Phänomenale Transparenz und kognitive Selbstbezugnahme

1. Einleitung

In diesem Beitrag möchte ich eine repräsentationalistische Analyse des reflexiven Selbstbewusstseins anbieten, die mit dem heute weit verbreiteten naturalistischen Programm in der Philosophie des Geistes kompatibel ist. Um die Grundlinien einer solchen repräsentationalistischen Analyse zu skizzieren, werde ich zwei neue Begriffe einführen: „phänomenale Transparenz“ (Abschnitt 2) und „attentionale Verfügbarkeit für die Introspektion“ (Abschnitt 3). Im vierten Teil dieses Aufsatzes werde ich diese neuen begrifflichen Werkzeuge dann auf ein interessantes Problem für den Naturalismus anwenden, das Lynne Baker (1998) aufgeworfen hat. Dabei werde ich potenzielle Äquivokationen zwischen phänomenologischen und erkenntnistheoretischen Lesarten von „Transparenz“ aufzeigen und dafür argumentieren, dass bestimmte cartesianische Gedanken (zum Beispiel Gedanken der Form [Ich bin mir gewiss, dass ich* existiere]) immer auf eine spezifische Form *phänomenalen* mentalen Inhalts reduzierbar sind und dass daher ihr Auftreten *per se* noch nicht den Schluss ermöglicht, dass wir durch sie mit einer genuinen oder sogar infalliblen Art von Selbsterkenntnis konfrontiert werden. Abschnitt 5 gibt einen kurzen Überblick über pathologische Fälle, in denen psychiatrische Patienten – obwohl sie noch als rationale Subjekte gelten müssen – sich konsistent über ihre eigene personale Identität täuschen oder die

unerschütterliche Überzeugung zum Ausdruck bringen, dass sie selbst *nicht* existieren. Ein Hauptziel meines Beitrags ist zu zeigen, dass ein zentraler *phenomenological constraint*, eine häufig übersehene einschränkende Bedingung für repräsentationalistische Theorien der kognitiven Selbstbezugnahme, darin besteht, dass auch hier phänomenal opake Zustände – erstens – tatsächlich existieren und – zweitens - häufig auf interessante Weise (nämlich innerhalb ein und desselben repräsentationalen Trägers) mit phänomenal transparenten Zuständen interagieren. Ein generelleres Ziel besteht darin, die Aufmerksamkeit darauf zu lenken, dass die Begriffe „phänomenale Transparenz“ und „phänomenale Opazität“ interessante Instrumente bei der Analyse bewusster, selbstrepräsentationaler Inhalte sind und dass ihre Relevanz für das Verständnis *kognitiver* Subjektivität in der Vergangenheit übersehen wurde. Ich behaupte, dass sich reflexives Selbstbewusstsein mit ihrer Hilfe naturalisieren lässt.

2. Phänomenale Transparenz

2.1 Der Begriff

Als *locus classicus* für die Verwendung des Begriffs der phänomenalen Transparenz wird gewöhnlich G.E. Moores Arbeit „Die Widerlegung des Idealismus“ („The refutation of idealism“, 1903) angeführt:

[...] the fact that when we refer to introspection and try to discover what the sensation of blue is, it is very easy to suppose that we have before us only a single term. The term ‚blue‘ is easy enough to distinguish, but the other element which I have called ‚consciousness‘ – that which a sensation of blue has in common with a sensation of green – is extremely difficult to fix. [...] And in general, that which makes the sensation of blue a mental fact seems

to escape us; it seems, if I may use a metaphor, to be transparent – we look through it and see nothing but the blue; we may be convinced that there *is something*, but *what* it is no philosopher, I think, has yet clearly recognised. [...] (Moore 1903, S. 446)

Heute besagt eine weit gefasste Standarddefinition - der wohl die meisten Fachleute in der Philosophie des Geistes zumindest im Sinne einer Diskussionsgrundlage zustimmen würden - phänomenale Transparenz bestehe darin, dass der Introspektion nur die Gehalteigenschaften einer bewussten mentalen Repräsentation verfügbar sind, nicht aber ihre nicht-intentionalen oder „Vehikeleigenschaften“. Introspektiv können wir ihren Gehalt erfassen, nicht aber den *Träger* dieses Gehalts. Normalerweise gehen wir dabei davon aus, dass Transparenz in diesem Sinne eine Eigenschaft *aller* phänomenalen Zustände ist.

- **DEF 1:** Phänomenale Zustände sind transparent, weil dem Erlebnissubjekt nur ihre Gehalteigenschaften introspektiv zugänglich sind.

Diese Definition ist phänomenologisch unplausibel und deshalb begrifflich unbefriedigend. Sie verstößt gegen für den Begriff des Bewusstseins wichtige einschränkende Bedingungen: Introspektive Unverfügbarkeit des Gehaltträgers ist keine notwendige Bedingung für Phänomenalität, denn nicht-intentionale Eigenschaften mentaler Repräsentationen *sind* häufig der Introspektion zugänglich. Es ist in diesem Zusammenhang vielleicht interessant, daran zu erinnern, dass Moore selbst in seiner Originalarbeit von der selben philosophischen Intuition geleitet wurde:

[...] that the moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a

mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous. Yet it *can* be distinguished if we look attentively enough, and if we know that there is something to look for. (Moore 1903, S. 450)

In § 275 seiner *Philosophischen Untersuchungen* hat auch Wittgenstein auf den naiven Realismus hingewiesen, der mit transparentem, phänomenalen Erleben zwangsläufig einhergeht.¹ Interessanterweise sind in letzter Zeit viele Autoren wieder auf den Begriff der Transparenz zurückgekommen und verwenden ihn als brauchbares begriffliches Werkzeug. Robert van Gulick hat bereits 1988 eine erste funktionalistische Analyse der phänomenalen Transparenz im Kontext seiner Theorie des Selbstbewusstseins entwickelt, basierend auf der Geschwindigkeit, Zuverlässigkeit und globalen Interdependenz der repräsentationalen Träger, wobei der introspektive oder linguistische Zugriff auf frühere Verarbeitungsstufen im System ausgeschlossen wird.² Was Qualia betrifft, hat Sydney Shoemaker

¹ § 275. Schau auf das Blau des Himmels, und sag zu dir selbst „Wie blau der Himmel ist!“ – Wenn du es spontan tust – nicht mit philosophischen Absichten – so kommt es dir nicht in den Sinn, diesen Farbeindruck gehöre nur *dir*. Und du hast kein Bedenken, diesen Ausruf an einen Andern zu richten. Und wenn du bei den Worten auf etwas zeigst, so ist es der Himmel. Ich meine: Du hast nicht das Gefühl des In-dich-selber-Zeigens, das oft das ‚Benennen der Empfindung‘ begleitet, wenn man über die ‚private Sprache‘ nachdenkt. Du denkst auch nicht, du solltest eigentlich nicht mit der Hand, sondern nur mit der Aufmerksamkeit auf die Farbe zeigen. (Überlege, was es heißt, „mit der Aufmerksamkeit auf etwas zeigen“.) (Wittgenstein 1952).

² Bei van Gulick wird „subjektiv erlebte Transparenz“ häufig auch als „semantische Transparenz“ bezeichnet, gemeint ist aber in fast allen Fällen das phänomenologische Element. Er schreibt zum Beispiel: “How can the functionalist account for subjectively experienced transparency? Possessing information always involves a capacity for appropriate behavior. Being informed about the content of a representation is being able to relate it to other representations and items in the world appropriate to its content. As long as understanding or being informed is analyzed as a behavioral capacity, even if the relevant behavior is all internal, the functionalist can hope to fit it within his theory.

Thus the functionalist should resist any view of phenomenal transparency as a form of nonbehavioral self-luminous understanding. He can undercut the intuitive appeal of that view, by explaining the subjective experience of understanding in terms of smooth and seemingly automatic transitions. The internal component of understanding need involve nothing beyond an ability to

zeigt, dass wir keinen introspektiven Zugriff auf die nicht-intentionalen Eigenschaften des Erlebens haben, die den betreffenden Inhalt kodieren.³ Gilbert Harman hat die Transparenz des Erlebens als die Tatsache definiert, dass wir seiner intrinsischen, nicht-intentionalen Eigenschaften nicht gewahr sind⁴; und Michael Tye verwendet den Begriff jetzt an vielen verschiedenen Stellen in seinen Arbeiten⁵. Er stellt dabei die starke (und interessante) Behauptung auf, der phänomenale Charakter sei tatsächlich als Ganzer *identisch* mit intentionalen Inhalten.

Ich möchte jetzt meine eigene Arbeitsdefinition für phänomenale

interrelate many diverse representations with great speed. [...] *How* this is done is not something to which I have linguistic or introspective access, but there must be powerful processors to produce these seemingly instantaneous transitions." (van Gulick 1988: 178f).

³ [...] The only thing that seems to answer the description „attending introspectively to one’s visual experience“ is attending to how things appear to one visually; and offhand this seem to tell one what the representational content of one’s experience is without telling one anything about what the nonintentional features of one’s experience are that encode this content. One may be inclined to say that one is revelling in the qualitative or phenomenal character of one’s experience when one „drinks in“ the blue of a summer sky or the red of a ripe tomato. But neither the blue nor the red is an object of introspective awareness; these are experienced, perceptually rather than introspectively, as located outside one, in the sky or in the tomato, not as features of one’s experience. G.E. Moore once complained that the sensation of blue is „as if it were diaphanous“; if one tries to introspect it one sees right through it, and sees only the blue. In a similar vein one might say that qualia, if there are such, are diaphanous; if one tries to attend to them, all one finds is the representative content of the experience. (Shoemaker 1990; 1996: 100f)

⁴ [...] in the case of her visual experience of a tree, I want to say she is not aware of, as it were, the mental paint by virtue of which her experience is an experience of seeing a tree. She is aware only of the intentional or relational features of her experience, not of its intrinsic non-intentional features.[...] When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree, including relational features of the tree „from here.“ (Harman 1990; 1997: 667)

⁵ Generalizing, introspection of your perceptual experiences seems to reveal only aspects of what you experience, further aspects of the scenes, as represented. Why? The answer, I suggest, is that your perceptual experiences have no introspectible features over and above those implicated in their intentional contents. So the phenomenal character of such experiences – itself is something that is introspectively accessible, assuming the appropriate concepts are possessed and there is no cognitive malfunction – is identical with, or contained within, their intentional contents. (Tye 1995,

Transparenz vorstellen, die Definition, auf die ich mich im weiteren beziehen werde. Transparenz in diesem Sinne ist, erstens, eine Eigenschaft aktiver mentaler Repräsentationen, die die minimal hinreichenden Bedingungen für das Auftreten bewusster Erfahrung bereits erfüllen. Manche dieser Bedingungen könnten sogar *notwendige* Bedingungen sein. Zum Beispiel werden phänomenal transparente Repräsentationen immer innerhalb eines virtuellen Gegenwartsfensters aktiviert und in ein einheitliches globales Weltmodell integriert. Ich werde diese einschränkenden Bedingungen und ihre logischen Beziehungen untereinander hier jedoch nicht diskutieren, sondern einfach voraussetzen, dass für jedes bewusste System eine Menge solcher Bedingungen existiert (eine kurze Liste findet sich in Metzinger 2000, Abb. 20.1; eine ausführlichere Darstellung in Metzinger 2003, Kapitel 3). Außerdem werde ich „phänomenale Transparenz“ und „phänomenale Durchsichtigkeit“ sowie „phänomenale Opazität“ und „phänomenale Undurchsichtigkeit“ ab jetzt als Synonyme verwenden.

Die zweite definierende Eigenschaft besteht darin, dass phänomenale Transparenz funktional dadurch charakterisiert ist, dass *frühere Verarbeitungsstufen im Gehirn* der Introspektion *attentional nicht verfügbar* sind. Letztlich resultiert sie aus einer strukturell-architektonischen Eigenschaft der neuronalen Informationsverarbeitung im Gehirn. Der Ausdruck „attentionale Verfügbarkeit“ bezieht sich dabei auf den Sonderfall der introspektiven *Aufmerksamkeit* und wird hier mit Bezug auf die Begriffe Introspektion₁ und Introspektion₃ verwendet, die in

Abschnitt 3 genauer eingeführt werden.

Was ist Aufmerksamkeit? Kurz gesagt: Aufmerksamkeit ist ein Vorgang der nicht-begrifflichen Metarepräsentation und der subsymbolischen Ressourcenallokation innerhalb eines Repräsentationssystems mit phänomenalen Zuständen, welcher auf bestimmten Teilen des gegenwärtig aktiven inneren Realitätsmodells operiert. Indem wir unsere Aufmerksamkeit auf ein bereits sinnlich wahrgenommenes Objekt lenken, führen wir zunächst eine *Selektion* und eine *Verstärkung* durch: Wir wählen die phänomenale Repräsentation dieses Objekts aus und intensivieren damit automatisch bestimmte dieser Repräsentation zu Grunde liegende Informationsverarbeitungsvorgänge in unserem Gehirn. Gleichzeitig erhöhen wir den Grad unserer *Wachheit* und *orientieren* uns auf den fraglichen Gegenstand. Wir tun dies jedoch ohne einen *Begriff* für das Objekt zu bilden. Typischerweise wird das Objekt dadurch prägnanter, detailreicher, und als kohärente Gestalt aus einem Hintergrund herausgelöst – es entsteht eine *neue* und bessere Repräsentation. Eine weitverbreitete Annahme in den Neurowissenschaften ist, dass das perzeptuelle Objekt durch die Aufmerksamkeit genaugenommen überhaupt *konstituiert* wird, indem einzelne, bereits aktive, Eigenschaftsrepräsentationen zu einem grösseren Ganzen integriert werden. Aufmerksamkeit ist also die Repräsentation einer oder mehrerer bereits aktiver Repräsentationen mit nicht-begrifflichen mentalen Mitteln, eben subsymbolische Metarepräsentation. Objektwahrnehmung ist keineswegs ein vollständig passiv-synthetischer Vorgang, denn viele Zielreize können überhaupt nur *entdeckt*

werden, wenn ihre Elementareigenschaften bereits miteinander verschmolzen worden sind. Dies ist eine weitere Hauptfunktion der Aufmerksamkeit: Sie *identifiziert* eine bestimmte Merkmalskombination innerhalb eines mehrdeutigen Wahrnehmungskontextes. Visuell-räumliche Aufmerksamkeit zum Beispiel segmentiert eine phänomenale Szene und trägt dadurch zur Reduktion von Ambiguität bei.

Innerhalb der repräsentationalen Architektur des menschlichen Geistes ist Aufmerksamkeitslenkung zudem eine *supramodale* Fähigkeit. Zum Beispiel kann sie unabhängig von Augenbewegungen im Raum verschoben werden, sich also innerhalb des Gesichtsfeldes bewegen obwohl die Position der Augen selbst stabil bleibt. Eine klassische Metapher auf der phänomenologischen Beschreibungsebene ist die des attentionalen „Blickstrahls“. Martin Eimer hat in einem hilfreichen und für Philosophen empfehlenswerten Überblicksartikel darauf hingewiesen, dass in der experimentalpsychologischen Literatur die visuell-räumliche Aufmerksamkeit häufig als mentaler „Lichtkegel“ beschrieben wird, der zu beliebigen Stellen des visuellen Feldes bewegt werden kann (Eimer 1996, S. 298). Die Lichtkegel-Metapher ist jedoch nicht nur eine phänomenologisch-repräsentationalistische Analogie, sondern ermöglicht auch eine *funktionale* Analyse von attentionaler Metarepräsentation: „Objekte, die sich innerhalb des „Lichtkegels der Aufmerksamkeit“ befinden, werden besser, schneller oder tiefer verarbeitet als Objekte, die außerhalb des Lichtkegels lokalisiert sind.“ (a.a.O.). Auf der Ebene der Neurobiologie schließlich zeigt sich, dass etwa die räumliche Aufmerksamkeit

frühe Stufen der visuellen Informationsverarbeitung funktional penetrieren kann, zum Beispiel indem bestimmte Bahnen der Informationsverarbeitung selektiv gehemmt, andere dagegen gestärkt werden. Obwohl es gegenwärtig noch keine umfassende neurowissenschaftliche Theorie der Aufmerksamkeit gibt, ist bereits klar, dass sie ein Prozess von begrenzter Kapazität ist, der das Antwortverhalten einzelner Neurone direkt modulieren kann und weder als eine Eigenschaft des Gehirns als Ganzem noch als Eigenschaft einer einzelnen, anatomisch eng umgrenzten Region beschrieben werden kann: Das Aufmerksamkeitssystem besteht aus einer Vielzahl einzelner und miteinander wechselwirkender Netzwerke, die auf sehr unterschiedliche Weise spezialisiert sein können (vgl. Posner 1995 und Teil V in Gazzaniga 1995).

Aufmerksamkeit kann auch nach innen gerichtet werden, zum Beispiel auf den Vorgang, durch den eine mentale *Selbstrepräsentation* erzeugt und in unser bewusstes Realitätsmodell eingebettet wird. Von besonderem philosophischem Interesse ist in diesem Zusammenhang die folgende Tatsache: Je mehr frühere Verarbeitungsstufen und je mehr Aspekte des inneren Konstruktionsvorgangs, der zum endgültigen, expliziten und desambiguierten phänomenalen Inhalt führt, für die introspektive Aufmerksamkeit zur Verfügung stehen, desto mehr wird das System als Ganzes imstande sein, diese phänomenalen Zustände *als* innere, selbsterzeugte Konstrukte zu erkennen. Totale Transparenz bedeutet totale attentionale Unverfügbarkeit früherer Verarbeitungsstufen. Grade von phänomenaler Undurchsichtigkeit treten als Grade attentionaler Verfügbarkeit

auf.

- **DEF 2:** Für jeden phänomenalen Zustand ist der Grad der Transparenz umgekehrt proportional zum Grad der attentionalen Verfügbarkeit früherer Verarbeitungsstufen.

Diese Definition weicht von früheren Begriffen der phänomenalen Transparenz darin ab, dass sie uns erlaubt, drei wichtige das phänomenale Bewusstsein betreffende Tatsachen zu beschreiben, die Philosophen häufig übersehen haben. Erstens reicht die *kognitive* Verfügbarkeit der Tatsache, dass momentan aktive phänomenale Inhalte das Endprodukt innerer repräsentationaler Vorgänge sind, nicht aus, um die phänomenale Transparenz dieser Inhalte aufzuheben oder zu schwächen. Der naive Realismus des Erlebens bleibt immer erhalten: Man kann sich nicht durch kognitive Operationen aus seinem phänomenalen Wirklichkeitsmodell „hinausdenken“. Einfach einen mental repräsentierten *Begriff* des Buches, das man gerade in seiner Hand hält, zu haben, und seinen Bewusstseinsinhalt damit mental vielleicht als eine besondere Form repräsentationalen Gehalts zu erfassen, ändert den unhintergebar realistischen Charakter der phänomenalen Erfahrung selbst praktisch nicht - wenigstens nicht in einer Weise, die in dem hier interessierenden Zusammenhang relevant wäre. Es scheint jedoch einen relevanten Unterschied zwischen kognitiver und attentionaler Verarbeitung zu geben, zwischen begrifflicher und nicht-begrifflicher Metarepräsentation von phänomenalen Zuständen erster Ordnung. Wenn man den Konstruktionsvorgang „selbst“ verfolgen könnte (also gewissermaßen auf der

„Objektebene“, durch einen subsymbolischen und eher wahrnehmungsartigen Vorgang der Metarepräsentation, welcher keine Trennung zwischen repräsentationalem Inhalt und repräsentationalem Träger mehr impliziert), würde man eine Verschiebung in der subjektiven Erfahrung bemerken, indem man nämlich seinem aktuellen Wirklichkeitsmodell einen neuen *nicht-begrifflichen* Inhalt hinzufügt.

Zweitens weicht diese Definition von der klassischen Unterscheidung zwischen Vehikel und Inhalt ab. Die herkömmliche Weise, Transparenz zu definieren, bestünde darin, dass man sagt: Nur die Gehalteigenschaften, nicht aber die Vehikeleigenschaften des phänomenal Repräsentierten sind dem System introspektiv verfügbar. Die Unterscheidung zwischen Vehikel und Inhalt ist ein äußerst nützliches begriffliches Instrument, aber sie enthält leicht zu übersehende Reste des cartesianischen Dualismus – weil sie uns verleitet, das Vehikel und den Inhalt zu verdinglichen, indem wir sie als epistemisch und ontologisch distinkte, unabhängige Entitäten auffassen. Eine rationale Theorie des Geistes wird sich aber immer auch daran orientieren müssen, was unsere gegenwärtig besten *empirischen* Theorien der mentalen Repräsentation darüber sagen, wie man die Beziehung zwischen Träger und Inhalt begrifflich am besten beschreibt. Solche Theorien gibt es bereits seit längerer Zeit. Der Konnektionismus hat uns nun schon seit einer Reihe von Jahren ein mathematisch präzises, psychologisch und neurobiologisch teilweise realistisches Verständnis davon vermittelt, wie sich in neuronalen Netzen Gehalteigenschaften und Vehikeleigenschaften gegenseitig bedingen und

wie sie analytisch eben nicht voneinander zu trennen sind. In neuronalen Netzen sind mentale Inhalte direkt in der physisch-kausalen Mikrostruktur des Systems verankert und können nicht losgelöst von ihr betrachtet werden. Ähnlichkeiten im Inhalt drücken sich direkt in Ähnlichkeiten der kausalen Rolle aus (Helm 1991). Ein empirisch gesehen plausibleres Modell repräsentationalen Inhalts wird diesen außerdem als einen Aspekt eines fortlaufenden *Prozesses* und nicht als eine Art abstrakten Objekts beschreiben müssen. Was wir für eine rationale Theorie der phänomenalen Repräsentation brauchen, ist *verkörperter* geistiger Inhalt, ein sich fortlaufend entfaltender und physikalisch realisierter Vorgang des *Beinhaltens*, nicht „ein“ Inhalt (siehe z.B. Churchland 1998, <unveröffentlichtes Manuskript>, Clark 1997, O'Brien und Opie 2001).

Drittens hat die Beschreibung phänomenaler Transparenz auf der Basis der attentionalen Verfügbarkeit früherer *Stufen* der Verarbeitung den Vorteil, dass sie erlaubt, viele verschiedene abgestufte Begriffe von *Graden* der Transparenz und Undurchsichtigkeit zu entwickeln. Für unterschiedliche Klassen phänomenaler Zustände, die sich aus verschiedenen Arten der Verarbeitung ergeben, ist es vielleicht auch möglich, nicht nur veränderliche Grade, sondern unterschiedene *Arten* der Transparenz und Undurchsichtigkeit zu beschreiben. Dies ermöglicht dann eine erheblich realistischere Beschreibung bestimmter phänomenaler Eigenschaften, die unterschiedlichen Klassen von Bewusstseinszuständen zugeordnet sind.

1.2 Die Phänomenologie der Transparenz

Transparenz ist eine besondere Form von Dunkelheit. Bezogen auf die Phänomenologie der visuellen Wahrnehmung bedeutet Transparenz, dass wir etwas nicht sehen können, weil es durchsichtig ist. Wir sehen nicht das Fenster, sondern nur den vorbeifliegenden Vogel. Im Sehen ist jedoch die negative Tatsache, dass wir das Fenster nicht sehen, nicht *explizit* repräsentiert.

Phänomenale Transparenz im Allgemeinen bedeutet dagegen, dass etwas Bestimmtes der subjektiven Erfahrung nicht zugänglich ist, und zwar der fundamental *repräsentationale* Charakter der Inhalte des bewussten Erlebens. Diese Analyse bezieht sich auf alle Formen der Sinneswahrnehmung und insbesondere auf unser integriertes phänomenales Weltbild als Ganzes – aber eben auch auf große Teile unseres Selbstmodells (Metzinger ²1999). Die *Instrumente* der Repräsentation selbst können nicht mehr als solche repräsentiert werden, und so ist das System, das die Erfahrung macht, auf dieser Stufe und mit begrifflicher Notwendigkeit in einen naiven Realismus verstrickt. Dazu kommt es, weil sich das System selbst jetzt zwangsläufig als im direkten Kontakt mit den aktuellen Inhalten seines eigenen Bewusstseins erfahren muss. Was genau ist es, was das System nicht erleben kann? Was der bewussten Erfahrung unzugänglich ist, ist die schlichte Tatsache, dass diese Erfahrung immer in einem *Medium* stattfindet. Dadurch führt die Transparenz phänomenaler Gehalte zu einem weiteren Charakteristikum des bewussten Erlebens, nämlich zum subjektiven Eindruck von Unmittelbarkeit. Viele schlechte philosophische Argumente hinsichtlich direkter

Bekanntheit, unfehlbarem Erste-Person-Wissen und direkter Referenz beruhen auf Äquivokationen von epistemischer und phänomenologischer Unmittelbarkeit: Aus der Tatsache, dass die bewusste Erfahrung - z.B. der Farbe eines Gegenstands - die Merkmale phänomenaler Unmittelbarkeit und direkter Gegebenheit trägt, folgt nicht, dass durch diesen Vorgang irgendeine Art von unvermittelter oder direkter *Erkenntnis* konstituiert wird. Phänomenaler Gehalt als solcher ist weder intentionaler noch epistemisch gerechtfertigter Gehalt, und es ist eine weithin akzeptierte und plausible Annahme, dass er lokal superveniert, nämlich auf Eigenschaften des Gehirns. Zu jeder Art von phänomenalem Inhalt beim Menschen gibt es nach dieser Annahme wenigstens ein minimal hinreichendes neuronales Korrelat (vgl. Chalmers 2000). Phänomenaler Inhalt lässt sich von intentionalem Inhalt trennen: Ein isoliertes und funktional adäquat stimuliertes Gehirn im Versuchslabor könnte phänomenal transparente Zustände besitzen, die die subjektiven Farben von Gegenständen als unmittelbar und direkt gegeben repräsentieren. Transparenz ist *phänomenale* Transparenz, und nicht epistemische Gewissheit. Jede vollständig transparente phänomenale Repräsentation ist dadurch gekennzeichnet, dass die den Träger des Gehalts kausal erzeugenden physischen Mechanismen, die zu ihrer Aktivierung geführt haben, und die Tatsache, dass es jetzt gerade einen *inneren* Zustand gibt, der ihren Inhalt trägt, selbst dem introspektiven Zugriff entzogen sind. Die Phänomenologie der Transparenz ist daher die Phänomenologie des naiven Realismus.

Es existieren natürlich auch Beispiele für phänomenal *undurchsichtige*

Repräsentationen, und sie sind wichtig, um den Begriff der „Transparenz“ noch besser zu verstehen. *Opake* phänomenale Repräsentationen entstehen zum Beispiel immer dann, wenn die Information, *dass* es sich bei ihrem Inhalt um das Resultat eines inneren Darstellungsvorgangs handelt, plötzlich global verfügbar wird. Wenn Sie entdecken, dass es das Buch in ihrer Hand in Wirklichkeit gar nicht gibt, dann wird die Halluzination zur Pseudohalluzination: Auch auf der Ebene des Erlebens selbst ist jetzt die Information verfügbar, dass Sie nicht auf die Welt schauen, sondern auf einen aktiven repräsentationalen Zustand, der im Moment allem Anschein nach kein gutes Instrument zur Wissensgewinnung ist. Der phänomenale Buchzustand wird undurchsichtig. Was Sie verlieren, ist *sensorische* Transparenz. Ihnen wird die Tatsache bewusst, dass Wahrnehmungen durch Sinnesorgane erzeugt werden und dass diese Organe nicht in allen Situationen absolut zuverlässig funktionieren. Weitere Beispiele für phänomenal undurchsichtige Zustandsklassen: Am deutlichsten werden bewusste Gedanken, aber auch einige Arten von Emotionen, Pseudohalluzinationen oder Tagträume subjektiv *als repräsentationale Vorgänge* erlebt. Solche Vorgänge erscheinen uns manchmal als absichtlich herbeigeführte kognitive oder repräsentationale Vorgänge, manchmal als spontan eintretende, begrenzte oder sogar (im Klartraum, bei manchen religiösen Erlebnissen und bestimmten Klassen von psychiatrischen Störungsbildern) das ganze Realitätsmodell umfassende phänomenale Simulationen, deren Inhalt sich häufig der Kontrolle durch das Subjekt der Erfahrung entziehen. In solchen Fällen wissen wir, dass wir keine

Wirklichkeit erleben, sondern bestenfalls eine Möglichkeit: Die Tatsache, dass es nur *repräsentationale* Vorgänge sind, deren Inhalt eine äußere Realität angemessen darstellen kann oder auch nicht, ist für die Aufmerksamkeitslenkung, die mentale Begriffsbildung und die Verhaltenskontrolle global verfügbar (zum Begriff der „globalen Verfügbarkeit“ vgl. Baars 1988, Chalmers 1997, Metzinger 2000). Viele Autoren beschreiben phänomenale Transparenz als ein Alles-oder-Nichts-Phänomen. Wenn man der Phänomenologie des bewussten Erlebens Gerechtigkeit widerfahren lassen will, benötigt man jedoch eine differenziertere Beschreibung. Wie ich am Ende dieses Beitrags zeigen werde, ergeben sich aus einer solchen Beschreibung dann aber auch weitreichende Konsequenzen für die Logik und den epistemischen Status des reflexiven *Selbstbewusstseins*.

Lassen Sie uns daher nochmals phänomenologische Beispiele näher betrachten, in denen die Bedingung der Transparenz in unterschiedlichen Graden erfüllt ist. Sinnliche Erfahrung ist das paradigmatische Beispiel für vollständig transparenten phänomenalen Gehalt. Es gibt jedoch auch Beispiele sensorischer Undurchsichtigkeit, etwa während der äußerst kurzen Übergangsphasen in bistabilen Phänomenen, z.B. wenn ein bewusst wahrgenommener Necker'scher Würfel von einer Interpretation zur anderen springt oder im binokulären Wettstreit (Leopold und Logothetis 1999). Auf der nächsten Stufe des sensorischen Bewusstseins finden wir gibt es das Phänomen der einfachen sensorischen Pseudohalluzination. Ein Beispiel. Wenn eine Versuchsperson in einem wissenschaftlichen Laborexperiment unter dem Einfluss einer halluzinogenen

psychoaktiven Substanz (etwa LSD oder 2-CB) abstrakte geometrische Muster an der Wand beobachtet, die zu atmen scheinen und von ihr wie ein lebendiges Fluten in immer tiefere und tiefere Formen von Schönheit erlebt werden, dann wird sie sich normalerweise des repräsentationalen Charakters ihrer visuellen Erfahrung in dem betreffenden Teilbereich des phänomenalen Raums bewusst sein: Sie merkt sofort, dass etwas „nicht stimmt“. Wir besitzen heute erste mathematische Modelle, die die selbstorganisierende Dynamik im visuellen Cortex beschreiben, durch die solche instabilen Zustände gekennzeichnet sind. Diese Modelle machen präzise Voraussagen über die phänomenologischen Formkonstanten, die solche einfachen, kontextinvarianten und sensorisch opaken Zustände kennzeichnen und sie lassen sogar Rückschlüsse auf die funktionale Architektur des visuellen Systems zu (vgl. etwa Bressloff, Cowan, Golubitsky, Thomas und Wiener 2001, Metzinger 2003, Abschnitt 4.2.4, insbesondere S. 243ff). Normalerweise wird die Versuchsperson unmittelbar Zweifel an der Veridikalität ihres Erfahrungszustands zweifeln, ihn sofort kognitiv „einklammern“ und die „Existenzannahme“ zurücknehmen, welche die visuelle Erfahrung in Standardsituationen ganz natürlich auf anstrengungslose und automatische Weise begleitet. Ich sage, dass unsere Versuchsperson das nur deshalb tun kann, weil bereits vorher etwas geschehen ist, das in Standardsituationen nur extrem selten passiert: Frühere Verarbeitungsstufen in ihrem visuellen System wurden zum Gegenstand der introspektiven Aufmerksamkeit. Die atmenden Muster *sind* diese früheren Verarbeitungsstufen. Die funktionale Architektur des Gehirns ist

vorübergehend verändert worden. Das pharmakologisch perturbierte System versucht nun verzweifelt wieder in einen stabilen Zustand zu gelangen, und es ist dieser Vorgang der dynamischen Selbstorganisation, den der Halluzinierende erstmals phänomenal – aber *subsymbolisch*, also *nicht-propositional* – repräsentieren kann. Es gibt allerdings auch komplexe, wahnhaftige Halluzinationen, die vollständig transparent sind und in denen sich die Versuchsperson in einem alternativen Modell der Wirklichkeit verliert (siehe z.B. Siegel und West 1975, Metzinger 2003, Kapitel 4 und 7). Wichtiger ist jedoch, dass das *paradigmatische* Beispiel für vollständig undurchsichtige Klassen von Zuständen absichtlich herbeigeführte Vorgänge bewussten Denkens sind. In diesen Fällen erleben wir uns als Wesen, die absichtlich abstrakte mentale Repräsentationen konstruieren und mit ihnen operieren. Weil wir erleben, dass sie sich jederzeit als falsch erweisen können, erkennen wir sie *als* Repräsentationen. Wir erleben uns gleichzeitig als kognitive und epistemische Agenten, die als denkende Subjekte eine Erweiterung ihres Wissens zu erlangen versuchen. Insbesondere sind wir uns introspektiv früherer Verarbeitungsstufen bewusst, nämlich der *Entstehung* von Gedanken: Wir können die introspektive Aufmerksamkeit auf die *Bildung* kognitiver Inhalte lenken. Kognitive Referenz ist phänomenal undurchsichtig.

Es gibt natürlich auch Klassen von bewussten Gedanken, die sich am anderen Ende des Spektrums zwischen phänomenaler Transparenz und phänomenaler Undurchsichtigkeit befinden. Wenn wir in manifeste Tagträume ableiten, erleben wir die kognitive Verarbeitung häufig nicht mehr *als* solche.

Eine weitere interessante Tatsache mit Blick auf die Phänomenologie von Transparenz und Opazität ist, dass nicht nur individuelle phänomenale Inhalte einen variablen Grad an Transparenz aufweisen, sondern zum Beispiel auch globale phänomenale Modelle der Realität als Ganzer. Direkt nach einem Verkehrsunfall kann uns die Welt „unwirklich“ oder „wie im Traum“ erscheinen. Verwandte Phänomene kennt man von Stress-Situationen und von Übergangsphasen im Verlauf bestimmter psychiatrischer Syndrome. Doch das beste und grundlegendste Beispiel für einen fast vollständig undurchsichtigen globalen phänomenalen Zustand ist der luzide Traum (LaBerge und Gackenbach 2000, Metzinger ²1999, 2003, Abschnitt 7.2.5).

Kehrt man von globalen phänomenalen Zustandsklassen zurück zu Beispielen mittlerer Komplexität, so ist die Analyse der Transparenz und Undurchsichtigkeit von *Emotionen* von besonderem Interesse. Im Gegensatz zu sinnlichen und kognitiven Zuständen sind Emotionen weder für Transparenz noch für Undurchsichtigkeit ein paradigmatisches Beispiel. Unsere emotionalen Zustände erscheinen häufig als unmittelbar gegebene Formen subjektiver Selbstrepräsentation: Ihren Inhalt ziehen wir nicht in Zweifel, sondern gehen einfach davon aus, dass es sich dabei um zuverlässige, unmittelbar gegebene Information über unseren eigenen aktuellen Zustand, über unsere Interessenlage und über unser Verhältnis zu anderen Menschen handelt. Doch mitunter wird uns plötzlich bewusst, dass unsere eigenen emotionale Reaktion unserem gegenwärtigen sozialen Umfeld in Wirklichkeit unangemessen sein könnte.

Typischerweise geschieht das, indem uns *Vorstufen* der eigentlichen Reaktion bewusst werden. Nehmen wir Eifersucht als Beispiel. Wir können den repräsentationalen Charakter unserer eigenen Eifersucht ganz plötzlich feststellen, wenn uns bewusst wird, dass es sich in Wirklichkeit um eine *Fehlrepräsentation* handeln könnte - beispielsweise derjenigen Personen in unserem sozialen Umfeld, auf die wir eifersüchtig sind. Dann wird – genau wie bei der Halluzination, die *als* Halluzination erkannt wird - das, was vorher naiv-realistisch als Eigenschaft der jeweiligen Person erlebt wurde, plötzlich als ein Problem dargestellt, das man möglicherweise *selbst* hat, als Inhalt eines *eigenen* Zustands. In Emotionen schwanken wir häufig zwischen Gewissheit und Ungewissheit, zwischen der Unmittelbarkeit einer auf der Hand liegenden Wahrnehmung und nicht-kognitiven Formen des Zweifels. Diese einfache phänomenologische Beobachtung enthält einen wichtigen Hinweis auf eine andere Eigenart undurchsichtiger phänomenaler Repräsentationen: Sie machen die Möglichkeit, dass es sich in Wirklichkeit um Fehlrepräsentationen handeln könnte, global verfügbar für Kognition, Aufmerksamkeit und Verhaltenskontrolle. In ihnen erlebt sich das Repräsentationssystem erstmals *als* Repräsentationssystem.

Zusammenfassend vertrete ich drei Thesen bezüglich der phänomenalen Transparenz:

- Transparenz ist keine notwendige Bedingung für Phänomenalität, denn opake Formen phänomenalen Gehalts existieren (Beispiele: Phänomenales *Weltmodell*: Pseudo-Halluzinationen, luzide Träume; Phänomenales

Selbstmodell: kognitive Bezugnahme).

- Transparenz ist eine *graduelle* Eigenschaft phänomenaler Zustände, denn die Verteilung von Transparenz/Opazität im phänomenalen Raum besitzt eine *Varianz*. (Beispiele: Phänomenales *Weltmodell*: Stresssituationen; Übergangsphänomene bei psychiatrischen Störungsbildern; Phänomenales *Selbstmodell*: Emotionen).
- Eine überzeugende Theorie phänomenaler Transparenz muss – phylo- und ontogenetisch – erklären können, wie und warum wir diese Eigenschaft *schrittweise* erwerben konnten. (Mögliche Strategien: Unterschiedliche evolutionäre Optimierungszeiträume für Sensorik, Emotionen und Kognition; begrenztes zeitliches Auflösungsvermögen der „introspektiven“ Metarepräsentation im Gehirn.)

2.3 Drei klassische Äquivokationen

Ich werde jetzt fortfahren, indem ich die drei wichtigsten Äquivokationen bzw. Missverständnisse bezüglich des Begriffs „phänomenale Transparenz“ beschreibe, wie er in **DEF 2** eingeführt wurde. Solche Missverständnisse gibt es, sie sind sogar recht verbreitet, aber eine kurze Klärung kann auch der weiteren Präzisierung und semantischen Anreicherung des Zielbegriffs dienen. Es gibt drei verschiedene und weit 00verbreitete Verwendungsweisen des Begriffs „Transparenz“, zwei in der Philosophie und eine in der Kommunikationstheorie.

Zunächst einmal ist Transparenz kein erkenntnistheoretischer, sondern ein

phänomenologischer Begriff. Er hat insbesondere nichts mit dem cartesianischen Begriff epistemischer Transparenz zu tun, also der philosophischen Intuition, dass ich mich über den Inhalt meines eigenen Bewusstseins nicht täuschen kann, weil die Vorstellung eines unbemerkten Irrtums beim introspektiven Zugriff auf den Inhalt des eigenen Bewusstseins in sich widersprüchlich ist. Descartes gab diesem Gedanken im letzten Abschnitt seiner zweiten Meditation Ausdruck, und der Fortschritt der klinischen Neuropsychologie des letzten Jahrhunderts hat diese klassische Annahme über den menschlichen Geist seit langem unhaltbar gemacht. Wie wir aus dem Studium von Anosognosien, Neglekten und vielen anderen Störungen wissen, sind unbemerkte (und unbemerkbare) Irrtümer bezüglich des Inhalts des eigenen Bewusstseins jederzeit möglich. Der moderne Begriff phänomenaler Transparenz steht jedoch insofern in einem systematischen Zusammenhang zum cartesianischen Projekt, als er einen wichtigen Baustein in einer Theorie darstellt, deren Ziel es unter anderem ist, die überwältigende *intuitive* Kraft hinter dieser falschen Annahme verständlich zu machen: Jedes bewusste System, das unter einem phänomenal transparenten Selbstmodell operiert, wird notwendig die phänomenale Eigenschaft der Selbstheit in einer Weise instantiieren, die das System selbst erlebnismässig nicht transzendieren kann (Metzinger ²1999, 2000, 2003, Abschnitt 6.2.6). Die cartesianische These der epistemischen Transparenz des Bewusstseins ist selbst erkenntnistheoretisch unhaltbar, aber sie hat den großen Vorteil, dass sie die *Phänomenologie* der Selbstgewissheit des Subjekts richtig beschreibt.

Weiter wird Transparenz hier aufgefasst als Eigenschaft phänomenaler Repräsentationen in einem subsymbolischen Medium, d.h. von nicht-linguistischen Entitäten im Lichte einer empirisch plausiblen Theorie mentaler Repräsentation, und nicht als Eigenschaft eines Kontextes. Die zweite mögliche Äquivokation ist die Extensionalitätsäquivokation: Der Begriff „phänomenale Transparenz“ wird nicht als ein Fachterminus verwendet, der in das Feld der formalen Semantik gehört, sondern als neues Konzept der philosophischen Neurophänomenologie. Transparenz als Eigenschaft von Kontexten ist etwas ganz anderes. Extensionale (d.h. *referentiell* transparente) Kontexte werden durch Sätze konstituiert, die durch die Intersubstitutivität koreferentieller Ausdrücke *salva veritate* gekennzeichnet sind und durch eine Implikation auf die Existenz derjenigen Entitäten, die in ihnen erwähnt werden. Intensionale (d.h. *referentiell* undurchsichtige) Kontexte werden dagegen durch Sätze konstituiert, die gekennzeichnet sind durch eine Intersubstitutivität von Ausdrücken mit identischer *Bedeutung salva veritate*. In solchen Kontexten bleibt der Wahrheitswert nicht erhalten, wenn koreferentielle Ausdrücke an der Individuenvariable x oder an der vom Prädikat F eingenommenen Stelle durcheinander ersetzt werden. Undurchsichtige Kontexte werden beispielsweise konstituiert durch eine Bezugnahme auf propositionale Einstellungen, auf zeitliche oder modale Ausdrücke oder durch indirekte Rede. Wer daraus, dass ein bestimmter Ausdruck x innerhalb eines intensionalen Kontextes nicht *salva veritate* durch einen koreferentiellen Ausdruck y substituiert werden kann, schließt, dass x und y nicht

auf denselben Aspekt der Wirklichkeit referieren, begeht den wohlbekannten „intensionalen Fehlschluss“.

Ich werde hier *nicht* über Transparenz als Eigenschaft von Kontexten sprechen. Phänomenale Transparenz kann in nicht-linguistischen Wesen existieren, denen jede Form von kognitiver Referenz oder Kontextgebundenheit fehlt. Phänomenal transparente Repräsentationen könnten sogar in einem Gehirn im Tank auftreten, was man von *referentiell* transparenten Repräsentationen nicht behaupten kann. Trotzdem können natürlich komplexe phänomenale Inhalte bestimmte Formen linguistischer Kontexte mitkonstituieren, insbesondere können dies undurchsichtige phänomenale Inhalte, die zu reflexiven Formen mentaler Selbstrepräsentation höherer Ordnung beitragen (siehe Abschnitt 4).

Interessanterweise gibt es auch hier eine Analogie, eine Verbindung, die zwischen referentieller und phänomenaler Transparenz besteht: In beiden Fällen haben wir so etwas wie eine Implikation der *Existenz* der repräsentierten Entitäten. Sätze, die extensionale Kontexte konstituieren, implizieren die Existenz der Entitäten, die in ihnen erwähnt werden. Vollständig transparente phänomenale Repräsentationen zwingen ein bewusstes System hinsichtlich ihrer Inhalte *funktional* zu einem naiven Realisten zu werden. Alles, was phänomenal transparent repräsentiert wird, muss – das ist die Parallele - von diesem System notwendigerweise als wirklich und zweifellos existierend erlebt werden.

Es gibt eine dritte weite Verwendungsweise des Begriffs „Transparenz“, die nicht mit dem Begriff, wie er hier verwendet werden soll, verwechselt werden

darf, die aber ebenfalls eine interessante semantische Parallele aufweist. Auch dieser Begriff ist gut eingebürgert: Transparenz kann als eine Eigenschaft von künstlichen *Medien* aufgefasst werden. In technischen Telekommunikationssystemen kann Transparenz die Eigenschaft eines Kanals oder ganz allgemein eines Systems für die Übertragung von Information sein. Hier ist ein einfaches Beispiel: Transparenz in diesem Sinne kann eine Eigenschaft eines Servers im Internet sein. Definierende Merkmale des kommunikationstheoretischen Begriffs von Transparenz sind erstens, dass er unmodifizierte Benutzerinformation als Eingabe zulässt; zweitens, dass er Benutzerinformation liefert, die auf der Ausgabeseite hinsichtlich Form und Informationsgehalt unmodifiziert ist; und drittens, dass Benutzerinformation intern durchaus geändert und auf vielfältige Weise weiterverarbeitet werden kann, aber immer in ihre ursprüngliche Form zurücktransformiert wird, bevor sie die Ausgabestufe erreicht, und zwar *ohne kausale Wechselwirkung mit dem Benutzer*. E-mail ist das offensichtlichste Beispiel: Eine einzelne elektronische Nachricht, die man an einen Freund schickt, kann mehrfach umkodiert und in viele unterschiedliche Portionen zerhackt werden, von denen jede dann viele verschiedene „hops“ auf jeweils anderen Pfaden durch das Netz macht, bevor sie auf der anderen Seite wieder zusammengefügt werden. Der Benutzer hat keinen Zugriff auf die subpersonalen Mechanismen, die einer erfolgreichen Kommunikation zwischen Personen zu Grunde liegen. Phänomenale Transparenz, wie sie hier diskutiert werden soll, ist klarerweise keine Eigenschaft technischer

Systeme. Dennoch ist es reizvoll, wieder auf die Parallele hinzuweisen, die sich ergibt, wenn wir das neuronale Korrelat von Bewusstsein oder das bewusste Modell der Wirklichkeit als *Medium der interpersonalen Kommunikation* betrachten: Auch dieses Medium ist transparent, weil die subpersonalen Verarbeitungsmechanismen, die zu seinem gegenwärtig aktiven Gehalt beitragen, attentional nicht für den introspektiven Zugriff verfügbar sind.

Als nächstes benötigen wir ein klareres Verständnis des Begriffs von *Internalität*, der diesem Begriff von introspektiver Verfügbarkeit zu Grunde liegt. Was genau bedeutet es, wenn man sagt, für jeden phänomenalen Zustand sei der Grad der phänomenalen Transparenz umgekehrt proportional zum *introspektiven* Grad der attentionalen Verfügbarkeit früherer Verarbeitungsstufen?

3. Introspektion

Es gibt viele verschiedene Begriffe von Internalität, die für die Analyse phänomenaler und kognitiver mentaler Inhalte relevant sind: *temporale* Internalität (die das „Jetzt“ des bewussten Erlebens konstituiert), *funktionale* Internalität (wie beim Schließen sensomotorischer Schleifen, wo sie komplexe Inhalte erzeugt, etwa das phänomenale Erleben von Leiblichkeit und Situiertheit) oder *repräsentationale* Internalität (wie sie für das Selbstbewusstsein relevant ist). In unserem Kontext ist es wichtig, zwischen der Introspektion als einem Prozess zu unterscheiden, der auf phänomenalen Zuständen erster Ordnung operiert, die lediglich auf *physikalisch* internen Systemzuständen supervenieren, und der Introspektion als

einem Vorgang, der den intentionalen Gehalt dieser Zustände noch einmal als einen Aspekt des Systems selbst *repräsentiert*.

Gleichzeitig müssen wir die begriffliche Unterscheidung zwischen Aufmerksamkeit und Kognition anwenden. Aufmerksamkeit ist ein Selektionsvorgang, der episodisch die Fähigkeit steigert, Information in einem bestimmten Bereich des repräsentationalen Raums zu verarbeiten. Funktional gesprochen ist Aufmerksamkeit eine Zuweisung von inneren, neurocomputationalen Ressourcen. Aufmerksamkeit ist sozusagen eine mentale Art und Weise des *Zoomens*, das innerhalb einer umfassenden Repräsentation der Welt zu einer lokalen Steigerung der Auflösung und der Erhöhung des Detailreichtums dient. Auf einfache sensorische Inhalte - wie zum Beispiel phänomenal erlebte Farben – gerichtet, erzeugt sie in fast allen Fällen einen nicht-begrifflichen mentalen Inhalt, für den wir keine transtemporalen Identitätskriterien besitzen und der deshalb unaussprechlich und kognitiv nicht verfügbar ist (Raffman 1995, Metzinger 1998, Metzinger und Walde 2000). Kognition ermöglicht echtes *Wiedererkennen*, unterstützt sprachliche Formen der Begriffsbildung und erzeugt eine Form von geistigem Inhalt, die Kompositionalität, Syntaktizität, Produktivität, Systematizität, usw. zumindest relativ stark approximiert. Es wird deshalb hilfreich sein, vier verschiedene Begriffe von Introspektion zu unterscheiden, da es nun zwei Lesarten von Internalität und zwei Typen von Metarepräsentation gibt, eine subsymbolische, attentionale (die ihre Gegenstände nur „hervorhebt“, aber keinen mentalen *Begriff*

bildet) und eine kognitive Art (die eine fortdauernde mentale „Kategorie“ für ihren Gegenstand bildet, oder zumindest eine um einen Prototypen zentrierte Region im Zustandsraum des Systems).

- **DEF 3: Introspektion₁ (“externe Aufmerksamkeit”).** Introspektion₁ ist subsymbolische Metarepräsentation, die auf einem präexistierenden, kohärenten Weltmodell operiert. Diese Form von Introspektion ist eine phänomenale Form von Informationsverarbeitung, bei der bestimmte Aspekte eines inneren Systemzustands attentional repräsentiert werden, wobei der intentionale Gehalt dieses Zustandes immer durch einen *als extern* dargestellten Teil der Welt konstituiert wird.
Die damit einhergehende Phänomenologie besteht darin, was wir gewöhnlich als Aufmerksamkeit beschreiben, d.h. die subjektive Erfahrung, den „Blickstrahl“ der Aufmerksamkeit auf einen Gegenstand in unserer Umgebung zu richten. **Introspektion₁** entspricht dem *alltagspsychologischen* Begriff von Aufmerksamkeit.
- **DEF 4: Introspektion₂ (“bewusst erlebte kognitive Bezugnahme”).** Dieser zweite Begriff bezieht sich auf eine begriffliche (oder quasi-begriffliche) Form der mentalen Metarepräsentation, die auf einem präexistierenden, kohärenten Weltmodell operiert. Diese Form von Introspektion ist eine phänomenale Form von Informationsverarbeitung, bei der kognitiv auf bestimmte Aspekte eines inneren Systemzustands Bezug genommen wird, wobei auch hier der intentionale Gehalt dieses Zustandes durch einen *als extern* dargestellten Teil der Welt konstituiert wird.
Phänomenologisch gesehen wird diese Zustandsklasse durch all die subjektiven Erlebnisse konstituiert, bei denen wir nicht nur den „Blickstrahl“ der Aufmerksamkeit auf einen Gegenstand in unserer Umgebung richten, sondern ihn gleichzeitig kategorisierend *wieder erkennen* oder einen neuen mentalen Begriff von ihm bilden: Es ist das bewusste Erleben aktuell stattfindender kognitiver Bezugnahme. Ein gutes Beispiel hierfür ist das, was Fred Dretske (1969) „epistemisches Sehen“ genannt hat.
- **DEF 5: Introspektion₃ (“nach innen gerichtete Aufmerksamkeit” und “innere Wahrnehmung”).** Dies ist eine mentale Form der subsymbolischen Metarepräsentation, die auf einem präexistierenden, kohärenten *Selbstmodell* operiert (zum Begriff „Selbstmodell“ siehe Metzinger ²1999, 2000, 2003). Diese Form von Introspektion nun ist eine phänomenale Form von Informationsverarbeitung, bei der wieder bestimmte Aspekte eines inneren Systemzustands attentional repräsentiert werden, wobei diesmal

aber der intentionale Gehalt dieses Zustandes durch einen *als intern* dargestellten Teil der Welt konstituiert wird. Sie erzeugt eine höherstufige, aber noch präreflexive Form des nicht-begrifflichen Selbstbewusstseins.

Die Phänomenologie dieser Klasse von Zuständen ist das, was wir in vortheoretischen Alltagskontexten ganz einfach als „nach innen gerichtete Aufmerksamkeit“ bezeichnen. Auf der Ebene philosophischer Theorien ist es diese Art von phänomenal erlebter Introspektion, die klassischen Theorien der *inneren Wahrnehmung*, z.B. bei John Locke oder Franz Brentano zu Grunde liegt (eine aktuelle kritische Diskussion findet sich bei Güzeldere 1995).

- **DEF 6: Introspektion₄ (“bewusst erlebte kognitive Selbstbezugnahme”).** Dieser vierte Begriff von Introspektion bezieht sich auf eine begriffliche (oder quasi-begriffliche) Form der mentalen Metarepräsentation, die wieder auf einem präexistierenden, kohärenten Selbstmodells operiert. Diese Form von Introspektion ist ein phänomenaler Vorgang, der begrifflich vermittelte Formen der Selbsterkenntnis begleitet und bei dem kognitiv auf bestimmte Aspekte eines inneren Systemzustands Bezug genommen wird, wobei der intentionale Gehalt dieses Zustandes durch einen *als intern* dargestellten Teil der Welt konstituiert wird.

Die allgemeine Phänomenologie, die zu dieser Art repräsentationaler Aktivität gehört, schließt alle Situationen ein, in denen wir bewusst über uns selbst *als uns selbst* nachdenken (d.h., wenn wir sogenannte Ich*-Gedanken denken; siehe z.B. Baker 1998 und den nächsten Abschnitt). Auf theoretischer Ebene stellt diese letzte Art introspektiver Erfahrung offensichtlich den Fall dar, an dem Bewusstseinsphilosophen traditionellerweise das größte Interesse haben: das Phänomen *kognitiver Selbstbezugnahme*, so wie es im reflexiven Selbstbewusstsein zum Tragen kommt.

Die ersten beiden Begriffe von Introspektion bzw. introspektiver Verfügbarkeit sind eher trivial, weil sie die Innerlichkeit der potentiellen Gegenstände von Introspektion allein über einen einfachen physikalischen Begriff von Internalität definieren. In unserem gegenwärtigen Zusammenhang ist Innerlichkeit *als phänomenal erlebte* von wesentlich größerer Relevanz. Trotzdem ist es vielleicht nützlich, an dieser Stelle noch einmal das ontologische „Prinzip der lokalen

Supervenienz“ für phänomenale Gehalte in Erinnerung zu rufen, das unter empirischen Gesichtspunkten äußerst plausibel ist: Für *alle* Formen von mentaler Repräsentation, auf die in **DEF 3 - 6** Bezug genommen wird, gilt, dass ihr *phänomenaler* Gehalt fixiert ist, sobald alle gleichzeitigen und internen physikalischen Eigenschaften des betreffenden Systems feststehen.

4. Cartesianische Gedanken und kognitive Selbstbezugnahme

Was ist eine *kognitive* Erste-Person-Perspektive im Gegensatz zu einer *phänomenalen* Erste-Person-Perspektive? Was ist ein *kognitives* Subjekt im Gegensatz zu einem *phänomenalen* Selbst? In diesem Abschnitt möchte ich zeigen, dass das erstere jeweils ein Sonderfall des letzteren ist: Der Besitz einer phänomenalen Erste-Person-Perspektive ist eine - und wahrscheinlich die relevanteste - notwendige Bedingung dafür, dass es in einem System mit mentalen Eigenschaften überhaupt zum Auftreten einer *kognitiven* Erste-Person-Perspektive kommen kann. Ich behaupte nun, dass - sofern alle anderen Bedingungen für perspektivische Phänomenalität erfüllt sind - ein bewusstes kognitives Subjekt genau dann entsteht, wenn eine global verfügbare, höherstufige Selbstrepräsentation des Systems als aktuell mit Hilfe selbsterzeugter, quasi-linguistischer und undurchsichtiger mentaler Repräsentationen operierend in das bereits vorhandene transparente Selbstmodell integriert wird. „Undurchsichtig“ und „transparent“ werden jetzt immer im Sinne von **DEF 2** verwendet. Es ist entscheidend, zu einer überzeugenden Analyse dieser besonderen Form

phänomenalen Inhalts zu kommen, weil wir dadurch überhaupt erst ein Verständnis davon erlangen, was es bedeutet, eine *kognitive* Erste-Person-Perspektive im Gegensatz zu einer bloß *phänomenalen* Erste-Person-Perspektive zu besitzen. Eine phänomenale Erste-Person-Perspektive könnte bereits in einem Tier realisiert sein oder auch in einem primitiven künstlichen System, das unter einem vollständig transparenten Mechanismus attentionaler Verarbeitung operiert, wie er im vorigen Abschnitt beschrieben wurde (also in einem System, das eine bewusste Repräsentation des „Lichtkegels der Aufmerksamkeit“ besitzt, diese aber nicht *als* Repräsentation erkennen kann; um Begriff eines „phänomenalen Modells der Intentionalitätsbeziehung“ im Allgemeinen, vgl. Metzinger 2000; 2003, Kap. 6.5). Das dieser Unterscheidung zu Grunde liegende philosophische Problem werde ich jetzt anhand eines Beispiels aus der aktuellen Diskussion illustrieren.

In einer wichtigen und hilfreichen Arbeit (Baker 1998; ich werde aus der gekürzten deutschen Übersetzung zitieren) hat Lynne Baker gezeigt, auf welche Weise der Besitz einer überzeugenden Konzeption der Erste-Person-Perspektive einen möglichen Test für jeden robusten Naturalismus darstellt. Baker differenziert zwischen zwei Klassen von Erste-Person-Phänomenen. Schwache Erste-Person-Phänomene sind diejenigen, bei denen beispielsweise manche Tiere begriffen werden können als Wesen, die unter einem egozentrischen Weltmodell operieren welches das Zentrum ihres eigenen Universums sowie den Ursprung ihrer eigenen Perspektive bildet. Wie Baker zeigt, sind alle empfindungsfähigen

Wesen bewusste Subjekte von Erfahrung, aber nicht alle haben Erste-Person-Begriffe von sich selbst. Für Baker sind nur diejenigen, die solche Begriffe haben, in einem interessanten Sinne selbstbewusst (Baker 1998, p. 328).⁶ Unter Voraussetzung der in Abschnitt 2 eingeführten begrifflichen Werkzeuge ist klar, was sich aus dieser Beschreibung ergibt: Solche Tiere würden ein integriertes, globales und transparentes Modell der Welt benutzen, das funktional durch ein transparentes Selbstmodell zentriert wird, um ihr eigenes Verhalten zu steuern. Wie Baker korrekt ausführt, könnte man von solchen Organismen sagen, sie lösten Probleme durch Anwendung perspektivischer Einstellungen, während sie noch keinen Begriff von sich selbst *als einem Subjekt* besitzen.

Erste-Person-Phänomene in einem stärkeren und interessanteren Sinne sind jedoch nicht nur charakterisiert durch die notwendige Bedingung des Besitzes einer Selbst-Welt-Grenze und der Fähigkeit, zwischen erster und dritter Person zu unterscheiden, sondern durch die Fähigkeit, über diese Unterscheidung auf der *begrifflichen* Ebene zu verfügen und auch tatsächlich Gebrauch von ihr zu machen. Unter der bisher eingeführten Terminologie bedeutet dies jetzt, dass die Existenz

⁶ Lynn Baker charakterisiert schwache Erste-Person-Phänomene folgendermaßen: „Zwei Punkte sind hinsichtlich der schwachen Ich-Phänomene zu beachten: (i) Sie treten bei empfindungsfähigen Organismen auf, die mittels perspektivischer Einstellungen Probleme lösen; diese Einstellungen erklären dann das problemlösende Verhalten. (Dies gilt unabhängig von jeder Theorie darüber, wie oder ob diese Einstellungen im Gehirn explizit repräsentiert werden.) (ii) Es bedarf keines Begriffs von einem Ich, damit Wunsch, Überzeugung und Verhalten demselben Organismus zugehören.[...] Obwohl solch ein Tier Wünsche und Überzeugungen hat, hat es weder einen Begriff von Wünschen und Überzeugungen noch einen Begriff seiner selbst als eines Subjekts oder Trägers von Wünschen und Überzeugungen. Es handelt aus seiner eigenen Perspektive heraus, ohne einen Begriff davon, dass es eine von anderen verschiedene Perspektive besitzt.“ (Baker 2000: 253f).

einer präattentiven Grenze zwischen Selbst und Welt und des Unterschiedes zwischen Zuschreibungen in der ersten und dritten Person gemäß DEF4 und DEF6 *kognitiv verfügbar* ist. Wie Baker darlegt, gehört dazu mehr, als Gedanken zu haben, die unter Verwendung des Wortes „ich“ ausgedrückt werden können.

Notwendig ist vor allem der Besitz eines Begriffs von sich selbst als dem *Denker* dieser Gedanken, als dem *Eigentümer* eines subjektiven Standpunkts. Kurz:

Zwingend erforderlich ist nicht nur die Bezugnahme vom Standpunkt der ersten Person aus, sondern auch die Fähigkeit, sich *selbst* diesen Akt der Referenz mental „zuzuschreiben“ – und zwar während er stattfindet. Baker selbst formuliert den Punkt folgendermaßen:

Ein bewusstes Wesen, das starke Ich-Phänomene zeigt, kann an sich selbst als an sich selbst denken. Es ist nicht ausreichend für ein starkes Ich-Phänomen, dass man zwischen erster Person und dritter Person *unterscheiden* kann; man muss zudem fähig sein, die Unterscheidung zu *konzeptualisieren*, also einen Begriff seiner selbst als seiner selbst zu haben. Die Fähigkeit, an sich selbst als an sich selbst zu denken, ist die Fähigkeit, unabhängig von einem Namen, einer Kennzeichnung oder einem Demonstrativpronomen der dritten Person an sich zu denken. Es ist die Fähigkeit, die Unterscheidung zwischen sich und allem anderen in der Welt zu *konzeptualisieren*. Sie beinhaltet nicht nur, Gedanken zu haben, die sich unter Gebrauch von „ich“ ausdrücken lassen, sondern auch, sich als den Träger dieser Gedanken zu begreifen. [...]...der Besitz einer Perspektive oder eines subjektiven Standpunkts allein genügt nicht für ein starkes Ich-Phänomen. Man muss sich vielmehr als jemanden mit einer Perspektive oder einem subjektiven Standpunkt begreifen können. (Baker 2000: 254f)

Diese begriffliche Unterscheidung ist wichtig für die Kognitionswissenschaft im Allgemeinen, aber auch für den philosophischen Begriff eines echten kognitiven Subjekts. Wie Baker anmerkt, könnte dieses Vermögen auch von Bedeutung sein

für sogenannte „Theorie-des-Geistes-Aufgaben“ [*theory-of-mind tasks*]: Nur wenn man in der Lage ist, über sich selbst als den Denker und das Subjekt von Erste-Person-Gedanken zu denken, ist man auch in der Lage den Begriff *anderer* Subjekte von Erste-Person-Gedanken zu bilden. Das zur Diskussion gestellte Vermögen würde einen auch befähigen, seine eigenen Wünsche *als seine eigenen* zu erfassen - beispielsweise als volitionale Akte zweiter Ordnung im Sinne von Harry Frankfurt (Frankfurt 1971).

Im Lichte des hier vorgeschlagenen Modells ist klar, was Baker in Wirklichkeit fordert. Die für kognitive Subjektivität erforderliche repräsentationale Architektur ist nicht nur eine der komplexen Selbstrepräsentation, die gewisse kognitive Aktivitäten als die *eigenen* darstellt und außerdem noch eine mentale Repräsentation des Systems selbst als *Initiator* dieser inneren Ereignisse beinhaltet, d.h. die mentale Repräsentation eines kognitiven Agenten. Was erforderlich ist, ist eine repräsentationale Struktur, die eben diese Tatsache noch einmal global und deshalb auch für *Kognition höherer Ordnung* verfügbar macht.

Interessanterweise wird mit dem bereits an anderer Stelle vorgeschlagenen Begriff eines phänomenalen Selbstmodells (PSM; siehe Metzinger ²1999, 2000, 2003, Kapitel 6) genau dieses Vermögen erfasst. Wenn ein System seine eigenen Operationen mit phänomenal undurchsichtigen mentalen Repräsentationen, d.h. mit mentalen *Simulationen* propositionaler Strukturen, die wahr oder falsch sein können, in sein bereits bestehendes transparentes Selbstmodell integriert und sich

selbst dabei zugleich als kausalen Ursprung dieser Repräsentationen darstellt, dann wird das System als Ganzes - falls alle anderen notwendigen Bedingungen für phänomenales Bewusstsein erfüllt sind – sich nicht nur automatisch als kognitiven Agenten erleben, als ein aktiv denkendes Wesen und Subjekt seiner *eigenen* Gedanken. Es wird diesen neuen phänomenalen Inhalt jetzt auch funktional für die kognitive Verarbeitung höherer Ordnung und für den linguistischen Zugriff verfügbar machen können: Es besitzt jetzt zusätzlich die Fähigkeit einen mentalen Begriff von sich selbst *als* einem kognitiven Agenten zu bilden. Was vorher ein subpersonaler repräsentationaler Prozess war wird jetzt zu etwas, das *das System als Ganzes* sich mit Hilfe seines transparenten Selbstmodells *angeeignet hat*. Einer der Vorteile der Selbstmodell-Theorie der Subjektivität (SMT) besteht deshalb darin, dass sie uns nicht nur eine empirisch plausible repräsentationalistische Analyse der kognitiven Selbstbezugnahme anzubieten hat, sondern auch darin, dass sie uns sagt, was propositionale Einstellungen *de se* (Chisholm 1981, Lewis 1979) in Wirklichkeit sind.

Doch, wie Baker schreibt:

Wenn ich mir einen Selbstbezug [im Original: *first-person reference*] zuschreibe, kann mein Satz nicht in angemessener Weise durch einen Satz wiedergegeben werden, in dem mir kein Selbstbezug [im Original: *first-person reference*] zugeschrieben wird: Die Selbstzuschreibung [im Original: *the attribution of first-person reference to oneself*] scheint nicht eliminierbar zu sein. (Baker 2000: 258)

An dieser Stelle möchte ich eine Alternative anbieten. Erstens: So etwas wie Selbste gibt es nicht in der Welt. Alles, was existiert, sind bewusste Systeme, die

unter phänomenal transparenten Selbstmodellen operieren. Zweitens: Es muss klarer zwischen dem Gehalt rein sprachlicher und dem Gehalt geistiger Repräsentationen unterscheiden. Sprachlicher Selbstbezug ist immer Bezug auf den phänomenalen Gehalt des Selbstmodells, meistens auf seinen transparenten Teil (das PSM ist das, was sprachlichen Selbstbezug funktional überhaupt erst ermöglicht). Weil der transparente Teil aber nicht *als* eine Form repräsentationalen Gehalts erlebt werden kann, kann auch die Tatsache, dass sich hier ein sprachlicher Gehalt auf einen mentalen Gehalt bezieht, nicht bewusst erlebt werden. Sie muss aber einen Eingang in die Theorie des Selbstbewusstseins finden. Eine Art, den von Baker gemeinten Typ von Satz im Sinne des hier zu Grunde gelegten Modells zu paraphrasieren, wäre deshalb: "Dieses System verwendet soeben einen Satz in einer öffentlichen Sprache, um sich auf eine gewisse Fähigkeit zu beziehen, nämlich die Fähigkeit, durch höherstufige kognitive Operationen auf den Inhalt gewisser undurchsichtiger, kognitiver Zustände zuzugreifen, welche in ein bereits existierendes transparentes Selbstmodell integriert wobei sind, wobei diese höherstufigen Operationen selbst ebenfalls automatisch in dieses Selbstmodell integriert werden ". Lassen Sie mich diesen Gedanken nun näher erläutern.

Nach Baker besteht der Besitz einer Erste-Person-Perspektive im philosophisch interessanteren Sinne im Besitz einer gewissen *Fähigkeit*. Es ist die Fähigkeit, Ich*-Sätze oder Ich*-Gedanken zu bilden (die Verwendung des Sterns

folgt hier der üblichen, von Hector-Neri Castañeda 1966 eingeführten Notation).⁷

Wie Baker korrekt hinzufügt, gehört dieses Vermögen zu einer vollständigen

Herrschaft über das „Ich“. Sie schreibt:

The first-person perspective is a necessary condition for any form of self-consciousness, and a sufficient condition for one form of self-consciousness as well (Baker 1998, p. 333).

Baker fährt nun fort, indem sie ihren Punkt durch ein besonders schönes Beispiel illustriert:

Zum Beispiel gibt es keine Möglichkeit, den cartesianischen Gedanken „Ich bin mir gewiss, dass ich* existiere“ in der dritten Person zu artikulieren. Descartes erhob den Anspruch auf die Gewissheit, dass er* existiere, nicht auf die Gewissheit, dass Descartes existiere. (Baker 2000: 269)

Was war in Wirklichkeit der Grund für Descartes' Gewissheit über seine eigene

Existenz? Er besaß ein transparentes Selbstmodell, das ihn zu einem

„autoepistemisch geschlossenen“ Wesen machte: Der repräsentationale Charakter

eines beträchtlichen Teils seines Selbstmodells war ihm prinzipiell nicht

verfügbar, d.h. er konnte keine *introspektive* Kenntnis darüber erlangen.

Autoepistemische Geschlossenheit, beschreibbar als maximale Erfülltheit der

Transparenzbedingung für große Teile des menschlichen Selbstmodells, führt mit

begrifflicher Notwendigkeit zu der präreflexiven phänomenalen Eigenschaft der

⁷ Die Fähigkeit, in dem eben beschriebenen Sinne einen Begriff seiner als seiner selbst* zu haben, ist sowohl notwendig als auch hinreichend für den starken Grad von Ich-Phänomenen: Ein Individuum, das der Ort von starken Ich-Phänomenen ist, kann sich als Träger von Ich-Gedanken verstehen. Es offenbart nicht nur die Fähigkeit zum Selbstbezug, sondern auch die zur Selbstzuschreibung von Selbstbezug. (Baker 2000: 259)

„Selbstheit“, die ganz einfach die Gewissheit über die eigene Existenz *ist*.⁸ Jedes

⁸ Die Transparenz des Selbstmodells ist eine besondere Form *innerer* Dunkelheit. Sie besteht darin, dass der repräsentationale Charakter der Inhalte des Selbstbewusstseins der subjektiven Erfahrung nicht zugänglich ist. Diese Aussage bezieht sich auf all die unterschiedlichen sinnlichen Modalitäten, die die inneren Informationsquellen, aus denen das PSM erzeugt wird, konstituieren: Viszerale Sensibilität, kinästhetische und propriozeptive, taktile und haptische oder auch vestibuläre Reize. Sie gilt aber auch für selbstgerichtete visuelle, auditive, taktile, olfaktorische oder gustatorische Erfahrung, für Selbstwahrnehmung durch Informationsquellen. Insbesondere gilt sie für das integrierte Selbstmodell des Organismus als ganzen: Das Instrument der Selbstrepräsentation ist als solches nicht global verfügbar, und das ist der Grund, weshalb das erfahrende System notwendig in einen naiven Realismus verwickelt wird hinsichtlich der Inhalte seiner eigenen Selbstrepräsentation. Vielen Formen der Selbstrepräsentation sind vorhergehende Verarbeitungsstufen nicht attentional verfügbar. Durch eine kontingente Eigenschaft ihrer repräsentationalen Architektur muss sie sich selbst einfach so erfahren, als hätte sie unmittelbaren Kontakt zu den Inhalten ihres Selbstbewusstseins. Die Tatsache, die dem System nicht zugänglich ist, wenn es von den ihm eigenen Ressourcen Gebrauch macht, ist die, dass seine Selbsterfahrung in einem Medium stattfindet. Vollständig transparente Selbstrepräsentation ist dadurch gekennzeichnet, dass die Mechanismen, die zu ihrer Aktivierung geführt haben, und die weitere Tatsache, dass eine konkreter innerer Zustand existiert, der als Träger ihres Inhalts fungiert, nicht mehr erkannt werden können. Daher ist die Phänomenologie der transparenten Selbstmodellierung die Phänomenologie der Selbstheit. Es ist die Phänomenologie eines Systems, das in einem naiv-realistische Selbstmissverständnis befangen ist. Phänomenale Selbstheit ist das Ergebnis *autoepistemischer* Abgeschlossenheit, sie ist ein Mangel an Information. Die präreflexive, präattentive Erfahrung jemand zu sein ergibt sich unmittelbar daraus, dass die Inhalte des gerade aktiven Selbstmodells transparent sind. Sofern alle weiteren notwendigen Bedingungen für die Emergenz phänomenaler Erfahrung in dem Feld, das durch diese Klasse von Systemen konstituiert wird, realisiert sind, so wird jedes System, das unter einem transparenten Selbstmodell operiert, sich selbst notwendig so erfahren, dass es sich in unmittelbarem Kontakt zu sich selbst befindet. Die phänomenale Eigenschaft der Selbstheit wird durch transparente, nicht-epistemische Selbstrepräsentation konstituiert, und auf dieser Ebene der repräsentationalistischen Analyse wird die Widerlegung des entsprechenden phänomenologischen Irrtums wirklich radikal, da sie eine direkte ontologische Interpretation hat: Nichts von der Art eines Selbst existiert in der Welt. Unter einem allgemeinen Prinzip ontologischer Sparsamkeit ist es nicht notwendig (oder vernünftig), die Existenz von Selbsten anzunehmen, da ihnen als theoretischen Entitäten kein unverzichtbarer Erklärungswert zukommt. Was existiert, sind informationsverarbeitende Systeme, die von einem transparenten Prozess phänomenaler Selbstmodellierung in Anspruch genommen sind. Alles, was durch den phänomenologischen Begriff eines „Selbst“ erklärt werden kann, lässt sich auch erklären, wenn man den repräsentationalistischen Begriff eines transparenten Selbstmodells verwendet.

Man beachte, dass es in Menschen Abschnitte des PSM gibt, die immer undurchsichtig sind. Während Formen der Selbstmodellierung höherer Ordnung wie selbstgerichtete Aufmerksamkeit immer noch in einem naiv-realistischen Selbstmissverständnis befangen sein mögen - man „sieht“ einfach auf oder „fühlt sich ein in“ den Teil seiner selbst, mit dem man in unmittelbarem Kontakt zu stehen glaubt -, ist selbstgerichtete *Kognition* typischerweise durch die Phänomenologie der Einsicht gekennzeichnet. Man scheint einfach zu wissen, dass all dies ein fortlaufender Prozess der Selbstrepräsentation ist, während man gleichzeitig gewiss ist, dass man selbst diesen Prozess initiiert hat. Dies ist natürlich die klassische Intuition Descartes', und die interessante Frage besteht darin, inwiefern es einfach eine andere Variante des phänomenologischen Irrtums ist - der ungerechtfertigte Gebrauch eines Existenzquantors in einem psychologischen Operator.

repräsentationale System, das die Bedingungen für bewusste Erfahrung erfüllt, während es unter einem transparenten Selbstmodell operiert, *muss* seiner eigenen Existenz erlebnismässig gewiss sein. Dies ist in allen möglichen Welten eine begriffliche Notwendigkeit für die beschriebene Klasse repräsentationaler Systeme. Der sprachliche Ausdruck „Ich*“ bezieht sich stets auf den Gehalt des bewussten, gegenwärtig aktiven und transparenten Selbstmodells.

Baker hat natürlich Recht, wenn sie darauf hinweist, dass Descartes' Behauptung sich nicht auf eine Dritte-Person-Repräsentation der Person Descartes bezog. Sie bezog sich auf den Gehalt einer mentalen Repräsentation von Descartes „im Modus der Selbstheit“, unter einer „phänomenalen EGO-Weise des Gegebenseins“ (siehe z.B., Newen 1997). Dies Weise des Gegebenseins ist der subsymbolische, nicht-begriffliche Modus der inneren, transparenten Selbstmodellierung. Sie führt zu einem phänomenalen Identitätserleben, setzt aber keine sprachliche Identifikation voraus. Der phänomenale Gehalt dieses transparenten Selbstmodells - insbesondere auch das charakteristische Merkmal seiner anscheinend unmittelbaren Gegebenheit, das „Sich-selbst-unendlich-nahe-sein“ - ist kognitiv verfügbar. Wenn das so ist, dann kann selbstverständlich auch ein zusätzlicher kognitiver Gehalt höherer Ordnung gebildet werden, der sich auf eben diese Tatsache bezieht. Wenn dieser Inhalt nun seinerseits (z.B. durch subpersonale Mechanismen der dynamischen Selbstorganisation repräsentationaler Strukturen im Gehirn) in das bereits bestehende phänomenale Selbstmodell eingebettet wird, dann wird das hier relevante bewusste Erlebnis,

dass man selbst diesen Gedanken *hat*, ebenso automatisch folgen. Ein zusätzliches und nicht-begriffliches Erleben einer Besitzbeziehung taucht auf, als eine neue Form aktiven phänomenalen Inhalts höherer Ordnung, als ein neuer Teil des phänomenalen Selbst. Dies ist Baker's Ich*-Gedanke. Sein Inhalt ist:

- **[Ich bin mir gewiss, dass ich* existiere]**

Ab jetzt werde ich mich auf aktiven phänomenalen Gehalt beziehen, indem ich (wie oben) eckige Klammern verwende, und auf linguistische Ausdrücke, die sich ihrerseits *auf* aktiven phänomenalen Inhalt beziehen, indem ich spitze Klammern verwende. Der *Ausdruck* <Ich> bezieht sich also auf die Tatsache, dass dieser aktuelle bewusste Gedanke eine Komponente des Selbstmodells ist. Diese Tatsache ist aber gleichzeitig der kognitiven Introspektion verfügbar, die gemäß **DEF 6** auf dem Selbstmodell operiert. Phänomenal ist sie dadurch repräsentiert, dass man sich als der Denker dieses Gedankens erfährt (mehr dazu weiter unten). <Ich*> dagegen bezieht sich auf den transparenten Teil des Selbstmodells. Die vom phänomenalen Selbst ausgesagte Gewissheit und insbesondere die Existenzannahme, die stets mit ihr einhergeht, bezieht sich auf eine global verfügbare mentale Repräsentation der phänomenalen Qualität „unendlicher Nähe“, auf das basale phänomenale Ichgefühl, also die unbezweifelbare (d.h. *phänomenal* notwendige) Erfahrung der Selbstheit, der unmittelbaren Sich-selbst-Gegebenheit die durch die Transparenz des Selbstmodells verursacht ist.⁹

⁹ Lassen Sie mich diesen Punkt illustrieren. Bakers Kritik an Flanagan (1992, 194 ff.) ist

Der Satz <Ich bin mir gewiss, dass ich* existiere> bezieht sich damit auf eine komplexe, integrierte und global verfügbare Form phänomenalen mentalen Gehalt, nämlich auf das aktuell aktive PSM der betreffenden Person. Mit einem geeigneten Ausgabegerät könnte er von einem isolierten Gehirn im Tank geäußert werden, weil dieser Inhalt lokal superveniert. Wir erinnern uns: Der phänomenale (nicht aber der intentionale oder epistemische) Gehalt einer mentalen Repräsentation steht dann fest, wenn alle gleichzeitigen funktionalen Eigenschaften des Gehirns feststehen. Der Inhalt dieses Selbstmodells ist also [Ich bin mir gewiss, dass ich* existiere]. Von jetzt an werde ich mich nur noch mit diesem phänomenalen mentalen Inhalt beschäftigen und nicht mit dem linguistischen Ausdruck, der, wie Lynne Baker es formuliert, die Existenz dieses Inhalts „anzeigt“. Ich möchte meine Leser aber auf die faszinierende Architektur dieser Struktur als Ganzer hinweisen: Wir stehen einer medienübergreifenden Form von Metarepräsentation gegenüber, die die Kluft zwischen propositionalem, sprachlichen und nicht-begrifflichem, phänomenalem Gehalt überbrückt, welche aber die zusätzliche Bedingung erfüllt, dass sie vom repräsentierenden System selbst *nicht als solche erlebt wird*. Eine interne, lokal supervenierende Form von phänomenalem Gehalts wird zum Gegenstand einer *linguistischen* Repräsentation im öffentlichen Raum, wobei der Sprecher sich wegen der Transparenz des

überzeugend, wenn sie darauf hinweist, dass man "*thinking about one's own model of one's own self*" analysieren muss nicht als die kognitive Bezugnahme auf ein Modell von jemand der bloß faktisch ma selbst ist, sondern als die Bezugnahme auf das Modell von einem selbst als einem selbst. [...*as thinking about one's model of oneself as oneself and not just as thinking of a model of someone-who-is-in-fact-onself* (Baker 1998, p. 342)]. Die Selbstmodell-Theorie erreicht genau dies durch Einführung

(mentalen) Inhalts erster Ordnung eben dieser Tatsache nicht bewusst wird.

Betrachten wir den letzten Aspekt nun etwas genauer.

Der Punkt, auf den ich die Aufmerksamkeit lenken möchte, besteht darin, dass der Inhalt des Selbstmodells eine phänomenal transparente und eine phänomenal undurchsichtige Komponente im Sinne von **DEF 2** besitzt. Der Grad der attentionalen Verfügbarkeit vorhergehender Verarbeitungsstufen ist auch hier variabel. Die transparente Komponente des Selbstmodells (z.B. der Inhalt des Körpermodells) wird als unmittelbar gegeben erfahren. Die undurchsichtige Komponente wird häufig, jedoch nicht notwendigerweise, als bewusst konstruiert erlebt, beispielsweise als das eigene Denken. *Beide* Formen von Inhalt sind subsymbolischer Inhalt. Wenn die opake Komponente nicht gerade aus bildhaften Vorstellungen usw. besteht, approximiert sie die charakteristischen Merkmalen sprachlich-symbolischer Repräsentationsmedien wie Konstituentenstruktur, Produktivität, Systematizität, propositionale Modularität (Vgl. Ramsey, Stich, und Garon 1991), usw. hinreichendem stark, um auf der Ebene äußerer autophänomenologischer Berichte typischerweise als eine quasi-linguistische Form mentalen Inhalts interpretiert zu werden („rationales Denken als eine Form inneren Sprechens“). Konzentrieren wir uns jedoch wieder auf den phänomenalen Inhalt selbst und beginnen dabei mit der transparenten Komponente von Bakers cartesianischem Gedanken.

- **[Ich*]**

ist der Inhalt des transparenten Selbstmodells. Dieser Inhalt wird gemäß dem oben bereits erwähnten „Prinzip der autoepistemischen Abgeschlossenheit“ aktiviert: Jedes bewusste System, das unter einem transparenten Selbstmodell operiert, wird mit Notwendigkeit ein phänomenales Selbst instantiiert, auf das es sich linguistisch unter Verwendung des <Ich*> beziehen *muss*. Es ist autoepistemisch gegen frühere Verarbeitungsstufen seines eigenen Selbstmodells abgeschlossen, da es seinen *repräsentationalen* Charakter nicht introspektiv³ entdecken kann. Anders ausgedrückt: In Bakers Beispiel wird der Ich*-Gedanke durch den Inhalt des transparenten Selbstmodells konstituiert, also von einer Teilmenge seiner Gehalteigenschaften. Eine natürliche Hintergrundannahme ist dabei, dass das System bereits alle anderen hinreichenden Bedingungen für einen zumindest minimalen Grad von Bewusstheit erfüllt.

- **[Ich* existiere]**

bezeichnet die entscheidende Eigenschaft des transparenten Selbstmodells (in diesem Kontext). Was sind allgemeine konstitutive Bedingungen für das phänomenale Erleben von „Existenz“? Lassen Sie uns einen *Minimalbegriff* von Selbstbewusstsein definieren: Der Inhalt des Selbstmodells muss in das aktuell aktive Weltmodell eingebettet sein; er muss in einem virtuellen Gegenwartsfenster aktiviert werden; und er muss transparent sein. Daher ist nicht nur die Tatsache, dass das Weltmodell ein *Modell* ist, sondern auch die Tatsache, dass die *zeitliche* Internalität der Inhalte des Gegenwartsfensters (die „Gegenwärtigkeit der Welt“)

ein inneres *Konstrukt* ist, für das Subjekt nicht introspektiv³ verfügbar. Analog dazu gibt es den Sonderfall phänomenaler *Gegenwart des Selbst*, die subjektive Erfahrung von *Anwesenheit*: Wieder ist die Tatsache, dass das Selbstmodell ein *Modell* ist, sowie die Tatsache, dass die zeitliche Internalität seines Inhalts als in das aktuelle Gegenwartsfenster integrierter Struktur ein repräsentationales Konstrukt ist, dem System als Ganzem nicht verfügbar. Diese Eigenschaft ist letztlich eine mikrofunktionale Eigenschaft (attentionale Unverfügbarkeit vorhergehender Verarbeitungsstufen bedeutet, dass die begriffliche Trennung zwischen Gehalteigenschaften und Trägereigenschaften nicht mehr durchgeführt wird).

- **[Ich]**

<Ich> bezieht sich auf den Sprecher eines Satzes. [Ich] dagegen ist der phänomenale Inhalt der *undurchsichtigen* Komponente des aktuellen Selbstmodells, als derjenige, der diesen Gedanken denkt. Die Tatsache, dass der aktuelle kognitive Inhalt nur der einer mentalen *Repräsentation* ist, ist sowohl der Aufmerksamkeit als auch der Kognition introspektiv verfügbar. Das mental selbstrepräsentierende System erfasst sich *als* mental selbstrepräsentierendes. Außerdem haben wir das strukturelle Merkmal kognitiven Handelns: Ich erlebe mich selbst als den *Denker* des Ich*-Gedankens. Die undurchsichtige Komponente des Selbstmodells repräsentiert das jeweilige Subjekt als ein Wesen, das den in Frage stehenden mentalen Inhalt absichtlich in sich selbst erzeugt. Daher modelliert [Ich] das System als Ganzes, es stellt den Inhalt der undurchsichtigen

Komponente dar, aber nicht isoliert, sondern immer schon als in den durchgängig aktiven Hintergrundzustand des transparenten Selbstmodells *eingebettet*. Der phänomenale Inhalt von [Ich] ist also der Denker, der sich aktuell als mit der mentalen Repräsentation operierend repräsentiert. Man beachte, dass der mentale Inhalt, auf den hier Bezug genommen wird, phänomenaler Inhalt ist und deshalb als solcher nicht als epistemisch gerechtfertigt gelten kann.

- **[bin mir gewiss, dass]**

ist das subsymbolische und mentale Äquivalent dessen, was ein „Einstellungsspezifikator“ (glauben, dass *p*; wollen, dass *p*; gewiss sein, dass *p*; usw.) bei der Analyse einer propositionalen Einstellung wäre. Dort wäre es die genaue Beschreibung der *Art* der Relation, die zwischen Person und Proposition besteht. Hier ist es die Beziehung zwischen Selbst und kognitivem Inhalt, wie sie aktuell *phänomenal* repräsentiert wird. Was sind die konstitutiven Bedingungen für phänomenal erlebte Gewissheit? Ich möchte die zwei definierenden Eigenschaften nennen: Die Objektkomponente der phänomenalen Erste-Person-Perspektive ist transparent, und die betreffende Person wird daher auf der Ebene der phänomenalen Erfahrung hinsichtlich des intentionalen Gehalts der Objektkomponente zu einer (epistemisch nicht gerechtfertigten) Existenzannahme gezwungen. Dasselbe gilt von der Subjektkomponente: Die zweite definierende Eigenschaft ist die Transparenz des Selbstmodells, die zu der Entstehung eines phänomenalen Selbst führt, das *als* selbstgewiss dargestellt wird. Eine phänomenale Erste-Person-Perspektive ist eine bewusste Repräsentation

dynamischer Subjekt-Objekt-Beziehungen: Einen Gegenstand zu sehen, einen Schmerz zu empfinden, sich selektiv „einen Gedanken zu eigen zu machen“, ein mögliches Handlungsziel zu wählen - oder seiner selbst als aktuell existierend *gewiss* zu sein.

Sich *seiner selbst* phänomenal gewiss zu sein ist nun ein weiterer Sonderfall. Er tritt dann auf, wenn die Objektkomponente der phänomenalen Erste-Person-Perspektive durch eine transparente *Selbstrepräsentation* gebildet wird, beispielsweise durch ein phänomenales Modell der betreffenden Person *als* eines jetzt existierenden Subjekts. Typischerweise wird die Objektkomponente einer phänomenalen Erste-Person-Perspektive transparent sein, weil genau dies einer der Hauptfaktoren ist, der sie zum Objekt *macht* – etwas, das nicht mehr als eine Repräsentation erfahren werden kann. Wir werden an dieser Stelle also mit einer bewusst erlebten Form von mentaler Selbstrepräsentation zweiter Ordnung konfrontiert. Falls man aus heuristischen Gründen weiter an der Unterscheidung zwischen den Trägern und den Inhalten von mentalen Repräsentationen festhalten will, dann ist es jetzt wichtig zu betonen, dass wir es unter empirischen Gesichtspunkten hier bei den neuronalen Trägern höchstwahrscheinlich nicht mehr mit selbständigen Entitäten zu tun haben, sondern mit Teil-Ganzes-Beziehungen innerhalb eines dynamischen Prozesses.

Auf der repräsentationalen Beschreibungsebene können jetzt starke Erste-Person-Phänomene und cartesianische Ich*-Gedanken als die *Integration* eines undurchsichtigen Selbstmodells in ein präexistentes, transparentes Selbstmodell

analysiert werden. Was wir während der kognitiven Selbstbezugnahme nicht bewusst erleben können, ist die Tatsache, dass wir uns sogar in dieser Situation auf den phänomenalen Inhalt einer Repräsentation beziehen, die „in uns selbst“ ist (weil sie lokal auf Eigenschaften des Gehirns superveniert). Dies ist eine notwendige Konsequenz der Transparenzbedingung für das subsymbolische Selbstmodell. Kognitive Selbstbezugnahme wird deshalb auch auf der phänomenalen Ebene notwendigerweise als unmittelbar und „direkt“ erlebt. Die zwei wichtigsten Gründe für diesen Zusammenhang sind die Tatsache, dass sie nicht durch einen sensorische Kanal vermittelt wird (sie ereignet sich in einem supramodalen Format) und die Tatsache, dass die Information, dass sie in Wirklichkeit ein Prozess phänomenaler Repräsentation *zweiter* Ordnung ist, prinzipiell nicht introspektiv verfügbar ist (das naiv-realistische Selbstmissverständnis). Deshalb ist uns die Möglichkeit, dass die Komponente erster Ordnung immer auch eine *Fehlrepräsentation* sein könnte, auf der Ebene der attentionalen Introspektion³ gemäß **DEF 5** nicht verfügbar.

Kehren wir kurz zur Ebene der linguistischen Selbstbezugnahme zurück. Der Satz <Ich bin gewiss, dass ich* existiere> als Ganzer zeigt an, dass jetzt gerade im Gehirn des Sprechers eine bestimmte, subpersonale Form phänomenalen Gehalts aktiv ist. Aus der Dritte-Person-Perspektive kann dieser Satz nun folgendermaßen analysiert werden:

- **<Im Gehirn des Sprechers dieses Satzes ist aktuell ein transparentes PSM aktiv, in das undurchsichtige Repräsentationen zweiter Ordnung eingebettet worden sind. Diese**

Repräsentationen sind durch drei Eigenschaften gekennzeichnet:
- **Erstens besitzen sie ein quasi-symbolisches Format (z.B. durch die konnektionistische Emulation von Konstituentenstruktur, propositionaler Modularität usw.);**
- **zweitens wird ihr Inhalt ausschließlich durch Operationen auf den transparenten Teilen des gegenwärtig aktiven PSM gebildet;**
- **drittens wird die sich ergebende Beziehung zwischen dem System als Ganzem und dem transparenten Gehalt *phänomenal* als die Beziehung der Gewissheit modelliert.<**

Jetzt kann man sehen, wie es das innere Modell der Beziehung zwischen System und Inhalt ist, welches – in diesem Fall - das bewusste Erleben einer kognitiven Erste-Person-Perspektive erzeugt. Diese kognitive Erste-Person-Perspektive kann aber immer auf einen Sonderfall der *phänomenalen* Erste-Person-Perspektive reduziert werden. Insofern sind alle Überzeugungen darüber *was* hier eigentlich mental repräsentiert wird – also Überzeugungen über den intentionalen oder epistemischen Gehalt starker Erste-Person-Phänomene und cartesianischer Ich*-Gedanken - epistemisch selbst zunächst in keiner Weise gerechtfertigt. Ich will selbstverständlich nicht behaupten, dass

5. Konsistente Fehlidentifikation und Leugnung der eigenen Existenz: Zwei kurze neurophänomenologische Fallstudien

In diesem Abschnitt werde ich im Sinne einer kleinen Serviceleistung für die *philosophical community* kurz auf in der Philosophie des Geistes weitgehend unbekanntes empirisches Material hinweisen, das überzeugend demonstriert, wie bewusste, rationale Subjekte tatsächlich a) sich konsistent mit *anderen* Personen verwechseln oder b) - mit phänomenal absoluter Gewissheit - behaupten können,

sie würden *nicht* existieren. Warum ist das relevant? Wäre Bakers cartesianischer Gedanke eine echte Form epistemischen, intentionalen Inhalts, dann dürfte es eigentlich keine mögliche Welt geben, in der solche Fälle existieren. Die aktuelle Welt ist aber eine mögliche Welt. Wie könnte ein rationales Subjekt sich konsistent über die eigene personale Identität täuschen? Wie könnte ein rationales Subjekt *wissen*, dass es nicht existiert? Wenn die relevante Form mentalen Gehalts nur *phänomenaler* Gehalt ist, der lokal auf seinem minimal hinreichenden neuronalen Korrelat superveniert, dann werden solche Möglichkeiten zumindest denkbar, etwa als Ergebnis einer schweren Fehlfunktion des Gehirns oder einer spezifischen psychiatrischen Störung. Wir könnten dann – mindestens im Prinzip – plausibel machen, wie es möglich ist, dass ein rationales Subjekt sich selbst als eine andere Person oder als nicht-existierend *erleben* könnte.

Das phänomenale Erleben von Selbstheit und kognitiver Selbstbezugnahme kann entlang einer großen Anzahl von Dimensionen variieren. Solche Variationen konstituieren wichtige phänomenologische *constraints*, die von begrifflichen Analysen der besonderen Formen mentalen Gehalts, der in der bewussten, kognitiven Selbstbezugnahme aktiv ist, erfüllt werden müssen. Es gibt zum Beispiel klar umschriebene, aber unbemerkte Verluste von phänomenalem Gehalt (wie bei der Leugnung der eigenen Blindheit, Anton's Syndrom; vgl. Anton 1898, 1899; Benson und Greenberg 1969). Es gibt auch verschiedene Typologien phänomenalen Zerfalls - zum Beispiel in der Schizophrenie, bei DID (von engl. *Dissociative Identity Disorder*) und verschiedene Depersonalisationsstörungen.

Solche Störungsbilder können mit einer Vervielfachung des phänomenalen Selbst in ein und demselben physischen System einhergehen oder mit teils dramatischen Umverteilungen der phänomenalen Eigenschaft der „Meinigkeit“ im phänomenalen Zustandsraum. Vor allem gibt es mindestens vier verschiedene Klassen von Störungen, die mit einer Fehlidentifikation von Personen einhergehen (*delusional misidentification*, DM¹⁰; nämlich das Capgras-Syndrom, das Frégoli-Syndrom, die Intermetamorphose und die reverse Intermetamorphose sowie die reduplikative Paramnesie). Sie sind zum Beispiel deshalb von Bedeutung für die philosophische Theorie des Selbstbewusstseins und der Selbstreferenz, weil sie das Wittgenstein/Shoemaker-Prinzip der Immunität gegen Irrtum durch Fehlidentifikation von verletzten (Wittgenstein 1955, S. 67, Shoemaker 1968, S. 9). Daher ist es wichtig, neue begriffliche Werkzeuge wenigstens anhand *einiger* Beispiele aus dem großen phänomenologischen Reichtum des Zielphänomens „Selbstbewusstsein“ zu prüfen. Was sind notwendige und hinreichende Bedingungen für die kognitiven Formen der Selbstmodellierung, die der linguistischen Selbstbezugnahme zu Grunde liegen? Die philosophische Relevanz von Fallstudien aus der kognitiven Neuropsychiatrie besteht darin, dass sie eine *reverse engineering*, eine umgekehrte funktionale und repräsentationale Rekonstruktion ermöglichen: Wenn wir eine empirisch plausible

¹⁰Indem ich die Abkürzung „DM“ - anstatt „DMS“ (für „delusional misidentification syndromes“) - verwende, übernehme ich eine von Nora Breen und Kollegen eingeführte terminologische Änderung, die die Ungenauigkeit zu vermeiden sucht, die in der Annahme eines ganzen Symptomclusters als definierende Eigenschaften für die verschiedenen Kategorien liegt. Vgl. Breen, Caine, Coltheart, Hendy und Roberts 2000, S. 75, Fußnote 1.

repräsentationalistische Analyse von Identitätsstörungen entwickeln können, werden wir automatisch zu einem besseren Verständnis davon gelangen, was es genau bedeutet, unter neurophänomenologischen Normalbedingungen *jemand zu sein*. Wie die Wissenschaftsgeschichte gezeigt hat, unterstützt bei der Erforschung komplexer Phänomene eine genaue Untersuchung von Grenzfällen die Aufdeckung von impliziten Annahmen, intuitiven Fehlschlüssen und begrifflichen Mängeln bestehender Theorien.

In unserem Zusammenhang sind mindestens zwei Typen von Identitätsstörungen philosophisch relevant: eine bestimmte Form von DM und die Cotard-Störung. Wir wollen mit der *umgekehrten Intermetamorphose* beginnen, die gewöhnlich als „die Überzeugung, man habe sich physisch und psychisch in eine andere Person verwandelt“ definiert wird (Breen, Caine, Coltheart, Hendy, und Roberts 2000, S. 75). Ich werde kurz dafür argumentieren, dass diese Störung auf einer devianten Form phänomenaler Selbstmodellierung beruht und dass sie grundsätzlich auch in einem nicht-linguistischen/nicht-kognitiven Wesen auftreten könnte, das überhaupt nicht in der Lage ist, so etwas wie „Überzeugungen“ im engeren philosophischen Sinn zu bilden. Hier ein Auszug aus einer jüngeren, von Nora Breen und ihren Kollegen vorgelegten Fallstudie über die Patientin Roslyn Z:

RZ, eine 40-jährige Frau hatte die wahnhafte Überzeugung, sie sei ein Mann. Der Wahn hatte sich zwei Monate lang stabil gehalten, bevor wir RZ untersuchten. Die meiste Zeit in diesen zwei Monaten glaubte sie, sie sei ihr Vater, doch gelegentlich pflegte sie zu sagen, sie sei ihr Großvater. Zu der Zeit, als wir RZ sahen, hatte sie die Persönlichkeit ihres Vaters angenommen.

Sie reagierte nur auf seinen Namen, und sie unterschrieb mit seinem Namen, wenn sie gebeten wurde irgendwelche Formulare zu unterschreiben. Sie berichtete kohärent aus der Lebensgeschichte ihres Vaters, wenn sie über ihre persönliche Lebensgeschichte befragt wurde. Zum Beispiel sagte sie, sie sei gerade in ihren Sechzigern. (...) Die folgenden Auszüge stammen aus einem Gespräch mit RZ. Während des ganzen Gesprächs saß RZs Mutter Lil neben ihr.

Untersuchender: Können Sie mir Ihren Namen sagen?

RZ: Douglas.

Untersuchender: Und Ihren Familiennamen?

RZ: B_____.

Untersuchender: Und wie alt sind Sie?

RZ: Daran kann ich mich nicht erinnern.

Untersuchender: Wie alt sind Sie ungefähr?

RZ: Etwas über sechzig.

Untersuchender: Etwas über sechzig. Und sind Sie verheiratet?

RZ: Nein.

Untersuchender: Nein. Waren Sie verheiratet?

RZ: Ja.

Untersuchender: Wie hieß Ihre Frau?

RZ: Ich kann mich nicht erinnern. Lil.

Untersuchender: Lil. Und haben Sie Kinder?

RZ: Vier.

Untersuchender: Und wie heißen sie?

RZ: Roslyn, Beverly, Sharon, Greg.

(...)

RZ steht vor einem Spiegel und betrachtet ihr eigenes Spiegelbild.

Untersuchender: Wenn Sie in den Spiegel dort sehen, wen sehen Sie?

RZ: Dougie B_____ (der Name ihres Vaters).

Untersuchender: Wie sieht dieses Spiegelbild aus?

RZ: Seine Haare sind unordentlich, er hat einen Vollbart, und seine Augen sind ganz matt.

Untersuchender: Ist das denn ein Mann oder eine Frau?

RZ: Ein Mann.

Untersuchender: Wie alt ist Dougie?

RZ: Etwas über sechzig.

Untersuchender: Und sieht das Spiegelbild, das Sie gerade ansehen, aus wie eine Person, die etwas über sechzig ist?

RZ: Ja.

Untersuchender: So alt sieht es aus?

RZ: Ja.

Untersuchender: Meinen Sie ein Mann, der etwas über sechzig ist, hätte graues Haar?

RZ: Nun, ich hatte in all den Jahren wenig Sorgen, daher wurde mein Haar nicht grau.

Untersuchender: Es ist also nicht grau?

RZ: Nein. Es ist braun.

(Breen, Caine, Coltheart, Hendy, und Roberts 2000, S. 94f/98f).

Die umgekehrte Intermetamorphose ist philosophisch interessant, da sie das Prinzip der Immunität gegen Irrtum durch Fehlidentifikation von Wittgenstein/Shoemaker in Frage stellt (Wittgenstein 1955, S. 67, Shoemaker 1968). Ich werde hier allerdings nicht in eine philosophische Diskussion über diese Fälle eintreten. Ich möchte nur illustrieren, was es bedeutet, wenn der Begriff der phänomenalen Transparenz auf das Selbstbewusstsein angewendet wird und auf die resultierende Selbstzuschreibung von *globalen* Eigenschaften des kognitiven Systems als Ganzem auf der personalen Beschreibungsebene. Ganz offensichtlich gibt es Fälle von vorsprachlicher, phänomenaler Selbstrepräsentation der eigenen *personalen Identität*, die Fehlrepräsentationen sind. Sie beruhen auf einer dramatischen Veränderung im subsymbolischen Selbstmodell, die zunächst nichts mit einer sprachlich durchgeführten Selbstidentifikation zu tun hat. Sie haben aber Folgen, wenn sie in sprachfähigen Systemen auftreten. Solche Sprecher, das ist meine These, beziehen sich linguistisch (oder auch kognitiv) immer noch auf den Gehalt eines transparenten Selbstmodells - nur mit dem Unterschied, dass sie jetzt das interne Modell einer *anderen* Person als ihre Selbstrepräsentation erster Ordnung verwenden. Dieses kohärente, aber wahnhaftes Selbstmodell ist der

Gegenstand sprachlicher oder kognitiver Selbstbezugnahme, und die Tatsache, dass es nur ein Modell ist, ist dem System als Ganzem weder kognitiv noch attentional verfügbar.

Man muss zunächst verstehen, dass es sich hierbei um *monothematische* Wahnstörungen handelt, die auf einen bestimmten Gegenstandsbereich beschränkt sind und die Rationalität der Person als Ganzer nicht tangieren (Davies, Coltheart, Langdon, und Breen 2001). Fehlidentifikation ist ein einzelnes *Symptom* und kein Syndrom, das eine stabile Ansammlung von Symptomen umfasst, und diese besondere Variante der konsistenten, vorsprachlichen Selbst-Fehlidentifikation tritt nur im Zusammenhang mit besonders ernsten psychotischen Zuständen auf (Förstl, Almeida, Owen, Burns und Howard 1991, S. 908f). Philosophisch interessant ist insbesondere die Tatsache, dass solche Patienten den Schritt vom Erleben zur Überzeugung auch dann machen, wenn es überwältigende Gründe dafür gibt, dem eigenen phänomenalen (Selbst)Erleben nicht zu vertrauen; dass ihnen vollständig bewusst ist, dass ihre Überzeugung *anderen* Personen als vollkommen unplausibel und irrational erscheinen muss; und dass es innerhalb des kognitiven Systems kein *ex falso quod libet* gibt: Typischerweise haben solche Patienten nur eine einzelne wahnhafte Überzeugung (oder eine sehr kleine Menge von solchen, miteinander in Beziehung stehenden Überzeugungen), ihr Überzeugungssystem als Ganzes ist aber in keiner anderen Weise abnormal oder inkohärent. Das Netz propositionaler Einstellungen bleibt stabil. Wollte man die wahnhafte Überzeugung als eine explanatorische

Hypothese interpretieren, dann wäre es ein isolierter doxastischer Zustand, der weder durch subjektives Erleben, noch durch enzyklopädisches/semantisches Wissen, noch durch rationale Argumente oder andere Personen widerlegt werden kann (Davies, Coltheart, Langdon, und Breen 2001).

Betrachten wir jetzt einen zweiten Typ von Identitätsstörung, der ebenfalls von besonderer theoretischer Relevanz ist. Ich werde sie ab jetzt auch *Existenzleugnung* nennen. Der französische Psychiater Jules Cotard führte 1880 den Ausdruck *délire de negation* ein, um eine bestimmte Art „nihilistischen“ Wahns zu bezeichnen, dessen zentrales definierendes Merkmal darin besteht, dass Patienten ihre eigene Existenz leugnen, häufig sogar die der Aussenwelt (siehe Cotard 1880, genauere Ausführungen finden sich in Cotard 1882). Ab 1879 wurde diese Krankheit in der wissenschaftlichen Literatur als „Cotard-Syndrom“ bezeichnet (Séglas 1879; ein knapper Überblick über die Literatur findet sich in Enoch und Trethowan 1991, S. 163ff). Obwohl der Begriff des Wahns als solcher immer noch recht kontrovers ist (siehe z.B. Young 1999), ebenso wie der begriffliche Status von „pathologischen“ Glaubenssystemen im Allgemeinen (siehe Coltheart und Davies 2000, Marshall und Halligan 1996), tendieren die meisten Forscher übereinstimmend zu der Auffassung, dass das Cotard-Syndrom wahrscheinlich eine distinkte theoretische Entität darstellt.¹¹ Ich werde es hier

¹¹ Young und Leafhead (1996, S. 154) vertreten die Auffassung, dass es kein bestimmtes Symptom gibt, das in allen reinen Cotard-Fällen auftritt, wohingegen Berrios und Luque (1995) eine statistische Analyse der klinischen Verwendung des Ausdrucks vorlegen und daraus schließen, dass ein reines Cotard-Syndrom (so genannte "Cotard Typ-1-Patienten") existiert und

einfach als eine neurophänomenologische Zustandsklasse behandeln, die durch eine bestimmte Form devianter Selbstmodellierung gekennzeichnet ist, ohne mich auf weiterführende empirische Spekulationen einzulassen. Als Form des Wahns ist die Cotard-Störung sicherlich recht dramatisch – zum Beispiel, weil sie auf lokaler Ebene die globale logische Kohärenz des „Überzeugungsnetzes“ eines Patienten verletzt (siehe Young 1999, S. 582f), dabei aber gleichzeitig seiner phänomenal repräsentierten biographischen Kohärenz *ein Ende setzt*. Andererseits stellt sie nur einen sehr begrenzten, modularisierten Verlust innerhalb des kognitiven Realitätsmodells dar, auf das der Patient bewusst zugreifen kann. Wie bei vielen Wahnerscheinungen ist es wieder der *isolierte* und eng umschriebene Charakter eines bestimmten Überzeugungsinhalts, der zunächst Grund gibt, den Status des Patienten als rationales Subjekt ernstlich in Zweifel zu ziehen. Wenn Descartes einem Cotard-Patienten gegenüber darauf hinweisen würde, dass er *selbst* es ist, der seine eigene Nicht-Existenz denkt, dann würde dies – nach allem was wir wissen – nicht zu einer Revision dieser Überzeugung führen.

Eine genauere Betrachtung der Daten macht jedoch schnell deutlich, dass ein Cotard-Patient durchaus als rationales Subjekt gelten kann – nämlich als eines, das den einzig möglichen Schluss aus einer dramatischen Verschiebung in seinem subkognitiven Selbstmodell zieht. Ein vielversprechender Versuch in Richtung auf eine überprüfbare und begrifflich überzeugende Hypothese könnte deshalb von der Annahme ausgehen, dass die Cotard-Störung einfach eine modularisierte

dass sein nosologischer Ursprung „in den wahnhaften und nicht in den affektiven Leiden

kognitive Reaktion auf ein sehr ungewöhnliches Wahrnehmungserleben ist (Young und Leafhead 1996). Es muss jedoch auch darauf hingewiesen werden, dass die Phänomenologie des festen Überzeugtseins von der eigenen Nicht-Existenz sich als zu subtil und zu komplex erweisen könnte, um überhaupt im Rahmen klassischer *belief-desire*-Ansätze behandelbar zu sein - beispielsweise, weil sie in entscheidender Weise auf einer Pathologie der *nicht-propositionalen* phänomenalen Selbstrepräsentation beruht.

Bei reinen, extremen Varianten des Cotard-Wahns kann es sein, dass Patienten nicht nur ausdrücklich behaupten, sie seien tot, sondern auch, dass sie überhaupt nicht *existieren*. Etwas, das aus logischen Gründen als *a priori* unmöglich erscheint - ein bewusstes Subjekt, das seine eigene Existenz ernsthaft leugnet - erweist sich als phänomenologische Realität. Und die Phänomenologie muss man ernst nehmen. Es gibt nomologisch mögliche repräsentationale Konfigurationen des menschlichen Gehirns, die rationale Subjekte zu wahrhaften, aber logisch inkohärenten autophänomenologischen Berichten führen. Im vorigen Abschnitt habe ich die folgende repräsentationalistische Analyse des cartesianischen Gedankens angeboten:

- **[Ich bin mir gewiss, dass ich* existiere].**

In extremen Formen der Cotard-Störung sind wir mit wahnhaften Überzeugungen des folgenden Typs konfrontiert:

- **[Ich bin mir gewiss, dass ich* *nicht* existiere.]**

liegt;..“ (ebd. S. 187). Ich möchte hier keine Stellung zu dieser Frage beziehen.

Schwächere Formen können folgendermaßen ausgedrückt werden:

- **[Ich bin gewiss, dass ich* tot bin.]**

Was ist das genauere phänomenologische Profil der *Cotard delusion*? Nach einer neueren Analyse von 100 Fällen (Berrios und Luque 1995, siehe Abb. 2, S. 186) wurde bei 89 % der Personen eine schwere Depression festgestellt, wobei sich die häufigsten Formen des „nihilistischen Wahns“ auf den Körper (86 %) und auf die Existenz als solche (69 %) bezogen. Andere sehr verbreitete Eigenheiten des berichteten Inhalts des PSM bei Cotard-Patienten sind Angst (65 %), Schuldgefühle (63 %), hypochondrische Wahnvorstellungen (58 %) und, noch überraschender, Unsterblichkeitswahn (55 %; siehe Berrios und Luque 1995, S. 187). In vielen Fällen scheinen Elemente des körperlichen Selbstmodells verschwunden oder zumindest attentional unverfügbar zu sein. Zum Beispiel sagte ein 59-jähriger Patient wiederholt „Ich habe kein Blut“ (Enoch und Trethowan 1991, Fall 5, S. 172), ein anderer Patient „Früher hatte ich ein Herz. Ich habe etwas, das an seiner Stelle schlägt“ (ebd., S. 173), während eine weitere Fallstudie (Ahleid 1968, zitiert nach Enoch und Trethowan 1991, S. 165) von einem Patienten berichtet, der darum bat „begraben zu werden, da er, wie er sagte ‚ein Leichnam, der schon stinkt‘ sei. Einen Monat später sagte er, er habe kein Fleisch und weder Beine noch Rumpf. Diese Vorstellungen waren unerschütterlich, so dass das klinische Bild monatelang unverändert blieb.“ Solche frühen Stufen entwickeln sich häufig in phänomenologische Zustände hinein, in denen der

Körper als ganzer geleugnet wird und der Patient sich wie ein „lebender Leichnam“ fühlt, z.B. wenn er sagt „Ich lebe nicht mehr“; „Ich bin tot“ (Enoch und Trethowan 1991, Fall 1, S. 168) oder „Ich habe keinen Körper, ich bin tot“ (Enoch und Trethowan 1991, Fall 2, S. 168) oder, wenn er, wie der Patient von Young und Leafhead (Young und Leafhead 1996, 154ff), auch einige Monate nach einem Motorradunfall (bei dem es zu Quetschungen in bestimmten temporo-parietalen Regionen der rechten Hemisphäre und einer bilateralen Schädigung des frontalen Cortex kam) noch fest davon überzeugt ist, tot zu sein, wobei sich diese Überzeugung nur mit der Zeit und allmählich auflöst.

Was an der Cotard-Störung einzigartig scheint, ist die Überzeugung, man sei tot. Der Cotard-Patient erwirbt auf der Ebene seines *kognitiven* Selbstmodells einen neuen, introspektiv⁴ verfügbaren mentalen Gehalt - und diesen Inhalt zeichnet aus, dass er extrem irrational und funktional immun gegen eine rationale Revision ist. Der Unterschied im phänomenalen Inhalt der transparenten und *subkognitiven* Schichten seines Selbstmodells lässt sich auf interessante Weise beschreiben, wenn man eine traditionelle begriffliche Unterscheidung anwendet, die nur in der deutschen Sprache zur Verfügung steht, nicht jedoch im Griechischen (z.B. verwendet Homer „soma“ nur, wenn es um Leichname geht), Lateinischen (*corpus* bezieht sich nur auf den Körper als Ding und ist die etymologische Wurzel des englischen „corpse“) und anderen wichtigen Sprachen wie Italienisch, Französisch, Spanisch und Englisch: Der Cotard-Patient hat nur ein phänomenales Selbstmodell als *Körper*, aber nicht eines als *Leib* (siehe auch

Ahleid 1968, und Enoch und Trethowan 1991, S. 179f). Worin besteht der Unterschied? Ein *Leib* ist ein *gelebter Körper*, ein Körper, der mit einer Seele verbunden ist, oder - um es moderner auszudrücken - der Körper *als Subjekt*, als Ort einer individuellen Erste-Person-Perspektive. Der *Körper als unbelebtes Objekt* ist das, was das PSM in der Cotard-Konfiguration darstellt. Der Cotard-Patient ist subjektiv entleibt, denn er kann introspektiv nur auf ein mentales *Körpermodell* zugreifen, nicht jedoch auf ein *Leibmodell*. Er besitzt keine Erste-Person-Perspektive mehr, sondern nur eine Erstes-Objekt-Perspektive.

Es ist interessant zu sehen, wie man vielleicht ein besseres phänomenologisches Verständnis der zu Grunde liegenden Zustandsklasse erlangen kann, in dem man einfach der traditionellen Linie folgt. Was für eine Art von Ausfall könnte ein *Leibmodell* zu einer Repräsentation von etwas Unbelebtem machen, von etwas, das nicht mehr an die innere Logik des Lebendigseins gebunden ist? Wenn es - wie empirisch plausibel - genau die Logik des Überlebens ist, die durch bewusste Emotionen als repräsentationaler Inhalt global verfügbar gemacht wird (vgl. Damasio 1999), dann folgt, dass ein vollständiger Verlust des *emotionalen* Selbstmodells genau diese Wirkung haben müsste. Philosophisch gesprochen würde dies bedeuten, dass dasjenige, worauf sich der Cotard-Patient in seinen Selbstzuschreibungen bezieht, wenn er behauptet, er sei tot oder ein Leichnam, der transparente Inhalt seines Selbstmodells ist, insbesondere diejenigen Schichten, die mit räumlicher Selbstrepräsentation, mit Propriozeption und Emotion zu tun haben. Dieser repräsentationale Inhalt porträtiert eine sich

bewegende *res extensa*, die einer lebendigen menschlichen Person sehr ähnlich ist, aber innerlich nicht mehr an die Logik des Überlebens gebunden ist. Die eigene Lebendigkeit kann sozusagen nicht mehr bewusst gefühlt werden, weil die emotionale Schicht des Selbstmodells nicht mehr introspektiv verfügbar ist. In traditioneller Terminologie hat der Patient eine korrespondierende Überzeugung *de se*. Da der nicht-begriffliche selbstrepräsentationale Inhalt erster Ordnung, in dem diese Überzeugung verankert ist, wegen der völligen Elimination der emotionalen Schicht auch die Information, dass die elementare biologische Selbstregulation in Wirklichkeit im weiter stattfindet, nicht mehr enthält, stellt der Patient eine Hypothese auf. Diese Hypothese ist unter Voraussetzung seiner internen Informationsquellen absolut kohärent: Er muss ein totes Objekt sein, das einem menschlichen Wesen ähnelt. Obwohl die Existenz dieses Objekts als Zentrum der Welt und Ursprung einer phänomenalen Perspektive erfahren wird, ist sie dem Patienten affektiv *gleichgültig*. Sie geht ihn nichts an. Er kann die *Valenz* dieses Objekts nicht mehr mental repräsentieren. Der australische Philosoph Philip Gerrans (Gerrans 2000, S. 112) schreibt: "The Cotard delusion, in its extreme form, is a rationalization of a feeling of disembodiment based on global suppression of affect resulting from extreme depression." Viele Cotard-Patienten äußern Sätze wie „Ich habe keine Gefühle“ (Enoch und Trethowan 1991, Fall 4, S. 171) oder sagen, wie der Patient KH von Young und Leafhead (Young und Leafhead 1996, 160ff), sie seien tot, da sie „innen nichts fühlen“. Wenn, wie zuvor erläutert, das bewusste Selbstmodell des Menschen Leiblichkeit nicht nur auf der

phänomenalen, sondern auch auf der funktionalen Ebene vermittelt, dann leiden diese Patienten genau deshalb an den beschriebenen funktionalen Defiziten, weil sie wegen einer schweren Störung in ihrem PSM *emotional entleibt* sind. Die Störung besteht darin, dass eine spezifische Teilmenge systembezogener Information nicht mehr durch das Selbstmodell global verfügbar gemacht werden kann. Dieser Umstand löst dann weitere Veränderungen auf der Ebene des kognitiven Selbstmodells aus.

Gerrans (1999, S. 590) hat darauf hingewiesen, dass die Cotard-Störung auf der Ebene repräsentationalen Inhalts begrifflich falsch eingeordnet wird, wenn sie als DM beschrieben wird. Zumindest bei schwächeren Formen sind die Berichte des Cotard-Patienten wahrhaftige Berichte über den Inhalt eines sehr ungewöhnlichen PSM. Dieses ungewöhnliche PSM resultiert aus einem globalen Affektverlust, der, so Gerrans, die globale Verteilung des neurochemischen Substrats widerspiegelt, dessen Mangel das repräsentationale Defizit verursacht, welches seinerseits dann kognitiv interpretiert und zum Gegenstand sprachlicher Selbstbezugnahme wird. Es geht deshalb auch darum, empirisch zu erklären, wie das fragliche PSM auf der subsymbolischen Ebene wirklich zustande kommen kann. Die klassische *belief-desire*-Psychologie und eine traditionelle philosophische Analyse, die propositionale Einstellungen postuliert, sind deshalb möglicherweise nicht sehr hilfreich bei der zu einem vollständigeren Verständnis des anvisierten Phänomens notwendigen Verknüpfung von personalen und subpersonalen Beschreibungsebenen. Es scheint beispielsweise plausibel, dass ein nicht-

linguistisches Wesen wie ein Affe durchaus an den subkognitiven Aspekten der Cotard-Phänomenologie leiden könnte, ohne jedoch in der Lage zu sein, inkohärente autophänomenologische Berichte von sich zu geben: Ein Tier könnte sich tot und emotional entkörperpert *fühlen*, ohne gleichzeitig die Fähigkeit zu besitzen, sich eben diese Tatsache selbst zuzuschreiben, sei es nun linguistisch oder auf der Ebene der kognitiven Selbstbezugnahme. Das Besondere im Fall des Menschen liegt darin, dass ein neues, isoliertes und funktional starres Element auf der Ebene des *kognitiven* Selbstmodells entsteht. Sobald uns eine überzeugende repräsentationale und funktionale Analyse des Selbstmodells beim menschlichen Cotard-Patienten gelingt, können wir zu der philosophischen Frage fortschreiten, ob und wie es möglich ist, starke Erste-Person-Phänomene im Sinne von Lynne Baker (1998) zu exemplifizieren und gleichzeitig die Überzeugung zu vertreten, dass man selbst in Wirklichkeit überhaupt nicht existiert.

Die Cotard-Störung kann man also analysieren als den Verlust einer ganzen Schicht nicht-begrifflichen, transparenten Gehalts (nämlich des emotionalen Selbstmodells) und das gleichzeitige Auftreten eines neuen, quasibegrifflichen und opaken Gehalts im (kognitiven) PSM des Patienten. Gibt es eine spezifischere Art, die kausale Rolle der Information, die jetzt auf der Ebene des phänomenalen Realitätsmodells des Patienten nicht mehr verfügbar ist, zu beschreiben? Was löst die lokale Umstrukturierung seines „Überzeugungsnetzes“ aus? Eine zweite und etwas spezifischere Hypothese ist hängt mit dem Phänomen der „emotionalen Entkörperung“ zusammen: Das PSM des Cotard-Patienten wird allen emotionalen

Inhalts entleert, und dies macht es ihm unmöglich, das bewusste Erlebnis der *Vertrautheit* mit sich selbst zu haben. Was der Cotard-Patient unwiederbringlich verliert, ist möglicherweise genau die eingangs erwähnte phänomenale Qualität der „präreflexiven Selbstvertrautheit“. Der Cotard-Patient ist sich selbst nicht unendlich nahe, sondern unendlich fern. Wenn es außerdem zutrifft, dass der emotionale Inhalt für einen Organismus in erster Linie die Logik des Überlebens repräsentiert, dann wird der *selbstrepräsentationale* emotionale Inhalt die innere Logik der autonomen Selbstregulierung, des *eigenen* Lebensprozesses für den Organismus global verfügbar machen – für die Handlungskontrolle, für die kognitive Weiterverarbeitung, für die introspektive Aufmerksamkeit. Für den Cotard-Patienten ist die Logik des eigenen Überlebens außer Kraft gesetzt: Obwohl sein biologischer Lebensprozess sich funktional als im physischen Körper realisierter auch weiterhin entfaltet, ist er nicht mehr der *eigene*, der unter einem PSM repräsentiert wird, und deshalb kann der Organismus sich seine eigene Lebendigkeit nicht mehr aneignen. Sein Lebensprozess hat nicht nur aufgehört phänomenal sein eigener zu sein, er kann sogar aufgehört haben, überhaupt Teil seiner phänomenalen *Realität* zu sein - je nachdem, wie schwer die psychotische Depression ist. Die Tatsache, dass die autonome Selbstregulierung - ein kontinuierlicher körperlicher Prozess, der auf die biologische Selbsterhaltung abzielt - fort dauert, ist nicht mehr global verfügbar. Tatsachen, die nicht global verfügbar sind, gehören nicht mehr zur Welt eines Menschen, denn die Grenzen des phänomenalen Realitätsmodells sind für jeden von uns die Grenzen der Welt.

Der dynamische Prozess der internen Selbstmodellierung ist ein andauernder Prozess des *Sich-selbst-Beinhaltens*. Ein Cotard-Patient lässt sich daher als ein lebendes, phänomenal selbstrepräsentierendes System beschreiben, das die Tatsache, dass es wirklich ein *lebendes* System ist, nicht mehr phänomenal beinhalten kann.

Die Cotard-Störung besteht jedoch in mehr als einer emotionalen Entkörperung, die entsprechend dem Prinzip der phänomenalen Selbstbezugnahme (welches besagt, dass jegliche kognitive Selbstbezugnahme sich letztlich immer nur Gehalt des aktuell bewussten Selbstmodells richten kann) zu einer persistierenden falschen Überzeugung *de se* führt. Zu behaupten, man sei tot - im Sinne eines *toten Körpers* - ist nicht dasselbe wie zu behaupten, man *existiere* nicht. Wie Enoch und Trethowan schreiben:

Subsequently the subject may proceed to deny her very existence, even dispensing altogether with the use of the personal pronoun 'I'. One patient even called herself 'Madam Zero' in order to emphasize her non-existence. One of Anderson's patients said, referring to herself, 'It's no use. Wrap it up and throw "it" in the dustbin'.¹² (...) If the delusion becomes completely *encapsulated*, the subject may even be able to assume a jovial mood and to engage in a philosophical discussion about her own existence or non-existence. (Enoch und Trethowan 1991, S. 173; 175).

In dieser phänomenalen Zustandsklasse ist die phänomenale Eigenschaft der Selbstheit nicht mehr instantiiert, während ein kohärentes Realitätsmodell als solches immer noch besteht. Ein Cotard-Patient kann jedoch seine dramatische

¹² Enoch and Trethowan beziehen sich bei dieser Behauptung auf Anderson 1968, S. 136. Doch befindet sich das angeführte Beispiel in Andersons Lehrbuch nicht an dieser Stelle.

und generalisierte emotionale Erfahrung der Unvertrautheit und des Valenzverlustes manchmal sogar dadurch zum Ausdruck bringen, dass er die Realität als Ganzes leugnet. Das Phänomen expliziter Existenzleugnung (in Bezug auf den Sprecher selbst) kann nicht ignoriert werden, weil es auf der Ebene explizit verneinten Inhalts das zweithäufigste repräsentationale Merkmal der Cotard-Störung darstellt und in 69 % der Fälle vorliegt (Berrios und Luque 1995, S. 187). Daher lautet die zentrale Frage: Was für eine neurophänomenologische Konfiguration könnte einen Menschen dazu bringen, a) wahrhaftig seine eigene Existenz zu leugnen und b) aufzuhören das Personalpronomen „ich“ (bzw. „ich*“) zu verwenden?

Betrachten wir zunächst die erste Teilfrage. Die phänomenale Repräsentation von Existenz wird beim Menschen über Transparenz kodiert. Phänomenal sind wir Wesen, die den Inhalt gewisser aktiver Repräsentationen genau dann als unhintergebar *real* erfahren, wenn uns frühere Verarbeitungsstufen dieser Repräsentation nicht attentional verfügbar sind. Dies führt zu der Vorhersage, dass eine Person deren Selbstmodell völlig undurchsichtig würde, sich selbst als nicht-existent erfahren würde – genauer: die phänomenale Eigenschaft der Selbstheit wäre nicht mehr instantiiert. Die Subjektkomponente des phänomenalen Modells der Intentionalitätsbeziehung wäre bei einem solchen Wesen vollständig opak und würde nur aus *funktionalen* Gründen weiterhin als der Ursprung seiner Erste-Person-Perspektive fungieren, weil die fortlaufende kontinuierliche Quelle internen, selbsterzeugten Inputs sie

weiterhin zum *funktionalen* Zentrum ihres repräsentationalen Raums macht (Metzinger ²1999, 2003). Da es kein phänomenales *Selbst als Subjekt* mehr gäbe, hätte ein solches System kein PSM im eigentlichen Sinne des Wortes mehr, und sein phänomenales Modell der Intentionalitätsbeziehung würde streng genommen nur eine funktionale, aber keine *phänomenale* Erste-Person-Perspektive mehr generieren. Diese Interpretation ist phänomenologisch unplausibel. Die zweite logische Möglichkeit besteht darin, dass in extremen Cotard-Konfigurationen die eigene Existenz deshalb geleugnet wird, weil es überhaupt kein PSM mehr gibt. Dies allerdings ist empirisch unplausibel, da Cotard-Patienten ein hohes Maß an sensorimotorischer Integration, flexibles Reaktionsvermögen, kohärentes Sprachverhalten, usw. aufweisen – sie befinden sich mit Sicherheit nicht im Koma oder im Tiefschlaf.

Eine dritte Möglichkeit besteht allerdings darin, dass ein teilweise transparentes und bewusstes Selbstmodell vorliegt, dass es aber aus bestimmten Gründen kein vollständiges *Subjektmodell* mehr ist, sondern nur noch ein *Objektmodell*. Etwas existiert noch - etwas, das wie das Modell einer Person aussieht, aber völlig unvertraut und affektiv unbesetzt ist, nicht lebendig und auf jeden Fall kein echtes Selbst im Akt des Lebens, Wahrnehmens, Aufmerksamseins und Denkens ist. In solch einem Fall gäbe es möglicherweise kein internes Modell einer Subjekt-Objekt-Beziehung, sondern nur das einer Objekt-Objekt-Beziehung. Es würde keine phänomenale *Erste-Person-Perspektive* konstituieren, sondern vielmehr eine *Erstes-Objekt-Perspektive*. Das „erste Objekt“ bleibt aus rein

funktionalen Gründen bestehen, als das invariante Zentrum der Realität, einfach weil es funktional in einer hochgradig invarianten Quelle internen Inputs verankert ist. Phänomenal ist dieses funktionale Zentrum der Ort, wo die Dinge geschehen, aber alle diese Dinge - wie auch das Zentrum selbst - werden nicht phänomenal *besessen*. Die phänomenale Eigenschaft der emotional vermittelten „Meinigkeit“, das nicht-begriffliche Gefühl des Besitzens ist aus der Realität des Patienten verschwunden. In den Worten von Philip Gerrans:

In this type of case the patient conceives of herself as nothing more than a locus, not of experience – because, due to the complete suppression of affect, her perceptions and cognitions are not annexed to her body – but of the registration of the passage of events. (Gerrans 2000, S. 118)

Gehen wir jetzt zum zweiten Teil der Frage über: Warum würde ein solches System aufhören, das Pronomen „ich“ zu verwenden, wenn es sich auf sich selbst bezieht? Die Annahme, dass das System immer noch unter einem funktional zentrierten Realitätsmodell operiert, ist plausibel. Motorische Kontrolle, attentionale Verarbeitung und kognitive Verfügbarkeit sind gegeben und zumindest prinzipiell gut integriert. Auch die sensorimotorische Integration ist erfolgreich. Phänomenologisch scheint jedoch ein *subjektloses* Realitätsmodell vorzuliegen. Auf der repräsentationalen Beschreibungsebene finden wir, dass es keine global verfügbare Repräsentation eines *Selbst als Subjekt* gibt. Ich behaupte, dass aus genau diesem Grund der betreffenden Person der Subjektgebrauch von „ich“ im Sinne von Wittgenstein und Shoemaker systematisch unmöglich ist. Vielleicht können wir auf die alte philosophische Frage, worauf sich das

Personalpronomen „bezieht“, für alle Fälle bewusst erlebter Selbstbezugnahme, die der oben diskutierten strukturell ähnlich sind, eine neue Antwort geben.

<Ich*> bezieht sich unvermeidlich auf eine besondere Form phänomenalen mentalen Inhalts: auf die transparente, *subkognitive* Partition des aktuell aktiven PSM, welche den Sprecher *als Subjekt* darstellt. Wenn diese Partition verloren geht oder undurchsichtig wird, dann hört der Sprecher auf <ich*> zu verwenden.

Wie wir gesehen haben, gibt es eine naturalistische Version von Descartes' Argument bezüglich der Gewissheit über die eigene Existenz: Solange ein kohärentes Selbstmodell aktiv und global verfügbar ist, müssen wir annehmen, dass *irgendeine* Art von physischem System existiert, das dieses Modell erzeugt. Aus der Dritte-Person-Perspektive ist dies einfach (und trivialerweise) eine „eingebaute“, aber zentrale theoretische Hintergrundannahme des philosophischen Naturalismus. Interessanterweise gibt es jetzt eine andere Lesart des „Eingebautseins“ – diesmal hinsichtlich der funktionalen Architektur, die der Phänomenologie des Selbstbewusstseins zu Grunde liegt: Aus der Erste-Person-Perspektive ist diese Annahme (bzw. das unhintergehbare *Erleben* der Selbstgewissheit des Subjekts) durch die Transparenz des PSM in unseren Gehirnen fest verankert. Der notwendige „Existenzquantor“ wird sozusagen in einem mentalen, aber nicht-propositionalen Format von den invariantesten und vollständig transparenten Partitionen des Selbstmodells getragen. Er resultiert funktional aus der attentionalen Unverfügbarkeit früherer Verarbeitungsstufen her und spiegelt sich im naiven Realismus des nicht-kognitiven

Selbstbewusstseins wider.

Ich erinnere kurz daran, dass eine der zentralen ontologischen Annahmen der hier vorgelegten Untersuchung besagt, dass es so etwas wie Selbste in der Welt nicht gibt. Alles, was in einem ontologischen Sinne existiert, sind gewisse Klassen informationsverarbeitender Systeme, die unter transparenten Selbstmodellen operieren. Es gibt nur *phänomenale* Selbste, aber keine Selbste. Das gilt auch für den Sonderfall der kognitiven Selbstbezugnahme. Daher sind alle Selbste entweder (phänomenologisch) halluziniert oder Elemente ungenauer, vergegenständlichender phänomenologischer Beschreibungen. Was sich in Standardsituationen nicht transzendieren lässt (eine Ausnahme stellt die oben skizzierte Cotard-Störung dar), ist die subjektive Erfahrung hier und jetzt tatsächlich zu *existieren*. Alle anderen Gehalteigenschaften, die vom aktuell aktiven Selbstmodell erzeugt werden, könnten jedoch grundsätzlich vollständig *phänomenale* Eigenschaften sein, die keinerlei Information über den realen Zustand des Systems tragen.

6. Schlussbemerkung

Eine interessante Konsequenz dieser kurzen repräsentationalistischen Analyse starker Erste-Person-Phänomene im Sinne von Lynne Baker besteht darin, dass die phänomenale Erste-Person-Perspektive eine Möglichkeitsbedingung für das Auftreten einer *kognitiven* Erste-Person-Perspektive ist. Kognitive Selbstbezugnahme ist immer Bezugnahme auf den phänomenalen Inhalt eines

transparenten Selbstmodells. Genauer gesagt: Es ist eine Variante phänomenaler Selbstmodellierung *zweiter Ordnung*, die jedoch durch *ein und dasselbe* integrierte Vehikel der Repräsentation vermittelt wird. Zwei verschiedene Formen mentalen Gehalts werden sozusagen durch denselben Träger realisiert (wobei die begriffliche Unterscheidung zwischen Gehalt und Träger instrumentalistisch verstanden werden sollte). Die Fähigkeit, sich selbst als mental und begrifflich als *sich selbst** zu erfassen, besteht also in der Fähigkeit ein dynamisches, „hybrides“ Selbstmodell zu aktivieren: Phänomenal undurchsichtige, quasi-symbolische Repräsentationen zweiter Ordnung, deren Gehalt durch den Inhalt eines bereits *zuvor* existierenden phänomenal transparenten Selbstmodells gebildet wird, werden aktiviert und dabei kontinuierlich wieder in eben dieses Selbstmodell eingebettet. Dieser Vorgang ist dann der Vorgang der Introspektion⁴, so wie er in **DEF 6** eingeführt wurde.

Was lässt sich jetzt über den Begriff eines bewussten kognitiven Subjekts sagen? Das transparente subsymbolische Selbstmodell ist ein Werkzeug, das von einem informationsverarbeitenden System verwendet wird, um eine bestimmte Art von Information *kognitiv verfügbar* zu machen. Das tut es, indem es erstmals undurchsichtige mentale Selbstsimulationen ermöglicht. Solche Selbstsimulationen können auf der phänomenalen Ebene ihren Ausdruck als bewusst erlebte Ich*-Gedanken finden, oder linguistisch als Ich*-Sätze, d.h. sie können zur kognitiven oder linguistischen Selbstzuschreibung einer präreflexiven Erste-Person-Perspektive führen. Durch den Akt der Selbstzuschreibung wird die

phänomenale zur kognitiven Erste-Person-Perspektive. Kognitive Selbstbezugnahme ist ein Prozess, in dem gewisse Aspekte des Inhalts eines zuvor existierenden transparenten Selbstmodells phänomenal modelliert werden und der dann begrifflich und aus der Außenperspektive als die Fähigkeit, sich selbst als sich selbst* zu erfassen, interpretiert werden kann.

Baker hat vollkommen Recht, wenn sie darauf hinweist, dass die Postulierung eines mentalen Symbols, das funktional intern noch einmal die Rolle von <Ich*> spielt, das Problem nur neu benennt, ohne es zu lösen. Wonach sie in Wirklichkeit sucht, ist der Besitz einer besonderen Fähigkeit, der Fähigkeit sich selbst geistig als sich selbst* zu erfassen, einer Fähigkeit, die über den bloßen Besitz einer selbstlokalisierenden Überzeugung à la Perry (1979) hinausgeht. Die Fähigkeit, die sie sucht, ist die Fähigkeit unter einem phänomenal transparenten Selbstmodell zu operieren, das *global verfügbar* ist, und zwar nicht nur für die selbstgerichtete Aufmerksamkeit und die selektive Handlungskontrolle, sondern gleichzeitig auch für die mentale Begriffsbildung. Es ist die Fähigkeit Introspektion₄ auf ein bereits existierendes, im Gehirn aktives Selbstmodell anzuwenden. Mit dem Menschen hat die natürliche Evolution auf unserem Planeten dieses Vermögen schon hervorgebracht - wie man an den Schriften von René Descartes sehen kann. Trotzdem stellen kognitives Handeln, begrifflich vermitteltes Selbstbewusstsein und starke Erste-Person-Phänomene im Sinne von Lynne Baker kein unüberwindliches Hindernis für den Naturalismus dar. Es ist möglich, diese Phänomene im Rahmen einer empirisch plausiblen Theorie der

mentalen Repräsentation nicht nur immer genauer, sondern auch erschöpfend zu analysieren. Das eigentliche philosophische Problem scheint mir eher darin zu bestehen, dass der zumeist stillschweigend vorausgesetzte *epistemische* Gehalt der kognitiven Selbstbezugnahme jetzt prinzipiell immer auf eine rein phänomenale Art mentalen Gehalts reduziert werden kann. Kognitives Selbstbewusstsein und die hier diskutierte Form der Selbstreferenz brauchen eine unabhängige epistemische Rechtfertigung.

Baker hat verlangt, dass jeder robuste, reduktive Naturalismus, der standhalten soll, eine überzeugende Analyse für die konstitutiven Bedingungen der oben erwähnten Fähigkeit in nicht-intentionalen und nicht-semantischen Begriffe vorlegen muss. Ich habe daher eine Analyse in *phänomenalen* Begriffen vorgelegt. Phänomenaler Gehalt ist nicht-intentionaler Gehalt und als solcher offen dafür, naturalisiert zu werden. Wenn es zutrifft, dass phänomenaler Inhalt lokal superveniert, dann kann man plausiblerweise annehmen, dass zu jeder phänomenalen Eigenschaft eine nomologisch koextensive physikalische Eigenschaft im menschlichen Gehirn existiert. Empirische Forschungsprogramme könnten dann *gegenstandsbereichsspezifische Identitäten* für jede einzelne Form von phänomenalem Inhalt entdecken. Das wäre dann auch für den phänomenalen Gehalt der kognitiven Selbstbezugnahme möglich.

Es gibt jedoch ein *Caveat*. Wenn phänomenaler Inhalt - wie heute viele Philosophen glauben – letztlich doch einfach eine besondere Variante von intentionalem Inhalt ist, dann wird das hier skizzierte Modell am Ende durch eine

vollständige naturalistische Theorie der Intentionalität ergänzt werden müssen, durch eine Theorie, die uns die *Beziehung* zwischen phänomenalem und intentionalem Gehalt begrifflich erläutert. Davon sind wir weit entfernt. Heute kann man, wie es aussieht, nur zwei Dinge mit Sicherheit sagen: Phänomenaler Gehalt als solcher ist kein epistemisch gerechtfertigter Inhalt, und er superveniert beim Menschen lokal auf Gehirneigenschaften. Ein isoliertes Gehirn im Tank könnte deshalb auch das spezifische bewusste *Erleben* generieren, das mit dem cartesianischen Gedanken [Ich bin mir gewiss, dass ich* existiere] einhergeht. Cotard-Patienten können bereits heute das spezifische bewusste *Erleben* generieren, das mit dem „nihilistischen“ Gedanken [Ich bin mir gewiss, dass ich* nicht existiere] einhergeht. Es ist daher plausibel, anzunehmen, dass ein minimal hinreichendes neuronales Korrelat für den phänomenalen Gehalt existiert, der diese spezifischen mentalen Zustände jeweils charakterisiert. Der phänomenale Gehalt ist in jedem einzelnen Fall durch dieses physische Korrelat vollständig determiniert.

Gewissheitserleben ist nicht dasselbe wie Gewissheit. Es wird viele Fälle geben, in denen der phänomenale Gehalt der kognitiven Selbstbezugnahme zuverlässig mit *intentionalem*, epistemisch gerechtfertigtem Inhalt korreliert ist. Unter naturalistischen Hintergrundannahmen muss trivialerweise *irgendein* repräsentationales System existieren, weil transparente Selbstmodelle begrifflich als durch physische Vehikel realisierte Strukturen eingeführt wurden. Wenn diese Hintergrundannahme eine unabhängige epistemische Rechtfertigung besitzt,

dann ist klar, dass jedes bewusste System in dem Ich*-Gedanken gerechtfertigt ist, dass *irgendein* physisches Repräsentationssystem existiert, welches jetzt die kognitive Selbstbezugnahme durchführt. Ungerechtfertigt ist dagegen die These, dass dieses System ein *Selbst* ist.

7. Literatur

- Anderson, E.W. (1964). *Psychiatry*. London: Baillière, Tindall and Cox Ltd.
- Anton, G. (1898). Ueber die Herderkrankungen des Gehirns, welche vom Patienten selbst nicht wahrgenommen werden. *Wiener klinische Wochenschrift* 11 (March 10, 1898): 227-9.
- Anton, G. (1899). Über die Selbstwahrnehmungen der Herderkrankungen des Gehirns durch den Kranken bei Rindenblindheit und Rindentaubheit. *Archiv für Psychiatrie und Nervenkrankheiten* 32: 86-127.
- Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baker, L.R. (1998). The first-person perspective: A test for naturalism. *American Philosophical Quarterly*, 35, 327-46. **Gekürzte deutsche Übersetzung:** Baker, L.R. (2000). Die Perspektive der ersten Person: Ein Test für den Naturalismus. In G. Keil und H. Schnädelbach (Hrsg.), *Naturalismus - Philosophische Beiträge*. Frankfurt am Main: Suhrkamp.
- Benson, D.F. und Greenberg, J.P. (1969). Visual form agnosia. *Archives of Neurology* 20: 82-89.
- Berrios, G.A. und Luque, R. (1995). Cotard's syndrome. Analysis of 100 cases. *Acta Psychiatrica Scandinavica* 91: 185-9.
- Breen, N., Caine, D., Coltheart, M., Hendy, J. und Roberts, C. (2000) Towards an understanding of delusions of misidentification: Four case studies. In M. Coltheart und M. Davies (eds), *Pathologies of Belief*. Oxford: Blackwell.
- Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., und Wiener, M.C. (2001). Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society London B* 356: 299-330.
- Campbell, J. (1999). Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist* 82: 609-25.
- Castañeda, H.N. (1966). »He«: A study on the logic of self-consciousness. *Ratio* 8: 130-157.
- Chalmers, D.J. (1997). Availability: The cognitive basis of experience? In N. Block, O. Flanagan und G. Güzeldere (eds), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- Chalmers, D.J. (2000). What is a neural correlate of consciousness? In T. Metzinger (ed), *Neural Correlates of Consciousness – Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Churchland, P.M. (1998). Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered. *Journal of Philosophy* 65: 5-32.
- Churchland, P.M. (unpublished manuscript). Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds.
- Clark, A. (1997). *Being There – Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.
- Coltheart, M. und Davies, M. (2000)[eds]. *Pathologies of Belief*. Oxford und Malden, MA: Blackwell.
- Cotard, J. (1880). Du délire hypocondriaque dans une form grave de la mélancolie anxieuse. *Annales Médico-Psychologiques* 38: 168-170.
- Cotard, J. (1882). Du délire des négations. *Archives de Neurologie* 4: 152-170/282-295.

- Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace und Company. Deutsche Übersetzung (2000): *Ich fühle, also bin ich*. München und Leipzig: List.
- Davies, M., Coltheart, M., Langdon, R., und Breen, N. (2001). Monothematic delusions: Towards a two-factor account. In C. Hoerl (Hrsg.), *On Understanding and explaining Schizophrenia*. Sonderband Nr. 8 von Philosophy, Psychiatry and Psychology.
- Dretske, F. (1969). *Seeing and Knowing*. Chicago: University of Chicago Press.
- Eimer, M. (1996). Wahrnehmung und Aufmerksamkeit. In G. Roth und W. Prinz (Hrsg.), *Kopf-Arbeit – Gehirnfunktionen und kognitive Leistungen*. Heidelberg, Berlin, Oxford: Spektrum Verlag.
- Enoch, M.D. und Trethowan, W.H. (1991). *Uncommon Psychiatric Syndromes*. Oxford: Butterworth-Heinemann.
- Förstl, H., Almeida, O.P., Owen, A., Burns, A. und Howard, R. (1991). Psychiatric, neurological and medical aspects of misidentifications syndromes: a review of 260 cases. *Psychological Medicine* 21: 905-910.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* LXVIII: 5-20.
- Gazzaniga, M. (1995)[ed]. *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.
- Gerrans, P. (1999). Delusional misidentification as subpersonal disintegration. *The Monist* 82: 590-608.
- Gerrans, P. (2000). Refining the explanation of Cotard's delusion. In M. Coltheart und M. Davies (eds), *Pathologies of Belief*. Oxford und Malden, MA: Blackwell.
- Güzeldere, G. (1995). Is consciousness the perception of what passes on one's own mind? In T. Metzinger (ed), *Conscious Experience*. Thorverton, UK: Imprint Academic & Paderborn: mentis.
- Halligan, P.W. und Marshall, J.C. (1996) [eds]. *Method in Madness: Case studies in Cognitive Neuropsychiatry*. Hove, UK: Psychology Press.
- Harman, G. (1990). The intrinsic quality of experience. In J. Tomberlin (ed), *Philosophical Perspectives, Vol. 4: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview Publishing. Reprinted in and quoted after N. Block, O. Flanagan und G. Güzeldere (eds), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press
- Helm, G. (1991). *Symbolische und konnektionistische Modelle der menschlichen Informationsverarbeitung*. Informatik-Fachberichte 288. Berlin: Springer.
- LaBerge, S., und Gackenbach, J. (2000). Lucid dreaming. In E. Cardeña, S.J. Lynne und S. Krippner (eds), *Varieties of anomalous experience: Examining the scientific evidence*. Washington, D.C. American Psychological Association.
- Leopold, D.A., und Logothetis, N.K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences* 3: 254-64.
- Metzinger, T. und Walde, B. (2000). Commentary on Jakob's „Ineffability of Qualia“. *Consciousness and Cognition* 9: 352-62. doi:10.1006/ccog.2000.0463
- Metzinger, T. (2000). The subjectivity of subjective experience: A representationalist analysis of the first- perspective. In T. Metzinger (ed), *Neural Correlates of Consciousness – Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Metzinger, T. (2003). *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Metzinger, T. (1999). *Subjekt und Selbstmodell*. Paderborn: mentis.
- Moore, G.E. (1903). The refutation of idealism. *Mind* 12: 433-53.
- Newen, A. (1997). The logic of indexical thoughts and the metaphysics of the 'self'. In W. Kühne, A. Newen und M. Anduschus, (eds), *Direct Reference, Indexicality and Propositional Attitudes*. Stanford: CSLI.
- O'Brien, G. und Opie, J. (2001). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines und P. Slezak (eds), *Representation in Mind: New Approaches to Mental Representation*. Greenwood Publishers.

- Perry, J. (1979). The problem of the essential indexical. *Noûs* 13: 3-22.
- Posner, M.I. (1995). Attention in cognitive neuroscience: An overview. In Gazzaniga 1995.
- Raffman, D. (1995). On the persistence of phenomenology. In T. Metzinger (ed), *Conscious Experience*. Thorverton, UK: Imprint Academic und Paderborn: mentis.
- Ramsey, W., Stich, S., und Garon, J. (1991). Connectionism, eliminativism, and the future of folk psychology. In W. Ramsey, S. Stich und D. E. Rumelhart, eds., *Philosophy and Connectionist Theory*. Hillsdale, NY: Erlbaum.
- Séglas, J. (1897). *Le Délire des Négations: Séméiologie et Diagnostic*. Paris: Masson, Gauthier-Villars.
- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy* 65: 555-78.
- Shoemaker, S. (1990). Qualities and qualia: What's in the mind? *Philosophy and Phenomenological Research*, 50, Supplement, 109-31. Reprinted in S. Shoemaker (1996), *The First-Person Perspective and Other Essays*. Cambridge, UK: Cambridge University Press.
- Siegel, R.K., und West, L.J., eds. (1975). *Hallucinations. Behavior, Experience, and Theory*. New York: Wiley.
- Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press.
- Tye, M. (1995). *Ten Problems of Consciousness – A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Tye, M. (1998). Inverted Earth, Swampman, and representationism. In J. Tomberlin (ed), *Philosophical Perspectives, Vol. 12: Language, Mind, and Ontology*. Malden, MA / Oxford, UK: Blackwell.
- Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
- Van Gulick, R. (1988). A functionalist plea for self-consciousness. *The Philosophical Review* XCVII (2): 149-81.
- Wittgenstein, L. (1953; 1982). *Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp.
- Young, A.W. und Leafhead, K.M. (1996). Betwixt of life and death: Case studies of the Cotard delusion. In P.W. Halligan und J.C. Marshall (eds), *Method in Madness: Case studies in cognitive neuropsychiatry*. Hove, UK: Psychology Press.
- Young, A.W. (1999). Delusions. *The Monist* 82: 571-90.