

11. Können mentale Phänomene neurobiologisch erklärt werden?

(In: Roth, G. und W. Prinz (Hg.) *Kopfarbeit*.
Heidelberg: Spektrum Verlag 1996, 413-425)

Ansgar Beckermann

Was im Kopf eines Menschen vorgeht, das heißt, was die Neurobiologie über die Funktionsweise des Gehirns und des ZNS herausbekommt, interessiert den Philosophen hauptsächlich deshalb, weil er sich von diesen Ergebnissen Fortschritte bei der Lösung eines der ältesten Probleme der Philosophie erhofft – Fortschritte bei der Lösung des Leib-Seele-Problems. Für den Philosophen lautete die zentrale Frage deshalb, ob uns die Neurobiologie helfen kann, das mentale Leben von Menschen und höheren Tieren zu verstehen. In dem Sinne, in dem uns die Physik geholfen hat, das Wesen der Wärme und des Magnetismus zu verstehen; und in dem uns die organische Chemie und die Physiologie geholfen haben, Vorgänge wie Wachstum, Fortpflanzung und Vererbung zu verstehen. Schließlich haben die Neurowissenschaften in den letzten Jahren erhebliche Fortschritte gemacht. Vielleicht ist jetzt also der Zeitpunkt gekommen, an dem es der Neurobiologie gelingt, eines der größten Rätsel der Menschheit einer Lösung näherzubringen, indem sie zeigt, wie mentale Phänomene durch neuronale Prozesse realisiert sind, wie diese Phänomene neurobiologisch erklärt werden können.

Gegen diese optimistische (manche würden wohl eher sagen: pessimistische) Auffassung ist jedoch von Philosophen und Wissenschaftlern immer wieder eingewandt worden, daß die Dinge im Bereich des Mentalen gänzlich anders liegen und daß sich mentale Phänomene daher jeder naturalistischen Erklärung grundsätzlich entziehen. Können die inzwischen immer schneller wachsenden Ergebnisse in der Neuroanatomie und Neurophysiologie die Frage entscheiden helfen, wer in diesem Streit recht hat? Auf diesen Punkt werde ich am Ende dieses Aufsatzes zurückkommen. Zunächst müssen jedoch zwei Vorfragen geklärt werden:

1. Was heißt es eigentlich, mentale Phänomene zu erklären? Unter welchen Bedingungen könnten wir zu recht behaupten, das mentale Leben von Menschen neurobiologisch erklärt zu haben?
2. Welche Gründe sprechen in den Augen der Skeptiker grundsätzlich gegen die Möglichkeit, mentale Phänomene naturalistisch zu erklären?

11.1 Wann wären mentale Phänomene neurobiologisch erklärt?

Wenn man im Hinblick auf diese Frage zu einer angemessenen Antwort kommen will, ist es von entscheidender Bedeutung, zwei grundsätzlich verschiedene Erklärungsprojekte voneinander zu unterscheiden: die *Erklärung von Ereignissen* auf der einen und die *Erklärung von Eigenschaften und Fähigkeiten* auf der anderen Seite.

Ereignis-Erklärungen sind Antworten auf Fragen wie „Was war die Ursache für diesen Autounfall?“, „Warum wurde dieses Stück Stahl rotglühend?“ oder „Warum fühlte Hans einen stechenden Schmerz in seinem Fuß?“. Solche Antworten könnten z.B so aussehen: „Eine regennasse Fahrbahn im Zusammenhang mit der Tatsache, daß der Fahrer zu viel Whisky getrunken hatte“, „Weil es auf eine Temperatur von 800° C erhitzt wurde“ und „Weil er am Strand auf eine Muschel getreten war“. Bei Ereignis-Erklärungen geht es also darum, die *Ursachen* für das zu erklärende Ereignis *e* anzugeben: Bedingungen, die zusammen mit anderen Bedingungen für das Auftreten von *e* *hinreichend* und die unter den gegebenen Bedingungen für *e* auch *notwendig* waren.

Wenn es um Erklärungen dieser Art ginge, wären mentale Phänomene also genau dann neurobiologisch erklärt, wenn sich für jedes mentale Ereignis *e* (jeden Schmerz, jeden Gedanken, jeden Wunsch, usw.) hinreichende und unter den gegebenen Bedingungen auch notwendige neuronale Bedingungen *n* angeben ließen, also Bedingungen, für die gilt: Immer wenn die Bedingungen *n* realisiert sind, tritt ein Ereignis vom Typ *e* auf, und unter den gegebenen Bedingungen wäre *e* nicht aufgetreten, wenn *n* nicht stattgefunden hätte. Erklärungen dieser Art (Erklärungen, die ich im folgenden *schwache Erklärungen* nennen werde) würden – obwohl sie für sich genommen außerordentlich interessant wären – den Philosophen jedoch nicht befriedigen. Denn ihn interessiert in erster Linie die Frage, ob mentale Phänomene auf irgendeine Weise mit physischen Phänomenen *identifiziert* oder auf solche Phänomene *reduziert* werden können.¹ Und auf diese Frage geben kausale Ereignis-Erklärungen keine Antwort. Denn natürlich können mentale Phänomene auch dann von physischen Phänomenen kausal abhängig sein, wenn sie selbst weder mit physischen Phänomenen identifiziert noch auf solche Phänomene reduziert werden können.

¹ Zum hier einschlägigen Begriff von „Reduktion“ s. Beckermann (1992a, 1992b).

Meines Wissens war R. Cummins (1983, S. 1-27) der erste, der mit Nachdruck darauf hingewiesen hat, daß es noch einen weiteren Erklärungstyp gibt, der in der wissenschaftlichen Praxis eine mindestens ebenso große Rolle spielt wie der Typ der kausalen Ereignis-Erklärung. In vielen Wissenschaften geht es weniger um die Ursachen von Ereignissen als vielmehr um Antworten auf Fragen wie „Worauf beruht die Härte von Diamanten oder die Reinigungsfähigkeit von Seifen?“, „Worauf beruht die Fähigkeit von Computern, in Sekundenbruchteilen Millionen von Berechnungen durchzuführen?“, „Wie schaffen es Zugvögel, den Weg in ihre Winterquartiere zu finden?“ oder „Wie schafft es unser Organismus, eine relativ konstante Kerntemperatur von ca. 37° C aufrechtzuerhalten?“. Antworten auf solche Fragen nennt Cummins *Eigenschafts-Erklärungen*. Denn in ihnen geht es darum zu erklären, worauf die Eigenschaften eines Gegenstandes *beruhen*, was an einem Gegenstand dazu führt, daß er eine bestimmte Eigenschaft *F* hat.

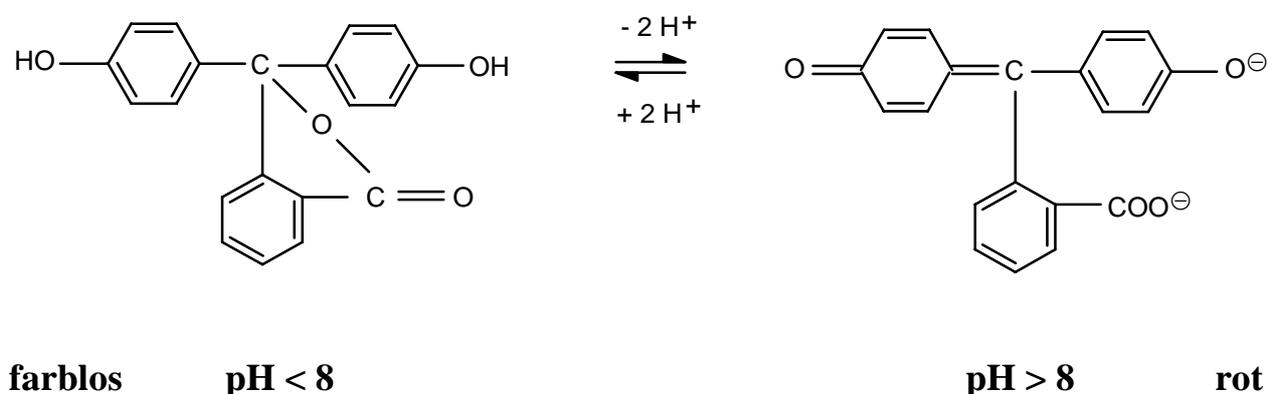
Many scientific theories are not designed to explain changes [or events – A.B.] but are rather designed to explain properties. The point of what I call a property theory is to explain the properties of a system not in the sense in which this means 'Why did S acquire P?' or 'What caused S to acquire P?' but, rather, 'What is it for S to instantiate P?', or, 'In virtue of what does S have P?'... Many of the most pressing and puzzling scientific questions are questions about properties, not about changes. We know a lot about what causes pain, but there is no very good theory of how pain is instantiated. Good property theories are wonderfully satisfying: we know how temperature is instantiated, how inheritance is instantiated, how electricity is instantiated, how solubility is instantiated. (1983, S. 14 f.)

Der Unterschied zwischen Ereignis- und Eigenschafts-Erklärungen beruht also darauf, daß man im Hinblick auf die Eigenschaften (oder Zustände) eines Gegenstandes zwei ganz verschiedene Fragen stellen kann. Man kann fragen, was dazu führte, daß der Gegenstand eine Eigenschaft bekam, die er vorher nicht hatte, d.h., was die *Ursache* dieser Veränderung war, und man kann fragen, worauf diese Eigenschaft *beruht*, d.h. aufgrund welcher Umstände er diese Eigenschaft hat.

Betrachten wir z.B. die Farbeigenschaften chemischer Verbindungen (vgl. dazu Glöckner 1973, 136ff.). Die Farbigekeit dieser Substanzen beruht darauf, daß sie einen Teil des Spektrums des sichtbaren Lichts absorbieren. (Werden alle Teile des Spektrums in gleicher Weise absorbiert, erscheint die Substanz grau oder schwarz.) Die Absorption entsteht dadurch, daß durch die Energie von auffallendem Licht Elektronen in einen höheren Energiezustand versetzt werden. Bei organischen Farbstoffen kommen in diesem Zusammenhang nur die π -Elektronen mesomerer Bindungssysteme in Frage. Intensiv gefärbte chemische Verbindungen sind dar-

über hinaus dadurch charakterisiert, daß sie zwei Arten von Gruppen enthalten: 1. ungesättigte Atomgruppen, die als Elektronenakzeptoren dienen können (diese werden Chromophore genannt); 2. Atomgruppen mit freien Elektronenpaaren, die als Elektronendonatoren dienen und die eine Farbverstärkung bewirken, ohne selbst eine farbgebende Wirkung zu besitzen (Auxochrome). Zwischen Mesomerie und Farbwirkung besteht folgende Beziehung: Organische Farbstoffe absorbieren Licht nur dann im sichtbaren Bereich, wenn sie mindestens drei mesomeriefähige konjugierte Doppelbindungen enthalten. Über die konjugierten Doppelbindungen hinweg kann die Auxochrom-Gruppe dem Chromophoren Elektronen liefern. Der Valenzausgleich tritt besonders leicht ein, wenn eine der Gruppen ionogen und eine neutral ist, das Farbstoffmolekül also als Ion vorliegt.

Nehmen wir als Beispiel das auch als Säure-Base-Indikator bekannte Phenolphthalein (Glöckner 1973, S. 141f.). Diese Substanz besitzt nicht immer die gleiche Farbe; vielmehr wird das bei einem pH-Wert < 8 farblose Phenolphthalein im alkalischen Bereich durch Öffnung des Laktoringes in ein mesomeriefähiges Anion von violettroter Farbe verwandelt (Abb.11.1).



11.1 Bei einem pH-Wert < 8 wird das ansonsten farblose Phenolphthalein durch Öffnung des Laktoringes in ein mesomeriefähiges Anion von violettroter Farbe verwandelt.

Ich habe dieses Beispiel gewählt, weil sich an ihm die beiden genannten Fragetypen sehr gut veranschaulichen lassen. Zum einen kann man fragen: Was ist die Ursache für die Farbveränderung dieser zuvor farblosen Probe Phenolphthalein? Zum anderen kann man aber auch fragen: Woran liegt es (was ist verantwortlich dafür), daß die Lösung violettrot ist? Als Antwort auf die erste Frage muß das Ereignis angegeben werden, das die Farbveränderung bewirkt hat, die Ursache dieser Veränderung – also: das Hinzugeben der Base. Was aber kann man als Antwort auf die zweite Frage anführen?

Das Beispiel macht klar, daß die angemessene Antwort hier darin besteht, daß man versucht, die zu erklärende Eigenschaft auf die *Mikrostruktur* des betreffenden Stoffs zurückzuführen. Hier geschieht das dadurch, daß man erstens die Struktur angibt, die Phenolphthaleinmoleküle bei einem pH-Wert > 8 besitzen, und daß man zweitens zeigt, daß Moleküle mit dieser Struktur bestimmte Bereiche des sichtbaren Lichts absorbieren, da in ihnen bestimmte Elektronen durch entsprechende Lichtquanten in einen höheren Energiezustand versetzt werden.

Eigenschafts-Erklärungen haben also im allgemeinen eine *mereologische* Struktur. Das heißt, in ihnen nimmt man auf die *Teile* eines Systems S und auf deren *Anordnung* Bezug und zeigt, daß jedes System, das aus den gleichen Teilen besteht, die in der gleichen Weise angeordnet sind, daß also jedes System mit derselben Mikrostruktur wie S alle die Merkmale aufweist, durch die die zu erklärende Eigenschaft F charakterisiert ist. Cummins zufolge kann man Eigenschafts-Erklärungen in Form eines deduktiven Schlusses darstellen:

- (EIG₁) (i) Für alle x gilt: wenn x aus den Teilen C_1, \dots, C_n besteht, die in der Weise O angeordnet sind, d.h. wenn x die Mikrostruktur $[C_1, \dots, C_n; O]$ besitzt, dann hat x die Eigenschaft F .
- (ii) S hat die Mikrostruktur $[C_1, \dots, C_n; O]$.
- (iii) S hat die Eigenschaft F .

Dieses Schema ist allerdings zu ungenau; denn es läßt den Status des Gesetzes (i) völlig offen. Es macht aber einen großen Unterschied, ob dieses Gesetz selbst nicht weiter erklärbar ist, so daß seine Geltung nur durch die Beobachtung entsprechender empirischer Korrelationen festgestellt werden kann, oder ob es möglich ist, aus den allgemein geltenden Naturgesetzen abzuleiten, daß Systeme mit der Mikrostruktur $[C_1, \dots, C_n; O]$ alle für die Eigenschaft F charakteristischen Merkmale besitzen. Denn dieser Unterschied markiert genau den Unterschied zwischen emergenten und reduzierbaren Eigenschaften (Beckermann 1992b, 1996). Nur im zweiten Fall kann man daher wirklich von einer Erklärung der Eigenschaft F sprechen oder sagen, daß das System S diese Eigenschaft *aufgrund* der Mikrostruktur $[C_1, \dots, C_n; O]$ hat bzw. daß die Eigenschaft F im System S durch diese Mikrostruktur *realisiert* ist. Wirkliche Eigenschafts-Erklärungen werden daher durch die im folgenden Schema (EIG₂) zusammengefaßten beiden Bedingungen besser charakterisiert als durch das Cumminsche Schema (EIG₁):

- (EIG₂) (i) Es kann aus den allgemein geltenden Naturgesetzen abgeleitet werden, daß Systeme mit der Mikrostruktur $[C_1, \dots, C_n; O]$ alle für die Eigenschaft F charakteristischen Merkmale besitzen.
- (ii) S hat die Mikrostruktur $[C_1, \dots, C_n; O]$.

Im Hinblick auf die Ausgangsfrage dieses Abschnitts ergibt sich damit die folgende Antwort:

- (NE) Ein mentales Phänomen (eine mentale Eigenschaft, ein mentaler Zustand) m in einem Organismus X ist genau dann (im *starken* Sinn) neurobiologisch erklärt, wenn
- (a) gezeigt werden kann, daß Organismen, die über ein Zentralnervensystem verfügen, das in der Weise O organisiert ist, alle für das mentale Phänomen m charakteristischen Merkmale besitzen, und
 - (b) X über ein Zentralnervensystem verfügt, das in der Weise O organisiert ist.

11.2 Was spricht gegen die Möglichkeit, mentale Phänomene neurobiologisch zu erklären?

Die beiden wichtigsten Merkmale, die in den Augen der Skeptiker grundsätzlich gegen die Möglichkeit sprechen, mentale Phänomene neurobiologisch bzw. allgemein: naturalistisch zu erklären, sind der *intentionale* und der *qualitative* Charakter dieser Phänomene. Um zu verstehen, was damit gemeint ist, ist es sinnvoll, sich zunächst einen groben Überblick über die verschiedenen Arten mentaler Phänomene zu verschaffen. Dabei ist es inzwischen allgemein üblich, zwei große Gruppen zu unterscheiden: intentionale Zustände (propositional attitudes) und Empfindungen (sensations).²

² Vergleiche zur folgenden Charakterisierung dieser Unterscheidung besonders McGinn 1982, S. 7ff. Der Terminus „Empfindung“ ist wegen seiner delikaten Geschichte in der deutschsprachigen Wahrnehmungsforschung an dieser Stelle möglicherweise etwas unglücklich; in der philosophischen Fachliteratur ist er allerdings schwer verzichtbar, da er sich inzwischen als Übersetzung des entsprechenden englischen Ausdrucks „sensation“ vollständig durchgesetzt hat. Um Mißverständnissen vorzubeugen, genügt vielleicht der Hinweis, daß „Empfindung“ in der philosophischen Literatur inzwischen ein Fachterminus

Intentionale Zustände (wie Glauben, Wünschen, Hoffen, Fürchten, etc.) sind dadurch gekennzeichnet, daß sie *auf etwas gerichtet* sind: „In der Vorstellung ist etwas vorgestellt, in dem Urteile etwas anerkannt oder verworfen, in der Liebe geliebt, in dem Hasse gehaßt, in dem Begehren begehrt usw.“ (Brentano 1924, S. 124)

Diese „intentionale (...) Inexistenz eines Gegenstandes“, diese „Beziehung auf einen Inhalt, die[se] Richtung auf ein Objekt (...), oder die[se] immanente Gegenständlichkeit“ wird seit Brentano als *Intentionalität* bezeichnet. Bei Brentano sieht es allerdings so aus, als sei ein Phänomen intentional, wenn es einen Gegenstand in einer bestimmten nicht räumlichen Weise als Objekt in sich enthält. Heute ist man sich dagegen weitgehend einig darüber, daß die Intentionalität von Wünschen, Überzeugungen usw. nicht darin besteht, daß sie eine schwer begreifliche Relation zu (möglicherweise nicht einmal existierenden) Gegenständen beinhalten, sondern darin, daß sie Wahrheits- bzw. Erfüllungsbedingungen haben. Dies drückt sich typischerweise darin aus, daß wir bei der Zuschreibung intentionaler Zustände „daß“-Sätze verwenden – z.B. wenn wir Hans eine bestimmte Überzeugung zuschreiben, indem wir sagen „Hans glaubt, daß es morgen regnen wird“. Grundsätzlich können bei jedem intentionalen Zustand somit zwei Aspekte unterschieden werden: die Art des Zustandes (Glauben, Wünschen, Hoffen, etc.) und sein – in „daß“-Sätzen formulierter – intentionaler Gehalt, d.h. das, was geglaubt, gewünscht oder gehofft wird.

Zu den *Empfindungen* sollen körperliche Empfindungen wie Schmerzen, Kitzel oder Übelkeit ebenso gehören wie Wahrnehmungseindrücke – etwa der Eindruck einer bestimmten Farbe, der Klang einer lauten Trompete oder der Geschmack einer reifen Birne. Zwischen diesen beiden Gruppen von Empfindungen gibt es zwar eine Reihe von Unterschieden; trotzdem ist es sinnvoll, sie zusammenzufassen. Denn alle Empfindungen sind auf den ersten Blick im wesentlichen durch ihre *phänomenalen* Eigenschaften definiert. Für Empfindungen ist ihr *qualitativer Gehalt* charakteristisch, d.h. *das, was man empfindet oder fühlt, wenn man eine solche Empfindung hat; die Art, wie es ist, eine solche Empfindung zu haben* (Nagel, 1974).

für phänomenale Zustände ist, also für alle mentalen Zustände, für die ihr qualitativer Gehalt (s.unten) entscheidend ist.

Offenbar sind intentionale Zustände, anders als Empfindungen, nicht so sehr durch spezifische phänomenale Qualitäten charakterisiert. Bei diesen Zuständen ist vielmehr das Verhalten von Personen von entscheidender Bedeutung. Man kann nicht die Absicht haben, ein Auto zu kaufen, ohne zumindest die Tendenz zu haben, das zu tun, was zum Erwerb eines Autos notwendig ist. Und wenn man feststellen will, ob eine Person die Überzeugung hat, daß es draußen kalt ist, fährt man in der Regel am besten, wenn man darauf achtet, ob sie sich warm anzieht, wenn sie nach draußen geht. Auch Empfindungen sind oft mit einem charakteristischen Verhalten verbunden. Wer Schmerzen hat, krümmt und windet sich in typischer Weise; er stöhnt, weint oder schreit; und er versucht, seine Schmerzen zu bekämpfen, indem er sich auf die Lippen beißt oder ins Bad geht und sich ein Aspirin holt. Trotzdem scheint dieses Verhalten nicht das Entscheidende. Ein perfekter Schauspieler kann dieses Verhalten so täuschend echt nachmachen, daß wir glauben mögen, er habe tatsächlich Schmerzen. Dennoch fehlt ihm das Entscheidende: Er *fühlt* den Schmerz nicht; dem Zustand, in dem er sich befindet, fehlt der typische qualitative Charakter. Und genau darin, d.h. in der Tatsache, daß sich Empfindungen nicht in Verhalten erschöpfen, scheint der Hauptgrund für die Schwierigkeit zu liegen, diese mentalen Phänomene naturalistisch zu erklären.

Diese Schwierigkeit ist von J. Levine (1983, S. 354-358) exemplarisch so formuliert worden. Wenn man die beiden Aussagen

- (1) Schmerz ist das Feuern von C-Fasern, und
- (2) Temperatur ist die mittlere kinetische Energie der Moleküle eines Gases

miteinander vergleicht, dann zeigt sich bei näherer Untersuchung ein entscheidender Unterschied. Auf der einen Seite ist es nämlich in einem bestimmten Sinn *undenkbar*, daß in einem Gas die mittlere kinetische Energie der Moleküle 6.21×10^{-21} Joule beträgt, das Gas aber nicht die Temperatur von 300 K hat, während es auf der anderen Seite sehr wohl denkbar zu sein scheint, daß in meinem Körper die C-Fasern feuern, ich aber keinen Schmerz empfinde. Nach Levine liegt dies daran, daß die Aussage (2) *vollständig explanatorisch* ist, die Aussage (1) dagegen nicht. Was ist damit gemeint?

Wenn man uns fragen würde, was wir mit dem Ausdruck ‘Temperatur’³ meinen, dann würden wir wahrscheinlich antworten:

- (2') Temperatur ist die Eigenschaft von Körpern, die in uns bestimmte Wärme- bzw. Kälteempfindungen hervorruft, die dazu führt, daß die Quecksilbersäule in Thermometern, die mit diesen Körpern in Berührung kommen, steigt oder fällt, die bestimmte chemische Reaktionen auslöst, und so weiter.

Mit anderen Worten: Wir würden Temperatur durch ihre *kausale Rolle* charakterisieren. Der Grund für den explanatorischen Charakter von (2) ist daher

... that our knowledge of chemistry and physics makes intelligible how it is that something like the motion of molecules could play the causal role we associate with heat. Furthermore, antecedent to our discovery of the essential nature of heat, its causal role, captured in statements like (2'), *exhausts our notion of it*. Once we understand how this causal role is carried out there is nothing more we need to understand. (Levine, 1983, S. 357; Hervorhebung A.B.)

Mit der Aussage (1) ist es auf den ersten Blick ähnlich. Denn auch mit dem Ausdruck „Schmerzen“ assoziieren wir eine bestimmte kausale Rolle. Schmerzen werden durch die Verletzung von Gewebe verursacht, sie führen dazu, daß wir schreien oder „Aua!“ sagen, sie bewirken, daß wir in Zukunft die schmerzauslösende Situation vermeiden, usw. Auch die Aussage (1) erklärt also das Phänomen Schmerz, indem sie uns sagt, welcher neuronale Zustand Träger der entsprechenden kausalen Rolle ist. Der entscheidende Unterschied ist jedoch, daß sich die Bedeutung von „Schmerz“ nicht in einer kausalen Rolle erschöpft.

However, there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does!* For there seems to be nothing about C-fiber firing which makes it naturally „fit“ the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties. Unlike its functional role, the identification of the qualitative side of pain with C-fiber firing (...) leaves the connection between it and what we identify it with completely mysterious. One might say, it makes the way pain feels into merely a brute fact. (ibd.)

³ Für das richtige Verständnis dieser Argumentation ist es wichtig zu sehen, daß mit „Temperatur“ hier eine objektive Eigenschaft von Gasen, Flüssigkeiten und festen Stoffen gemeint ist und nicht die Sinnesempfindungen der Wärme und Kälte, die durch diese objektive Eigenschaft in uns hervorgerufen werden.

Der wesentliche Grund dafür, daß die Aussage (1) nicht vollständig explanatorisch ist, ist also, daß a) die Bedeutung des Ausdrucks „Schmerz“ außer einer kausalen Rolle auch einen qualitativen Aspekt umfaßt und daß man b) allein mit den Mitteln der Physik, Chemie und Neurobiologie *unmöglich zeigen kann*, daß es sich für einen Organismus in der für Schmerzen charakteristischen Weise *anfühlt*, wenn seine C-Fasern feuern. Ja, es scheint außerhalb der Reichweite jeder möglichen Naturwissenschaft zu zeigen, daß sich das Feuern von C-Fasern auf irgendeine Weise anfühlt (vgl. dazu auch Broad 1925; Stephan 1993). Vor dem Hintergrund der im letzten Abschnitt vorgetragenen Überlegungen zum Begriff der naturalistischen Erklärung kann man dieses Ergebnis verallgemeinernd so formulieren: Es scheint unmöglich, Empfindungen naturalistisch zu erklären, weil Empfindungen mit ihrem qualitativen Charakter ein Merkmal haben, das physische Zustände nicht besitzen können bzw. von dem mit naturwissenschaftlichen Mitteln niemals gezeigt werden kann, daß physische Zustände dieses Merkmal besitzen (vgl. Bieri 1992).

Wenn das so ist, dann scheint die Situation für intentionale Zustände jedoch günstiger. Denn diese mentalen Zustände sind, wie gesagt, weniger durch bestimmte phänomenale Qualitäten, sondern eher durch die Art charakterisiert, in der sie unser Verhalten beeinflussen. Intentionale Zustände sind daher vielleicht besser naturalistisch erklärbar als Empfindungen. Allerdings: Auch intentionale Zustände besitzen einen Aspekt, der in diesem Zusammenhang zu Problemen führt. Denn bei ihrer Zuschreibung spielen Prinzipien der Rationalität eine Rolle, für die ebenfalls im Bereich des Physischen kein Platz zu sein scheint. Wenn wir einer Person bestimmte Wünsche und Überzeugungen zuschreiben, dann geht es uns auch darum, das Verhalten dieser Person als *sinnvoll* zu verstehen, indem wir zeigen, daß es für diese Person (bei den Wünschen und Überzeugungen, die sie hat) *rational* ist, so zu handeln, wie sie es tut. Aus diesem Grunde können wir Personen nur intentionale Zustände zuschreiben, die auch im Verhältnis zueinander rational sind, d.h. die insgesamt ein kohärentes und konsistentes Muster von Einstellungen bilden. Nur wenn wir diese Standards der Rationalität anlegen, kann uns der Nachweis gelingen, daß eine Einstellung oder ein Verhalten sinnvoll ist. Rationalitätserwägungen dieser Art scheinen aber nur im Bereich des Mentalen möglich. Bei der Charakterisierung der Zustände unseres ZNS oder unseres Körpers spielen solche Erwägungen normalerweise keine Rolle. Wie bei den Empfindungen ihr phänomenaler Charakter scheint es also bei den intentionalen Zuständen das mit ihnen untrennbar verbundene Element von Rationalität unmöglich zu machen, diese Zustände naturalistisch zu erklären.

11.3 Kann die Neurobiologie dennoch helfen, mentale Phänomene naturalistisch zu erklären?

Angesichts der Batterie der im letzten Abschnitt ins Feld geführten Argumente mag es völlig aussichtslos erscheinen, weiter nach einer starken neurobiologischen Erklärung mentaler Phänomene zu suchen. Aber dies sollte einen erstens natürlich nicht davon abhalten, weiter nach schwachen Erklärungen zu suchen. (Und solche Erklärungen zu finden, wäre ebenfalls ein beträchtlicher Erfolg.) Und zweitens: Man sollte die Flinte niemals zu früh ins Korn werfen. Manchmal erschließt sich die *Möglichkeit* einer naturalistischen Erklärung erst, wenn man das Funktionieren eines komplexen Systems richtig verstanden hat. In der Geschichte der Wissenschaften gab es dafür immer wieder beeindruckende Beispiele. Seit der Antike galt es etwa als unmöglich, die charakteristischen Merkmale des *Lebens* (Selbsterhaltung, Wachstum, Reproduktion, zielgerichtetes Verhalten usw.) naturalistisch zu erklären. Zur Erklärung dieser Merkmale wurden deshalb immer wieder nicht-naturalistische Entitäten postuliert – die Seele, aristotelische Formen oder der *élan vital*. Erst zu Beginn der Neuzeit wurde unter anderem von Descartes versuchsweise eine naturalistische Gegenposition vertreten. Doch erst die Ergebnisse der modernen Molekularbiologie verhalfen uns zu einem – zumindest weitgehend – adäquaten naturalistischen Verständnis der Phänomene des Lebens. Zu ihrer Zeit außerordentlich plausible a priori-Argumente gegen die Möglichkeit, Leben zu erklären, wurden also durch den Fortschritt der Wissenschaften widerlegt.

Warum sollte es also – trotz aller gegenteiligen Argumente – unmöglich sein, daß ein umfassendes Verständnis der Struktur und des Funktionierens unseres ZNS eines Tages auch zu einer adäquaten naturalistischen Erklärung unseres mentalen Lebens führt? Schließlich kann es gar keinen Zweifel geben, daß auch in diesem Bereich der Naturwissenschaften in den letzten Jahren beeindruckende Fortschritte erzielt wurden. Und a priori-Argumenten gegen die Möglichkeit einer solchen Erklärung sollte man, wie schon gesagt, angesichts der Tatsache, daß solche Argumente in anderen Wissenschaftsbereichen durch die Entwicklung der Forschung einfach überholt wurden, keine allzu große Bedeutung beimessen.

Wenn man allerdings die konkreten Ergebnisse der empirischen Forschung betrachtet, scheint eine neurobiologische Erklärung mentaler Phänomene im Augenblick doch noch in weiter Ferne zu liegen. Traditionell beruhte das Wissen um den Zusammenhang zwischen Gehirnfunktionen und mentalen Zuständen fast ausschließlich auf Untersuchungen darüber, wie bestimmte geistige Phänomene durch

die Schädigung oder Zerstörung bestimmter Teile des Gehirns oder ZNS ganz oder teilweise beeinträchtigt werden. Grundlage solcher Untersuchungen bildeten beim Menschen z.B. Gehirnschäden, die durch Unfälle und Kriegsverwundungen oder durch Schlaganfälle und Tumore verursacht wurden. Typisch ist in diesem Zusammenhang die Untersuchung von Paul Broca im Jahre 1865, der bei der Obduktion eines vor seinem Tode an einer Aphasie leidenden Patienten entdeckte, daß ein bestimmter Teil des Cortex dieses Patienten – die später nach ihrem Entdecker benannte Brocasche Area – zerstört war. Bei Tieren wurden manchmal auch bestimmte Teile des Gehirns chirurgisch entfernt, um die entstehenden Folgen genau beobachten zu können (Flourens, Munk).

Was kann mit Untersuchungen dieser Art gezeigt werden? Auf den ersten Blick nur sehr wenig. Denn selbst wenn es gelingt, eindeutig nachzuweisen, daß die Schädigung oder der Ausfall bestimmter Gehirnregionen unweigerlich die Beeinträchtigung oder den Ausfall bestimmter mentaler Leistungen nach sich zieht, zeigt dies nur, daß in diesen Regionen Gehirnprozesse ablaufen, die für die entsprechenden mentalen Leistungen *notwendig* sind. Es sagt nichts aus über die *Art* dieser Prozesse; es kann daher auf gar keinen Fall zeigen, daß diese Prozesse diese mentalen Leistungen in dem oben erläuterten Sinn erklären; ja es zeigt nicht einmal, daß diese Prozesse hinreichende Bedingungen für diese Leistungen darstellen.

Kann man mit Hilfe neuerer Verfahren mehr zeigen? Ja und nein. Bildgebende Verfahren wie das Lassensche Verfahren zur Darstellung des corticalen Blutflusses oder das PET-Scanning führen auch nicht zu viel aussagekräftigeren Ergebnissen. Alle diese Verfahren sind im wesentlichen dazu geeignet festzustellen, in welchen Regionen des Gehirns bei bestimmten mentalen Phänomenen eine höhere Aktivität vorliegt. So läßt sich mit den genannten Verfahren z.B. zeigen, daß eine erhöhte Aktivität im Occipitallappen des Cortex feststellbar ist, wenn eine Person ein bewegtes Objekt mit den Augen verfolgt; daß beim Zuhören die Aktivität über dem Temporallappen zunimmt; daß Bewegungen mit einer Erhöhung der Aktivität im motorischen Cortex korreliert sind, usw. Alle diese Ergebnisse gehen zwar einerseits deutlich über das hinaus, was mit Hilfe der Erforschung von Läsionsfolgen festgestellt werden kann. Denn sie zeigen im günstigsten Fall tatsächlich *alle* Regionen des Cortex, die bei einer bestimmten Leistung benötigt werden. Andererseits sagen sie aber genauso wenig über die spezifischen neuronalen Prozesse aus, auf denen diese Leistung beruht. Denn auch sie zeigen nur, *wo* (und im günstigsten Fall auch *wann*) etwas stattfindet, aber nicht, *was* stattfindet. Insofern bringen auch sie uns einer wirklichen Erklärung der entsprechenden Phänomene noch nicht viel

näher. Vielmehr bringen sie uns in eine Situation, die vielleicht mit der vergleichbar ist, daß wir die Funktionsweise eines Computers erklären sollen *allein* aufgrund der Information, *wann* und *in welchen Teilen* des Computers bei bestimmten Aufgaben die elektrische Aktivität besonders hoch ist.

In gewisser Weise gilt das auch für mit dem EEG ermittelte ereigniskorrelierte Potentiale. Denn aus diesen kann man zwar erschließen, daß mit einem bestimmten Ereignis tatsächlich ein spezifischer neuronaler Prozeß korreliert ist. In gewissem Maße läßt sich sogar die räumliche und zeitliche Struktur dieses Prozesses erschließen. Aber der Prozeß selbst bleibt wieder unsichtbar. Wir wissen, daß etwas vorgeht, wo es sich abspielt, wann es sich abspielt, aber wir wissen immer noch nicht, was sich im einzelnen abspielt. Grundsätzlich scheint also das Problem zu bestehen, daß es aufgrund der überaus komplexen Struktur des Gehirns außerordentlich schwierig ist, die neuronalen Prozesse und Mechanismen, die bestimmten mentalen Phänomenen zugrunde liegen, so genau zu beschreiben, daß es möglich wird, diese Phänomene durch diese Prozesse im starken Sinne zu erklären.

Möglicherweise führen hier stärker ins Detail gehende Methoden – etwa Einzel- und Multizellableitungen – wenigstens ein Stück weiter. Immerhin konnten Hubel und Wiesel ihre bahnbrechenden Ergebnisse mit Hilfe dieser Methoden erreichen – Ergebnisse, die meiner Meinung nach einen wichtigen Schritt in die erforderliche Richtung darstellen. Warum? In erster Linie deshalb, weil sie zeigen, daß bei der Verarbeitung der von der Retina bzw. dem Corpus geniculatum laterale kommenden Signale in der primären Sehrinde Orientierungssäulen eine entscheidende Rolle spielen, deren Zellen im wesentlichen dann reagieren, wenn die Reize auf der entsprechenden Region der Retina von Kanten einer bestimmten Orientierung hervorgerufen werden. Dies ist deshalb von großer Bedeutung, weil man solchen Kantendetektoren eine klare funktionale Bedeutung bei der Analyse des auf der Retina ankommenden Bildes zuschreiben kann. Wenn es darum geht, aus einem (zweidimensionalen) Bild die (dreidimensionale) Szene zu rekonstruieren, die zu diesem Bild geführt hat, ist es nämlich sinnvoll, in einem ersten Schritt die zentralen Bildelemente herauszuarbeiten. Und dabei kommt gerade der Entdeckung von Kanten eine entscheidende Bedeutung zu. Allerdings ist auch dies nicht mehr als ein erster Schritt. Wie die in der primären Sehrinde enthaltenen Informationen so weiter verarbeitet werden, daß etwa Repräsentationen der wahrgenommenen Umweltszene entstehen, ist, soweit ich weiß, ziemlich unklar. Außerdem enthalten die Ergebnisse von Hubel und Wiesel auch nicht den kleinsten Hinweis darauf, wie der qualitative Aspekt visueller Eindrücke neurobiologisch erklärt werden könnte.

Dies gilt sicher auch für einen zweiten – für sich genommen äußerst interessanten – Ansatz: die Untersuchungen zur zeitlichen Synchronisation corticaler Zellen, wie sie etwa von Eckhorn und Bauer in Marburg und von der Arbeitsgruppe von Wolf Singer am MPI für Hirnforschung in Frankfurt/M. durchgeführt werden. Auch hier besteht der interessante Punkt darin, daß dieser Synchronisation eine funktionale Bedeutung bei bestimmten mentalen Leistungen zugeschrieben werden kann. Christoph von der Malsburg z.B. hat die Auffassung vertreten, daß durch die Synchronisation verschiedener Feature-Detektoren möglicherweise die Merkmale zusammengebunden werden, die jeweils zu einem wahrgenommenen Objekt gehören (in diesem Sinne auch Engel, König & Singer 1993). Und Wolf Singer nimmt an, daß dieser Synchronisation auch eine wichtige Aufgabe bei der Trennung von Figur und Hintergrund zukommen könnte. Ebenso wie bei den von Hubel und Wiesel entdeckten Orientierungssäulen handelt es sich allerdings auch hier nur um einen kleinen Teilaspekt des offenbar sehr komplexen Geschehens, das der visuellen Wahrnehmung zugrunde liegt. Von einer adäquaten neurobiologischen Erklärung zumindest von Wahrnehmungsphänomenen sind wir also auch nach diesen Untersuchungen noch meilenweit entfernt. Und dies gilt umso mehr, als die kritischen Aspekte *Rationalität* und *phänomenale Qualitäten* noch gar nicht in das Blickfeld dieser Untersuchungen getreten sind.

Steht es also doch schlecht um die Aussichten auf eine adäquate neurobiologische Erklärung unseres mentalen Lebens? Ich denke, diese Frage läßt sich im Augenblick noch nicht beantworten. Klar ist nur, daß wir trotz aller Erfolge der Neurowissenschaften zum gegenwärtigen Zeitpunkt von einer solchen Erklärung noch weit entfernt sind. Und klar ist auch, daß wir nicht viel weiter kommen werden, wenn die entscheidenden Aspekte *Rationalität* und *phänomenale Qualitäten* nicht explizit thematisiert werden. Ich möchte deshalb noch kurz auf eine Theorie eingehen, die in den letzten Jahren von Hans Flohr entwickelte wurde, da diese Theorie mit dem expliziten Anspruch auftritt, gerade den qualitativen Charakter des Bewußtseins zu erklären (Flohr 1992, 1994; zum Problem der Erklärung des Merkmals der Rationalität s. Beckermann 1990b, 1991).

Flohr geht bei seinen Überlegungen von der Feststellung aus, daß wir zwar in vielen Fällen wissen, unter welchen Bedingungen sich Bewußtsein verändert oder ausfällt (z.B. durch Psychopharmaka, Narkose, Sauerstoffmangel, Verletzungen des Hirnstammes und cortikale Läsionen), daß es aber bisher nicht gelungen ist, diese Daten auf einen gemeinsamen Nenner zu bringen. Das heißt, wir wissen nicht, *warum* unter bestimmten physiologischen Bedingungen Bewußtsein auftritt

und unter anderen fehlt. Im Hinblick auf die Frage nach einer möglichen neurobiologischen Erklärung von Bewußtsein muß man daher zwei Teilprobleme unterscheiden: ein empirisches und ein theoretisches oder konzeptionelles Teilproblem. Bei dem empirischen Teilproblem geht es darum herauszufinden, unter welchen neurophysiologischen Bedingungen Bewußtsein auftritt, welche dieser Bedingungen (einzeln) notwendig und (zusammen) hinreichend für Bewußtsein sind. Bei dem theoretischen Teilproblem dagegen geht es um ein Verständnis dieses Zusammenhangs, d.h. um die Frage, *warum* das so ist. Flohr trifft hier also eine Unterscheidung, die genau der oben eingeführten Unterscheidung zwischen schwachen und starken neurobiologischen Erklärungen von Bewußtsein entspricht.

Empirisch konzentriert sich Flohr auf die Tatsache, daß auch sogenannte NMDA-Antagonisten wie Phencyclidin und Ketamin Bewußtseinsstörungen bzw. Bewußtlosigkeit verursachen. *Warum ist das so?* Flohrs Antwort auf diese entscheidende theoretische Frage läßt sich – stark verkürzend – so zusammenfassen. 1) Nicht blockierte NMDA-Kanäle verleihen den entsprechenden Synapsen eine Hebb-Charakteristik, d.h. die synaptischen Verbindungen werden bei gleichzeitiger prä- und postsynaptischer Aktivität verstärkt. Nicht blockierte NMDA-Kanäle ermöglichen damit die Bildung neuronaler Assemblies. 2) Das Gewicht von Synapsen mit nicht blockierten NMDA-Kanälen kann sehr schnell verändert werden. Im Effekt ergibt sich also: 3) Nicht blockierte NMDA-Kanäle ermöglichen die schnelle Bildung und Auflösung neuronaler Assemblies. Genau dieser Effekt ist für Flohr die entscheidende Bedingung für die Entstehung bewußter mentaler Zustände. Denn er geht, was die theoretische Analyse des Problems angeht, von drei zusätzlichen Annahmen aus: 4) Neuronale Assemblies sind Repräsentationen. 5) Einige dieser Repräsentationen können als Metarepräsentationen aufgefaßt werden. Und 6) Bewußtsein – auch phänomenales Bewußtsein – beruht im wesentlichen auf der Fähigkeit des Organismus, Metarepräsentationen zu bilden, d.h. Repräsentationen, die sich auf die eigenen mentalen Zustände beziehen.

Ohne Zweifel gibt es in dieser Argumentation eine ganze Reihe von kritischen Punkten, und es ist daher in meinen Augen mehr als zweifelhaft, daß es Flohr wirklich gelungen ist, den qualitativen Charakter von Empfindungszuständen neurobiologisch zu erklären (vgl. hierzu Beckermann 1994; die Grundzüge einer alternativen Theorie finden sich in Beckermann 1995). Aber das ist nicht entscheidend. Entscheidend ist vielmehr, daß in der Flohrschen Theorie empirische und theoretische Analysen konsequent miteinander verknüpft werden und daß in den theoretischen Teil auch einschlägige philosophische Ansätze angemessen integriert sind.

Es ist gerade die Pointe der Flohrschen Theorie, daß sie versucht, eine Brücke zu schlagen zwischen der empirischen Tatsache, daß NMDA-Antagonisten Bewußtseinsstörungen und Bewußtlosigkeit verursachen, und der philosophischen Annahme, daß Bewußtsein genau dann vorliegt, wenn ein Organismus über Metarepräsentationen verfügt.⁴

Eine Antwort auf die Frage, ob bzw. inwieweit mentale Phänomene neurobiologisch erklärt werden können, wird sich meiner Meinung nach nur finden lassen, wenn die empirische Forschung genau in dieser Weise noch ein ganzes Stück weiter vorangetrieben wird.

Literatur

Beckermann, A. *Zur Logik der Identitätstheorie*. In: Pasternack, G. (Hrsg.) *Philosophie und Wissenschaften*. Frankfurt/M. (Peter Lang) 1990a. S. 87-110.

Beckermann, A. *Monismus – Herausforderung an die Philosophie*. In: *Aus Forschung und Medizin* (Schering) 5/1 (1990b) S. 25-32.

Beckermann, A. *Der endgültige Todesstoß für den Repräsentationalismus? – Eine Replik auf Andreas Kemmerlings Artikel 'Mentale Repräsentationen'*. In: *Kognitionswissenschaft* 2 (1991) S. 91-98.

Beckermann, A. *Introduction – Reductive and Nonreductive Physicalism*. In: Beckermann, A.; Flohr, H.; Kim, J. (Hrsg.) *Emergence or Reduction?* Berlin, New York (de Gruyter) 1992a. S. 1-21.

Beckermann, A. *Supervenience, Emergence, and Reduction*. In: Beckermann, A.; Flohr, H.; Kim, J. (Hrsg.) *Emergence or Reduction?* Berlin, New York (de Gruyter) 1992b. S. 94-118.

⁴ Für das Verständnis der Flohrschen Argumentation ist es wichtig zu wissen, daß die z.B. von D. Rosenthal (vgl. etwa Rosenthal 1986) vertretene metarepräsentationalistische Analyse von Bewußtsein in der Tat eine der wichtigsten Bewußtseinstheorien darstellt, die zur Zeit in der Philosophie des Geistes diskutiert werden. Darin daß Flohr sich gerade auf diesen Ansatz stützt, liegt möglicherweise allerdings auch die größte Schwäche seiner Theorie; denn man kann mit guten Gründen die Auffassung vertreten, daß die metarepräsentationalistische Analyse von Bewußtsein ihrerseits den phänomenalen Charakter von Empfindungen nicht nur nicht erklärt, sondern völlig außer acht läßt.

- Beckermann, A. *Metarepräsentationen und phänomenale Zustände*. In: Lenk, H.; Poser, H. (Hrsg.) *Neue Realitäten – Herausforderung der Philosophie. XVI. Deutscher Kongreß für Philosophie. Vorträge und Kolloquien*. Berlin (Akademie Verlag) 1994. S. 236-245.
- Beckermann, A. *Visuelle Informationsverarbeitung und phänomenales Bewußtsein*. In: Metzinger, T. (Hrsg.) *Bewußtsein – Beiträge aus der Gegenwartsphilosophie*. Paderborn (Schöningh) 1995. S. 663-679.
- Beckermann, A. *Eigenschafts-Physikalismus*. In: *Zeitschrift für Philosophische Forschung* 50 (1996). S. 3-25.
- Beckermann, A.; Flohr, H.; Kim, J. (Hrsg.) *Emergence or Reduction?* Berlin, New York (de Gruyter) 1992.
- Bieri, P. *Was macht Bewußtsein zu einem Rätsel?* In: *Spektrum der Wissenschaft* (Okt. 1992) S. 48-56.
- Brentano, F. *Psychologie vom empirischen Standpunkt*. (Hrsg. Kraus, O.) Leipzig (Meiner) 1924.
- Broad, C.D. *The Mind and its Place in Nature*. London (Allen & Unwin) 1925.
- Cummins, R. *The Nature of Psychological Explanation*. Cambridge, MA (MIT) 1983.
- Engel, A.K.; König P.; Singer, W. *Bildung repräsentationaler Zustände im Gehirn*. In: *Spektrum der Wissenschaft* (Sept. 1993) S. 42-47.
- Flohr, H. *Qualia and Brain Processes*. In: Beckermann, A.; Flohr, H.; Kim, J. (Hrsg.) *Emergence or Reduction?* Berlin, New York (de Gruyter) 1992. S. 220-238.
- Flohr, H. *Die physiologischen Bedingungen des Bewußtseins*. In: Lenk, H.; Poser, H. (Hrsg.) *Neue Realitäten – Herausforderung der Philosophie. XVI. Deutscher Kongreß für Philosophie. Vorträge und Kolloquien*. Berlin (Akademie Verlag) 1994. S. 222-235.
- Glöckner, W. (Hrsg.) *Fischer Kolleg Chemie*. Frankfurt/M. (Fischer) 1973.
- Levine, J. *Materialism and Qualia: The Explanatory Gap*. In: *Pacific Philosophical Quarterly* 64 (1983) S. 354-361.
- McGinn, C. *The Character of Mind*. Oxford (Oxford University Press) 1982.
- Nagel, Th. *What is it like to be a bat?* In: *Philosophical Review* 83 (1974) S. 435-450.

Rosenthal, D. *Two Concepts of Consciousness*. In: *Philosophical Studies* 94 (1986) S. 329-359.

Stephan, A. C.D. *Broads a priori-Argument für die Emergenz phänomenaler Qualitäten*. In: Lenk, H.; Poser, H. (Hrsg.) *Neue Realitäten – Herausforderung der Philosophie. Sektionsbeiträge I zum XVI. Deutschen Kongreß für Philosophie*. Berlin (1993). S. 176-183.

Weiterführende Literatur

Beckermann, A. *Eigenschafts-Physikalismus*. In: *Zeitschrift für Philosophische Forschung* 50 (1996), S. 3-25.

Bieri, P. (Hrsg.) *Analytische Philosophie des Geistes*. 2. Aufl., Königstein/Ts. (Hain) 1993.

Block, N. (Hrsg.) *Readings in Philosophy of Psychology, 2 Vol.* London (Methuen) 1980.

Churchland, P.M. *Matter and Consciousness* (Neuauf.) Cambridge, MA (MIT) 1988.

Guttenplan, S. (Hrsg.) *A Companion to the Philosophy of Mind*. Oxford (Blackwell) 1994.

Kim, J. *Philosophy of Mind*. Boulder, COL (Westview Press) 1996.

Lycan, W.G. (Hrsg.) *Mind and Cognition*. Oxford (Blackwell) 1990.

McGinn, C. *The Character of Mind*. Oxford (Oxford University Press) 1982.