Is consciousness a brain process?", in Borst (Hrsg.) (1970): 42-51.
- Place, U.T. (1999) "Token- versus Type-Identity Physicalism", in *Anthropology & Philosophy*, Vol. 3 No. 2.
- Putnam, Hilary (1991) *Repräsentation und Realität*, Frankfurt.
- Quine, W.V.O. (1953) „Two Dogmas of Empiricism“, in *From a Logical Point of view*, Harvard.
- Quine, W.V.O. (1960) *Word & Object*, Cambridge MA.
- Ramsey, Frank P. (1931) "Theories", in *Foundations of Mathematics*, London.
- Rorty, Richard (1965) "Mind-body identity, privacy, and categories", in Borst (Hrsg.) (1970): 187-213.
- Rosenthal, David (1986) „Two Concepts of Consciousness”, *Philosophical Studies* 94: 329-359
- Russell, Bertrand (1918) "The Philosophy of Logical Atomism", in Robert Charles Marsh (Hrsg.) (1964)
- Russell, Bertrand (1905) "On Denoting", in Robert Charles Marsh (Hrsg.) (1964)

- Russell, Bertrand (1927) *The Analysis of Matter*, London.
- Schmitz, Michael (2004) "Consciousness: Central, but not Essential for Intentionality", in Johann Christian Marek & Maria Elisabeth Reicher (Hrsg.), *Erfahrung und Analyse. Beiträge des 27. int. Wittgenstein Symposiums Band XII*, Kirchberg am Wechsel.
- Schmitz, Michael (Manuskript) „Having and Instantiating Properties“.
- Searle, John R. (1983) *Intentionality*, Cambridge MA.
- Searle, John R. (1992) *The Rediscovery of the Mind*, Cambridge MA.
- Seebaß, Gottfried (1993) *Wollen*, Frankfurt am Main.
- Shaffer, Jerome (1961) "Could mental states be brain processes?", in Borst (Hrsg.) (1970).
- Smart, J.J.C. (1959) *Sensations and brain processes*, in Borst (Hrsg.) (1970): 52-66.
- Smart, J.J.C. (1961) "Further remarks on sensations and brain processes", in Borst (Hrsg.) (1970): 113-23.
- Smart, J.J.C. (1963) "Materialism", in Borst (Hrsg.) (1970): 159-70.
- Stevenson, J.T. (1960) "'Sensation and brain processes': A reply to J.J.C. Smart", in Borst (Hrsg.) (1970): 87-92.
- Strawson, Galen (1994) *Mental Reality*, Cambridge.
- Strawson, Galen (2002) "Real Materialism", in *Chomsky and his Critics*, Louise Antony & Norbert Hornstein (Hrsg.), Oxford.
- Stroud, Barry (2000) *The Quest for Reality: Subjectivism and the Metaphysics of Colour*, Oxford.
- Stubenberg, Leopold (2005) "Neutral Monism", *The Stanford Encyclopedia of Philosophy (Spring 2005 Edition)*, Edward N. Zalta (Hrsg.), <http://plato.stanford.edu/archives/spr2005/entries/neutral-monism/>
- Sturgeon, Scott (2000) *Matters of Mind*, London.
- Tye, Michael (1995) *Ten Problems of Consciousness*, Cambridge, MA.
- Tye, Michael (2003) "A theory of phenomenal concepts", in A. O'Hear (Hrsg.) *Minds and Persons*, Cambridge.
- Williams, C. J. F. (1989) *What is Identity?*, Oxford.
- Wilkes, K.V. (1988) "—, Yishi, Duh, Um, and Consciousness", in A.J. Marcel & E. Bisiach (Hrsg.) *Consciousness in Contemporary Science*, Oxford.
- Wittgenstein, Ludwig (1984) *Philosophische Untersuchungen. Tractatus-logico-philosophicus. Werkausgabe Band 1*, Frankfurt am Main.
- Wittgenstein, Ludwig (1984) *Über Gewißheit. Zettel. Werkausgabe Band 8*, Frankfurt am Main.