

## 6. (tatsächlich: 4.) Vorlesung vom 15. Mai 2007

Ich möchte Ihnen heute am Beispiel eines bekannten, jedenfalls viel diskutierten Aufsatzes von David Rosenthal 1. eine interessante Überzeugung über und 2. ein falsches Erklärungsmodell für Bewusstsein vorstellen („Two Concepts of Consciousness“, in: D. Rosenthal [ed.], *The Nature of Mind*, New York-Oxford: University Press, 1991, 452-477, zuerst 1986 in *Philosophical Studies* 94, 3 veröffentlicht, 329-59).

Die Überzeugung ist: Bewusstsein lässt sich nicht als einseitig auffassen. Es liegt immer nur vor, wenn auch noch bewusst ist, *dass* es vorliegt. Diese Auffassung hatten wir auch schon bei Husserl angetroffen („innere Wahrnehmung“) und ansatzweise auch bei Ned Block. *P*-Bewusstsein – denken Sie an das unaufmerksame Nebenbei-Hören der Dampfmaschine, die vor Ihrem Arbeitszimmer dröhnt – tritt nur dann auf, wenn sein Auftreten selbst noch einmal bewusst ist. Diese Intuition liefert auch das Motiv dafür, dass man Bewusstsein mit einer Form von Selbstbewusstsein gleichsetzt, nämlich dem Bewusstsein davon, dass ein bestimmter mentaler Zustand vorliegt. – Nun nimmt Rosenthal gerade dies nicht an. Seiner Meinung nach gehört es *nicht* zu den notwendigen und zureichenden Eigenschaften unseres mentalen Leben, dass wir uns ihrer bewusst sein müssen. Wir müssen uns ihrer nur bewusst werden *können* (463). (So meinte auch Kant, dass der Gedanke ‚Ich denke‘ nicht alle unsere Vorstellungen notwendig begleitet, sondern dass er das nur können muss: *KrV* B 131 f.)

Aber Rosenthals Aufsatz ist aus einem zweiten Grunde interessant. Er liefert ein geradezu klassisches Beispiel für das so genannte ‚Reflexionsmodell‘ des Selbstbewusstseins. Danach wird ein mentaler Zustand erster Ordnung erst dadurch *bewusst*,

dass ein zweiter (höherstufiger) sich auf ihn beugt und dabei feststellt: Dieser Zustand liegt tatsächlich vor. Der erste (niederstufige) Zustand wäre dann der Wahrmacher des zweiten, höherstufigen. – Ich werde nun im Folgenden Rosenthal nicht in seiner ersten Überzeugung widersprechen, wonach Bewusstsein typischerweise seiner selbst bewusst ist oder es doch werden können muss. Diese Überzeugung halte ich vielmehr für richtig. Ich werde Rosenthal aber darin widersprechen, dass Selbstbewusstsein als eine Selbstreflexion, als ein nachträglicher Akt höherer Ordnung interpretiert werden kann.

Aber beginnen wir ohne Überstürzung. Rosenthal ist mir schon darum sympathisch, weil er sich gleich eingangs ein eindrucksvolles Bekenntnis zur Zentralität von Bewusstsein für jede denkbare Philosophie des Geistes abgibt:

No mental phenomenon is more central than consciousness to an adequate understanding of the mind. Nor does any mental phenomenon seem more stubbornly to resist theoretical treatment.

Consciousness is so basic to the way we think about the mind that it can be tempting to suppose that no mental states exist that are not conscious states. Indeed, it may even seem mysterious what sort of thing a mental state might be if it is not a conscious state. On this way of looking at things, if any mental states do lack consciousness, they are exceptional cases that all call for special explanation or qualification. Perhaps dispositional or cognitive states exist that are not conscious, but nonetheless count as mental states. But if so, such states would be derivatively mental, owing their mental status solely to their connection with mental states. And perhaps it makes sense to postulate nonconscious versions of ordinary mental states, as some psychological theories do. But if consciousness is central to mentality in the way this picture insists, any such states are at best degenerate examples of mentality, and thus peripheral to our concept of mind (462 l. 1).

Aber ist – so fragt sich auf einen zweiten, kritischeren Blick – dies Bild wirklich angemessen? Wie bewusst sind uns denn eigentlich mentale Zustände? Die Frage wird noch aufsässiger, wenn wir bedenken, dass mentale Zustände zwar keine physischen Eigenschaften aufweisen (sie sind z. B. nicht räumlich; Blau-Wahrnehmungen sind nicht selbst blau, usw.). Aber gerade visuelle Wahrnehmungen sind so offen-

kundig an physisch-chemische Eigenschaften des Sinnesorgans Auge geknüpft, dass wir genauer über den Zusammenhang des Bewussten mit dem Nicht-Bewussten nachdenken müssen. Aber Sinneswahrnehmungen sind nicht die einzigen mentalen Phänomene, bei denen wir enge Verflechtungen mit unserem Körper erleben. Für einige Gefühle oder Emotionen gilt das erst recht (und schon Descartes hat das in seinen *Passions de l'âme* so gesehen). Wenn wir vor Furcht zittern, gelb vor Neid werden oder uns vor Ekel erbrechen müssen, dann erleben wir einen sehr eindrucksvoll engen Zusammenhang einiger mentaler Zustände mit der Beschaffenheit der Welt und unseres Körpers. – Damit widerspricht Rosenthal der häufig anzutreffenden cartesischen Überzeugung, es gebe keine kontinuierlichen Übergang vom Physischen zum Mentalen:

Mind and consciousness are continuous with other natural phenomena of which we can give impressively powerful explanations (463).

Rosenthals 1. These ist also: Nicht alle Zustände, die in unserem Geist vorkommen, sind *ipso facto* bewusst. Wenn wir – wie viele Philosophen annehmen (Rosenthal nennt Descartes und Th.. Nagel – annehmen, Bewusstsein sei eine notwendige und zureichende, also eine erschöpfende Bedingung für den Geist („consciousness [is] what makes a state a mental state“ [464], geraten wir in zirkuläre Verwicklungen. Wir müssen dann über Bewusstsein immer schon verständigt sein, wenn wir von Mentalem sprechen. Und – Sie erinnern sich – die Voraussetzung machen auch viele Philosophers of Mind. Sie erinnern sich: Block beginnt seinen Text „On a Confusion [...]“ mit dem Zugeständnis, dass er keine nicht-zirkuläre Definition von Bewusstsein geben könne.

Rosenthal ist aber höflich genug, mit unseren cartesianischen Alltags-Intuitionen nicht einfach zu brechen. Er will uns vielmehr über den Mechanismus aufklären, wie wir nicht-zirkulär und in informativer Weise von mentalen zu bewussten mentalen Zuständen übergehen können. [Wiederholt erklären, was ein *circulus in probando* und was ein *infiniter Regress* ist!] Wir werden sehen, dass Rosenthals Erklärungsversuch selbst ein besonders hilfloses Beispiel für Zirkularität liefert. Aus Rosenthals Scheitern lässt sich aber viel lernen über eine angemessene Theorie des Bewusstseins.

Wir stimmen also aus Höflichkeit Rosenthals Trennung des Mentalen vom Bewussten erst einmal zu und erwarten nun von *ihm* eine Auskunft, was einen mentalen Zustand zu einem bewussten mache (wenn er es nicht immer schon ist). Und hier ist seine Auskunft schlicht und direkt: „On [my] explanation, a mental state is conscious if one has a suitable second-order thought“ (465 r.). Wir hätten dann einen mentalen (selbst nicht bewussten) Zustand erster Ordnung; und über diesen erststufigen mentalen Zustand hätten wir einen Gedanken höherer Stufe. Der Gehalt dieses höherstufigen Gedankens wird ungefähr lauten: Da liegt ein erststufiger (*vor* meinem Gedanken unbewusster) mentaler Zustand vor. (Es ist übrigens gar nicht zu sehen, warum ein unbewusster Zustand durch Reflexion – d. h. durch Abspiegelung – bewusst werden sollte. Husserl hatte gesagt: Die Reflexion kann nur finden, was schon da war; sonst haben wir nicht mit einer Reflexion, sondern mit einer Art Gehirnwäsche zu tun, die das ins Bewusstsein zu Hebende vielmehr verändert als abspiegelt: Ich spare mir diesen wichtigen Einwand für eine bessere Gelegenheit auf.)

Rosenthal fügt hinzu, dieser höherstufige Akt sollte möglichst ungefähr gleichzeitig („roughly contemporaneous“) mit dem ersten erfolgen – vielleicht, weil er denkt, mit größerer Zeitentfernung könne der erststufige Zustand in Vergessenheit geraten, oder – da wir sehr viele Akte zugleich nicht zum Bewusstsein heben können – eine Fehlindividuierung erfolgen: „we cannot be conscious of every thing at once“ (466, r. 1). Das war, nach Auskunft seines Interviewers (Burman), auch Descartes' Ansicht. Er meinte aber, zwei übereinandergelagerte mentale Akte könne das Bewusstsein verkraften; ebenso, wie es gleichzeitig repräsentieren könne, dass es rede und esse. Bewusstsein bestehe geradezu in solcher Gleichzeitigkeit des Vorstellens („penser“) und des Sich-seines-Vorstellens-reflexiv-bewusst-Seins (1358 f).

Descartes vertritt, obwohl Rosenthal ihn was ganz anderes meinen lässt, also *grosso modo* Rosenthals eigenen Standpunkt. Von *bewusst sein* sagt Descartes (auf Französisch): „Être conscient, c'est assurément penser et réfléchir sur sa pensée“ (Descartes 1953, 1359). Also: Bewusst sein heißt: denken und auf sein Denken reflektieren. Auch Rosenthal nennt ein Denken über einen mentalen Zustand ‚reflektieren‘ (Rosenthal 1991, 466 l., 2. Abschn.). Nun muss man wissen, dass Descartes unter ‚cogitare‘ jeden mentalen Zustand versteht, sofern er auf diese Weise zum Bewusstsein gehoben werden kann: Sinneswahrnehmungen, Gefühle, Wollungen, natürlich auch Gedanken im engeren Sinne: genau dasselbe, was Kant pauschal ‚Vorstellungen‘ und was Rosenthal *mental states* nennt. Descartes fügt hinzu, er nenne – innerhalb der Menge der *cogitationes* – ‚Idee‘

tout ce qui est immédiatement conçu par l'esprit: en sorte que, lorsque je veux et que je crains, parce que je conçois en même temps que je veux et que je crains, ce vouloir et cette crainte sont mis par moi au nombre des idées (vgl. 390 und 286,2).

Sie haben bemerkt: Wie Rosenthal besteht Descartes auf der Gleichzeitigkeit des erststufigen Zustands und des zweitstufigen Gedankens. Das hatte auch Spinoza – in Descartes' Nachfolge – angenommen (zit. in Frank 2002, 186-88). Hobbes hatte eingewandt, dass, wenn Bewusstsein *gleichzeitig* bestehen soll zu dem mentalen Akt, *dessen* reflexives Bewusstsein es ist, ein unendlicher Regress sich bilden würde:

[...] il est tout à fait impossible de penser qu'on pense, ni de savoir qu'on sait; car ce serait une interrogation qui ne finirait jamais: d'où savez-vous que vous savez que vous savez ce que vous savez, etc.? (Hobbes in: Descartes 1953, 401).

Hobbes sieht den Regress, wie gesagt, nur ausbrechen, wenn die Reflexion als zeitgleich mit dem Reflektierten auftretend gedacht werde; als Erinnerung an vergangene Geisteszustände lässt er sie zu. Dabei sieht er nicht, dass ich mich nur an das erinnern könnte, was einmal gegenwärtig bewusst war, und zwar so, dass das Bewusstsein mit dem bewussten Erlebnis zeitgleich bestanden hat (so lautet denn auch etwas Descartes' Erwiderung auf Hobbes' 5. Einwand [Descartes 1953, 407]). Das sah Spinoza ebenso (Spinoza 1925 b, *Ethica* II, Prop. 21, Schol.).

Aber Descartes' und Spinozas Auskunft ist natürlich nicht überzeugend; denn der Regress tritt schon darum in einer vitiösen Gestalt auf, wenn und weil er den mentalen Zustand ("idea") und das Bewusstsein von ihm ("idea ideae") als zweierlei zählt. Zeitgleichheit oder -ungleichheit sind dagegen belanglos. Zwei gleichzeitig auftretende Zustände bestehen noch lange nicht in der Einheit eines Ereignisses. Diesen – bei Spinoza verdeckten – Fehlschluss bringt die These einer möglichen Ungleichzeitigkeit von Ereignis und Ereignis-Bewusstsein unfreiwillig ans volle Licht.

Selbst die Gefahr eines (mit dem Reflexionsmodell von Selbstbewusstsein einhergehenden) Regresses hat Descartes in den „Réponses de l'auteur aux sixièmes objections“ hellichtig erkannt, wenn auch nicht gelöst.

*C'est une chose très assurée que personne ne peut être certain s'il pense et s'il existe, si, premièrement il ne connaît la nature de la pensée et de l'existence. Non que pour cela il soit besoin d'une science réfléchie, ou acquise par une démonstration, et beaucoup moins de / la science de cette science, par laquelle il connaisse qu'il sait, et derechef qu'il sait qu'il sait, et ainsi jusqu'à l'infini, étant impossible qu'on en puisse jamais avoir une telle d'aucune chose que ce soit; mais il suffit qu'il sache cela par cette sorte de connaissance intérieure qui précède toujours l'acquise [...] (526 f.)*

Aber jetzt sind wir von Rosenthals Erklärung abgeschweift, wie denn mentale Akte im Allgemeinen zu bewussten im Besonderen werden können. Ich habe nur betonen wollen, dass seine Charakterisierung über die Position des Descartes ganz oberflächlich (und aus Textkenntnis schlecht) belehrt ist. In Wahrheit hat er eine ähnliche, vielleicht noch etwas naivere Vorstellung vom Aufkommen von Selbstbewusstsein als Descartes.

Hören wir ihn darum genauer an. Nach der Unterscheidung der beiden Auffassungen des Mentalen (nach der ja sein ganzer Aufsatz heißt) schreibt Rosenthal:

*Although both marks of the mental have enjoyed widespread acceptance, it is crucial which mark we adopt if our goal is to give an explanation of consciousness. Conscious states are simply mental states we are conscious of being in. And, in general, our being conscious of something is just a matter of our having a thought of some sort about it. [Das war Descartes' Definition.] Accordingly, it is natural to identify a mental state's being conscious with one's having a roughly contemporaneous thought that one is in that mental state. When a mental state is conscious, one's awareness of it is, intuitively, immediate in some way. So we can stipulate that the contemporaneous thought one has is not mediated by any inference or perceptual input. We are then in a position to advance a useful, informative explanation of what makes conscious states conscious. Since a mental state is conscious if it is accompanied by a suitable higher-order thought, we can explain a mental state's being conscious by hypothesizing that the mental state itself causes that higher order thought to occur (Rosenthal 1991, 465, r., 2).*

Achten Sie auf eine paar charakteristische begriffliche Unsauberkeiten: ‚roughly contemporaneous‘, ‚suitable second-order thought‘; die unkommentierte Einführung des Alternativ-Ausdrucks für ‚consciousness‘: ‚awareness‘, das Nebeneinanderstehen der Begriffe ‚unmittelbar‘ bzw. ‚nicht-inferentiell‘ neben ‚begleitet von‘ (später noch die unklare Abgrenzung der ‚Introspektion‘ gegen ‚höherstufige Gedanken‘). Über die ‚ungefähre Gleichzeitigkeit‘ der höherstufigen Begleitgedankens haben wir uns schon mokiert. Wenn aber ein Zustand einen anderen Zustand ‚begleitet‘ und dieser auch noch als Zustand *zweiter* Stufe gezählt wird, dann haben wir mit einer Relation mit zwei Gliedern zu tun – und die darf nicht ‚unmittelbar‘ heißen. Das sagt Rosenthal zum Überfluss selbst: „[C]onsciousness must be a relational property of being accompanied by higher-order thoughts“ (474 r. u.; die entgegengesetzte Auffassung, Bewusstseinszustände seien unrelational und darum nicht weiter analysierbar, verurteilt er als ‚cartesianisch‘: 468 l. 3). Rosenthal nimmt aber auch an der Konsequenz keinen Anstoß, dass auf diesem Wege gar kein Ende der Reflexionskette abzusehen ist: Der zweite Akt ist ja selbst unbewusst und wird bewusst erst durch einen Akt dritter Ordnung (465, r. u.). Solche Akte, tröstet uns Rosenthal, kämen aber relativ selten vor („we would expect, [...], that the third-order thoughts that confer consciousness on such second-order thoughts would be relatively rare“). Aber wenn der Gedanke dritter Stufe seinerseits wieder unbewusst ist, bedarf er zur Bewusstwerdung eines Gedankens vierter Stufe – und so bliebe der jeweils letzte Gedanke selbst unbewusst – die ganze Kette endete im Unbewussten.

Rosenthal gibt das zu: „But higher-order thoughts are not automatically conscious, any more than other mental states“ (466, l. u.). Und was das Regress-Argument mit

dem Vorwurf betrifft, die ganze Kette laufe ins Unbewusste aus, sagt Rosenthal treuherzig:

It may seem slightly odd that each of these hierarchies of conscious mental states has a nonconscious thought at its top. But whatever air of paradox there seems to be here is dispelled by the common-sense truism that we cannot be conscious of everything at once (466, r. o.).<sup>1</sup>

Nun ist es sicher wahr, dass wir nicht alles auf einmal wissen können. Aber ein in-finiten Regress, der ‚an seiner Spitze‘ einen unbewussten Gedanken hat, ist natürlich als Explanans für Bewusstsein unduldbar. Einen solchen Regress nennt man in der Philosophie vitiös. Schelling verglich eine ähnliche Bewusstseins-Theorie wie die Rosenthal’sche einmal mit der englischen Staatsschuld: Einer borgt bei einem anderem, der selbst nichts hat und bei einem dritten borgt, usw. (SW I/4, 344 Anm.). Irgendwann muss der Check einmal eingelöst werden, sonst wird die Schuld eben nicht abgetragen, sondern auf den Sankt-Nimmerleins-Tag gefristet. Entsprechend muss irgendwann einmal ein Zustand auftreten, der *unmittelbar seiner selbst bewusst* ist, sonst kommt Bewusstsein in der ganzen Kette nirgendwo auf, und der von Rosenthal als ‚schlicht‘ und ‚simpel‘ charakterisierte Erklärungsversuch bricht in sich zusammen.

Aus diesem Zusammenbruch können wir aber eine wichtige Lehre darüber ziehen, wie eine Bewusstseinstheorie *nicht* aussehen darf. Eine erste Einsicht könnte

---

<sup>1</sup> Rosenthal unterscheidet übrigens höherstufige Gedanken von Introspektionen (Rosenthal 1991, 465 f.; 474). Bei introspektivem Wissen sei Aufmerksamkeit im Spiel. Dennoch dürfe diese nicht mit dem Vorliegen eines höherstufigen Gedankens verwechselt werden, der ja selbst unbewusst bleibe. „Introspection is the reflective consciousness of our mental states“ (474). Genauer: Introspektionen reflektieren höherstufige Gedanken, sind als drittstufig („a yet higher-order thought“ [466 r. u.]). Wie kann man sich Aufmerksamkeit als etwas nicht Kognitives vorstellen? Ferner: Introspektion verhilft uns nicht nur zum Bewusstsein, sondern auch zum Bewusstsein des jeweils höchsten Bewusstseins. Wie wäre das möglich, wenn entweder das jeweils letzte Bewusstsein unbewusst bleibt oder aber seiner bewusst ist? Ist es seiner bewusst, dann gibt es eben doch ein unmittelbares, nicht-reflexives Bewusstsein von Bewusstsein, und die common-sense-Ansicht von Bewusstsein, die Rosenthal bekämpft, hätte sich als die richtige herausgestellt. – Ich finde das alles reichlich dunkel und *ad hoc*.

so lauten, dass wir beim Selbstbewusstsein mit einem von Gegenstandsbewusstsein spezifisch verschiedenen Bewusstsein zu tun haben. Denn *second-order*-Gedanken haben mentale Zustände zu ihrem *Gegenstand*, sie richten sich intentional auf sie. Das war, wie Dieter Henrich 1966 gezeigt hat (Henrich 1967), eine fundamentale Einsicht Johann Gottlieb Fichtes, zuerst formuliert 1797 im „Versuch einer neuen Darstellung der Wissenschaftslehre“. Sie bestand in der Entdeckung, dass das *so genannte Reflexionsmodell* an der Beschaffenheit von Selbstbewusstsein scheitert. Dies Modell besagt, Selbstbewusstsein komme durch eine Art höherstufigen Bewusstseins zustande, das auf ein Bewusstsein ersten Grades gerichtet werde. Auf die Weise entdeckte das Bewusstsein sich selbst in der Position eines Gegenstandes. Und da alles Wissen im Wissen von einem Gegenstand bestehe (dieser sei ein Einzelgegenstand oder ein Universale oder ein Sachverhalt), sei damit das Phänomen des Selbstwissens als das Wissen von einem besonderen Typ von Gegenstand erklärt.

Der Zirkel in diesem Erklärungsversuch springt in die Augen; doch blieb er, soviel wir wissen, in den klassischen Erklärungsversuchen von Selbstbewusstsein, die vor Fichte unternommen worden sind, unbemerkt.<sup>2</sup> Nun *besteht* Selbstbewusstsein, sagt Fichte. Bestehende Sachverhalte können nicht falsch sein. Das kann nur ihre (in Zirkel oder Regresse führende) Erklärung. In der Tat besteht die Erklärung von Selbstbewusstsein aus einer Reflexion die folgende Überprüfung nicht.

---

<sup>2</sup> Vielleicht – wie Udo Thiel (1996) gezeigt hat – mit der Ausnahme von Johann Bernhard Merian (1749a und 1749b). Ausführlich dazu: Frank 2002, 154 ff.

Vergegenwärtigen wir uns zunächst so genau wie möglich, was die falsche Reflexionstheorie behauptet: Ich gewinne – sagt sie – Kenntnis vom Ich (sagt Fichte; vom Bewusstsein, sagt Rosenthal) durch Reflexion, d. h. dadurch, dass das Ich (bzw. der mentale Zustand) in einen Bezug zu sich selbst eintritt (oder sich auf sich selbst zurück wendet), wodurch es seiner ansichtig oder gewahr („aware“) wird. Aber wie kann man das Subjekt, inwiefern es nichts ist als reines Subjekt, erkennen? Kant hat zwischen der reinen Apperzeption und dem Gedanken ‚Ich denke‘ wohl unterschieden und sogar gemeint, jene bringe diesen hervor (*KrV* B 132). Dem entspricht seine Unterscheidung zwischen dem reinen urständlichen Sein und dem gegenständlichen Sich-Erscheinen des Ich (§§ 24/25 der Transzendentalen Deduktion). Wenn gilt (und Kant hat diese Überzeugung nie in Frage gestellt), dass Kenntnis nur von Phänomenen besteht, die sich einem Kennenden gegenüber in Gegenstandsposition befinden, kann vom Ich-Subjekt keine objektive Erkenntnis besessen werden, es bleibe eine vitiöse Voraussetzung.

Die Reflexionstheorie, wie Kant sie von Descartes und Leibniz (aber ebenso von zahlreichen Denkern des britischen Empirismus) übernimmt (Frank 2002), muss also das Phänomen voraussetzen, dessen Struktur sie aufzuklären antrat. Fichte konnte ihr darum in seinen Vorlesungen über die *Wissenschaftslehre nova methodo* (zwischen 1796 und 1799) „Sophisterei“ vorwerfen (Fichte 1797; in: Frank 1991).

Ich paraphrasiere die Grundzüge des Fichteschen Arguments. Wäre das Bewusstsein unserer selbst das Ergebnis einer Verschachtelung mehrerer Bewusstseine, deren jeweils vorletztes durch das jeweils letzte bezeugt würde, so gäbe es überhaupt kein Selbstbewusstsein. Denn für das jeweils letzte gälte abermals die glei-

che Bedingung, dass, um sich selbst bewusst zu werden, ein nachfolgendes Bewusstsein es zum Gegenstand machen müsste – ein Bewusstsein, das selbst unbewusst wäre und sein Bewusstsein von einem es vergegenständlichenden weiteren Bewusstsein zuerkannt bekäme, und so *ad infinitum* (so auch Sartre 1947; in: Frank 1991, 379 ff.). Nun *besteht* Bewusstsein, also ist das Modell falsch. Es ist falsch, heißt: Bewusstsein muss *unmittelbar*, jeder Vergegenständlichung durch ein nachfolgendes Bewusstsein zuvor, mit sich vertraut gewesen sein. Das drückt Fichte aus durch die Wendung von der völligen Nichtunterschiedenheit von Subjekt und Objekt im Selbstbewusstsein (so wiederum auch Sartre 1947; in: Frank 1991, 382<sub>,3</sub>). Nun ist ein unmittelbares Bewusstsein – in kantischer Terminologie – eine Anschauung. Da sie, anders als die sinnliche Anschauung, kein Seiendes in Raum und Zeit, sondern das Sein der lautereren Spontaneität der Apperzeption anschaut, muss ihr der Status der Intellektualität zugesprochen werden.

\*

Wenden wir uns dem *Second-order*-Modell von Bewusstsein noch einmal in anderem und größeren Zusammenhang zu, und versuchen wir dann eine Kritik mehr im Stil der analytischen Geistphilosophen, besonders Richard Morans und Lynne Rudder Bakers.

Das Reflexionsmodell ist eine Variante des Modells der inneren Wahrnehmung, das wir von Husserl her kennen. In Rosenthals Version kommt es dem von Fichte attackierten Reflexionsmodell in der Sache sogar noch näher, weil es den Be-

wusstseinsmodus, der für die Vergegenwärtigung des erststufigen („first order“) Bewusstseins zuständig ist, im Dunkeln belässt (während die früher – bei Gelegenheit von Husserl – abgewiesene Theorie die Wahrnehmung auszeichnet). Das gilt ebenso für die Leistungen der so genannten Aufmerksamkeit, die viele Philosophers of Mind mit dem Bewusstsein schlechthin verwechseln, ohne zu erklären, wie sie sich zu den Befunden des elementaren (unaufmerksamen) Erlebens verhält (so auch Rosenthal selbst: 1991, 466, l. 2).

Ein Verfechter dieses Modells war vor Rosenthal schon Hugh Mellor (1977/78). Mellor nimmt an, Überzeugungen („beliefs“) seien Dispositionen, die inferentiell unabsehbar mit anderen verflochten und in den wenigsten Fällen bewusst (= reflektiert, explizit erkannt, ins Licht der Aufmerksamkeit getaucht) sind. Bewusst werden sie durch „einen neuen geistigen Zustand“, der dem vorangehenden unbewussten eine Art „Zustimmung (assent)“ erteilt. Man ist jetzt *überzeugt, dass man überzeugt ist*.

Moran (2001, 28) sagt dazu: Erst mit der *Bewusstwerdung* werden mentale Zustände *kognitiv* interessant, vorher nicht. Stillschweigende Überzeugungen sind solange kein Wissen (geschweige ein Selbstwissen), „wie nicht explizit auf sie reflektiert wird“ (27). Außerdem: falschen/fehleitenden Überzeugungen gebührt unter Reflexion nicht Zustimmung, sondern Ablehnung („dissent“ [28]; dieser Punkt könnte auf einem Missverständnis Mellors beruhen; „assent“ meint vielleicht nur so etwas wie Freges „behauptende Kraft“ oder Brentanos „Bestätigung/Setzung“). Der entscheidende Einwand: Wenn Überzeugungen prinzipiell unbewusst sind, so sind es auch die *second-order*-Überzeugungen (die *Reflexionen*

auf die Überzeugungen erster Ordnung). Durch Verschachtelung oder Übereinander-Stülpen unbewusster Zustände entsteht aber kein Bewusstsein („without its ever occurring to consciousness“ [29 f.; vgl. xiii]). – Und schließlich: Das Auftauchen einer Überzeugung zweiter Ordnung liefert keinen Beitrag zur Aufklärung des spezifischen Unterschieds zwischen erstpösonlichem oder drittpösonlich-objektivem Bewusstsein eigener Überzeugungen (das ist in Kürze auch Frau Bakers Einwand). Nur das würde als eine solche Erklärung zählen, was offen legt, wie solches ‚unmittelbare‘ Bewusstsein eigener Einstellungen möglich ist und warum *nur die eigenen* Einstellungen auf solche Weise wissbar sind. Bewusstsein eigener Zustände *modifiziert* diese Zustände, „steht nicht außer ihnen, sondern dringt in sie ein“ (sie bestehen jetzt nicht nur, sondern gehen *mich* an), während das Betrachten eines Baumes oder fremden Verhaltens diesen bzw. dieses qualitativ nicht verändert ([30 f.: xiii]: ein Punkt, den Moran S. 36 mit Sartre belegt). Moran spricht von der „adverbialen Funktion“ des Bewusstseins.

Aber kehren wir zu Rosenthal als dem bekanntesten Vertreter des *second order*-oder Zwei-Stufen-Modells zurück. Wir hörten: Seine ‚Theorie höherstufiger Gedanken‘ erklärt das Aufkommen von Bewusstsein aus bewusstlosen Zuständen durch eine Iteration nicht-bewusster mentaler Zustände. ‚Nicht-bewusster‘ darum, weil ja das jeweils letzte höherstufige (*second order*) Bewusstsein der Kette selbst ein Bewusstsein erster Stufe ist, solange es nicht in den Skopus eines reflektierenden (höherstufigen, ‚*second order*‘) Bewusstsein gerät, das dann abermals wieder unbewusst ist. Subjekt S befindet sich also zunächst (zum Zeitpunkt  $t_1$ ) im mentalen Zustand M. Zum Zeitpunkt  $t_2$  denkt es, S habe den Gedanken, sich in M zu befinden. Und zu  $t_3$  denkt es, S denke, dass S sich im mentalen Zustand M befinde.

Lynne R. Baker („The First-Person Perspective: A Test for Naturalism“, in: *American Philosophical Quarterly*, Volume 35, Number 4, October 1998, 327-347) hat nun gezeigt, dass es zwei mögliche Interpretationen dieses Theorievorschlags gibt: Nach der 1. sind alle Gedanken der entsprechenden Iteration drittpersönlich, nach der 2. muss spätestens auf der dritten Stufe (3.) ein erfolgreich reflexiver Gedanke auftauchen (Baker 1998, 338).

Das wollen wir uns aber erst einmal genauer ansehen:

**First interpretation:** Mental state M is introspectively conscious if and only if there is some thinker or subject S and some time t such that:

1. at t, S is in a mental state M.
2. At t, S has a thought that S is in mental state M.
3. At t, S has a thought that S has a thought that S is in mental state M,

where S's reporting the thought that she had in (2) is sufficient for the truth of (3).

**Second interpretation:** Mental state M is conscious if and only if there is some thinker or subject S and some time t such that:

1. At t, S is in mental state M.
- 2a. At t, S has a thought that she\* is in mental state M.
- 3a. At t, S has a thought that she\* has a thought that she\* is in mental state M,

where „she“ attributes a first-person reference to S and where S's reporting the thought that she has in (2a) is sufficient for the truth of (3a). As we have seen at length, sentences like (3a), which [self-]attribute first-person reference, cannot be replaced without loss by any sentences – like (1) or iterations of (1) – that are free of the first person (338, rechte Spalte).

Wenn Sie die erste mit der zweiten Interpretation vergleichen, werden Sie sehen:

Der einzige Unterschied besteht im Austausch von „S“ durch „she\*“. „S“ ist eine anonyme Bezeichnung für das Wissens-Subjekt, „she\*“ aber eine solche, die notwendige Kenntnis dessen einschließt, dass S *selbst* es ist, die da auf sich Bezug nimmt. Den Asterisk hinter/über dem *she\** hat Castañeda in einem berühmten

Aufsatz 1966 typographisches Signal eingeführt; man spricht von notwendig wahren (also unfehlbar treffenden) intentionalen Selbstbezügen auch als von *he, himself locutions*. Eine andere – wie David Lewis favorisierte – Redeweise ist. *Attributions De se*. Wir werden uns nächste Woche – am Beispiel Chisholms – ausführlich mit ihnen auseinandersetzen.

Hier – zur Kritik an Rosenthal *second-order*-Modell – genügt Folgendes: S muss wissen, dass *er/sie selbst* es ist, die den Gedanken über sich fassen. Der Gedanke war: ‚S befindet sich um t im mentalen Zustand M.‘ Dieser Gedanke wird aber zu einem Gedanken *de se* dann und nur dann, wenn er spätestens auf der zweiten Stufe zustande kommt. Denn sonst hätte das Subjekt der dritten Stufe kein Kriterium dafür, dass das reflektierte S der zweiten Stufe *es selbst*, und nicht ein anonymer Gegenstand ist. Anonyme Selbstreferenzen implizieren kein *Selbstwissen*.  
Wenn ich den Satz:

- a) Marie glaubt, dass *sie selbst* das beste Abitur von Schriesheim bestehen wird

durch ein einen Intentional-Satz mit anonymen Selbstverweis ersetze:

- b) Es gibt ein x, x ist identisch mit die Marie, und x wird von x für die beste Abiturientin von Schriesheim gehalten,

dann habe ich die Garantie des notwendigen Selbstwissens eingebüßt; denn natürlich kann Marie auf sich selbst verweisen, ohne 1. zu wissen, dass sie selbst

Marie ist (man hat nach der Geburt ihr Namensschildchen am Arm verwechselt, und nun hält sie sich für Karoline), und ohne 2. zu wissen, dass von ihr selbst die Rede ist, wenn man von der besten Schriesheimer Abiturientin spricht. Im Beispielsatz a) – mit der Rede *de se* – hatte sie aber ein notwendiges Wissen davon.

Rosenthals Modell kann also keinerlei Rechenschaft ablegen über das Zustandekommen berechtigter erstpersönlicher Gedanken.<sup>3</sup> Anders gesagt: Ohne Selbstwissen je schon vorauszusetzen, kann die Wahrheit der Einsicht dritter Stufe nicht verständlich gemacht werden: „Satisfaction of conditions (1)-(3) does not account for a thinker’s being in a mental state of which she is introspectively conscious“ (339).<sup>4</sup>

Wie aber ein(e) Denker(in) erfolgreich in einen solcherart auf *de re* irreduziblen introspektiven bzw. selbstbewussten Zustand kommen kann, das zeigt Frau Baker so wenig wie die von ihr angegriffenen Autoren.

---

<sup>3</sup> Während der Vorlesung machte mich ein Teilnehmer völlig mit Recht auf einen wesentlichen Unterschied meiner Kritik an Rosenthals Reflexions-Modell und der von Frau Baker aufmerksam. Ihr geht es um die Irreduzibilität von der *De-se*-Bezügen auf *de re*, mir aber nur darum, den Zirkel aufzuzeigen, der sich ergibt, wenn man Bewusstsein durch Reflexion auf Unbewusstes erklären will. Die beiden Zirkel sind von sehr ähnlicher Natur, aber sie unterscheiden sich doch deutlich. Die über-übernächste Vorlesung über Chisholm, Castañeda und die *he, himself locution* wird hier Klarheit bringen.

<sup>4</sup> Baker zeigt, dass analoge Einwände auch gegen David M. Armstrongs Idee eines sich selbst scannenden Gehirns oder gegen Daniel Dennetts Benutzung von auf Tonbandträgern gespeicherten aufrichtigen ‚ich‘-Äußerungen zur angeblich rein drittpersönlichen Erklärung der first person perspective gelten (Baker 1998, 341 f.), und sieht darin einen grundsätzlichen „challenge to Naturalism“. Mir genügt es zu sagen, dass Bakers Argument einen grundsätzlichen Einwand gegen den Objektualismus von Selbstbewusstsein begründet.

## 7. (de facto 5.) Vorlesung vom 22. Mai

Der erste Satz eines berühmten Artikels aus den 70er Jahren zum Leib-Seele-Problem lautet: „Consciousness is what makes the mind-body problem really intractable“ („What is it like to be a bat?“, in: ders., *Mortal Questions*, Cambridge University Press: 1979, 164-180, hier: 164). Man könnte ihn *ad hoc* so umakzentuieren: Es gäbe kein Leib-Seele-Problem, wenn es kein Bewusstsein gäbe. Wenig später schreibt Nagel:

Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless. The most important and characteristic features of conscious mental phenomena is very poorly understood. Most reductionist theories do not even try to explain it. And careful examination will show that no currently available concept of reduction is applicable to it (166).

Das Bewusstsein verhält sich zu den Leib-Seele-Theorien etwa wie in einem Gedicht Enzensbergers die Leute zur Umerziehung:

Einfach vortrefflich  
All diese großen Pläne:  
Das Goldene Zeitalter  
Das Reich Gottes auf Erden  
Das Absterben des Staates.  
Durchaus einleuchtend.

Wenn nur die Leute nicht wären!  
Immer und überall stören die Leute.  
Alles bringen sie durcheinander.

Wenn es um die Befreiung der Menschheit geht  
Laufen sie zum Friseur.  
Statt begeistert hinter der Vorhut herzutrippeln  
Sagen sie: Jetzt wär ein Bier gut.  
Statt um die gerechte Sache  
Kämpfen sie mit Kramfadern und mit Masern.  
Im entscheidenden Augenblick  
Suchen sie einen Briefkasten oder ein Bett.  
Kurz bevor das Millennium anbricht  
Kochen sie Windeln.

An den Leuten scheitert eben alles.

Mit denen ist kein Staat zu machen.  
Ein Sack Flöhe ist nichts dagegen.

[...]

(*Gedichte 1950-1985*, Frankfurt/M.: Suhrkamp, 1986, st 1360, 68)

Unsere Vorlesung hat ‚Bewusstsein‘ zum Thema. Darum sollten wir zunächst einmal das Motiv von Thomas Nagels Aufsatz zu verstehen versuchen. Ich ergänze dazu den zitierten ersten Satz des Fledermaus-Aufsatzes um den ganzen ersten Abschnitt:

Consciousness is what makes the mind-body problem really intractable. Perhaps that is why current discussions of the problem give it little attention or get it obviously wrong. The recent wave of reductionist euphoria has produced several analyses of mental phenomena and mental concepts designed to explain the possibility of some variety of materialism, psychophysical identification, or reduction. But the problems dealt with are those common to this type of reductionism and other types, and what makes the mind-body problem unique, and unlike the water-H<sub>2</sub>O problem or the Turing machine-IBM machine problem or the lightning-electrical discharge problem or the gene-DNA problem or the oak tree-hydrocarbon problem, is ignored (Nagel 1979, 165).

Hier wird gleich eine ganze Flut theoretischer Einstellungen zum Leib-Seele-Problem aufgerufen: die Identitätstheorie, die seelische Zustände eins zu eins mit geeigneten elektromagnetischen Vorgängen identifiziert; der Funktionalismus, der sich an einem Computer-Modell orientiert und mentale Zustände über ihre ‚kausale Rolle‘ identifiziert, die gleichsam alternativ mental oder materiell implementiert sein kann und obendrein in verschiedenen Materien: statt in Aminosäuren etwa in Silikon (oder, wie Hilary Putnam scherzt: in Schweizer Käse).

Alle diese Theorien, sagt Nagel, haben aber eines gemeinsam. Sie versuchen, die offenkundige intuitive Unähnlichkeit zwischen geistigen und materiellen Prozessen zu reduzieren auf objektiv bestehende Identitäten, deren beide Glieder auf den ersten Blick ebenso unähnlich erscheinen, in Wahrheit aber eben identisch

sind: Wasser und H<sub>2</sub>O, Wärme und mittlere Molekülbewegung, Wellen und Korpuskel, Kohlenwasserstoff und Eichenholz usw. Und von diesen Identifikationen des scheinbar Unähnlichen sagt Nagel, dass sie in allen Beispielen misslingt, in denen auf der einen Seite der Gleichung Bewusstsein steht. Nagel geht noch weiter: Unsere zeitgenössische Naturwissenschaft hat auch nicht einmal eine Ahnung davon, wie zwischen Zuständen, die wir ‚bewusst‘ nennen, und Hirnzuständen eine Identitäts-Relation angenommen werden könnte. Ein neuerer und sehr erfolgreicher Bewusstseinstheoretiker, David J. Chalmers, bläst in dasselbe Horn und schreibt:

I argue that reductive explanation of consciousness is impossible, and I even argue for a form of dualism.

[...]

Materialism is a beautiful and compelling view of the world, but to account for consciousness, we have to go beyond the resources it provides.

[...]

Insofar as consciousness is the really difficult problem for a science of the mind, it is phenomenal consciousness [*P*-consciousness] that is central. The problems here are of a different order of magnitude. Even after we have explained the physical and computational functioning of a conscious system, we still need to explain why the system has conscious experiences (*The Conscious Mind. In Search of a Fundamental Theory*, Oxford University Press, 1996, xiv und 29).

Das ist eine sehr starke These. Und ich schlage Ihnen vor, mit der einfacheren Frage zu beginnen: Wie könnte denn ein Leib-Seele-Theorie *ohne Bewusstsein* überhaupt aussehen? Denn Bewusstsein ist ja nach Nagel die große Hintertreiberin gelingender Leib-Seele-Theorien.

Darauf haben wir aber sowohl von Ned Block wie von David Rosenthal schon eine Antwort bekommen: Es könnte doch sein – meinen sie –, dass es geistige Zustände gibt, die irgendwie mit materiellen Zuständen identisch sind und die uns nicht bewusst sind. Dafür muss man nicht unbedingt Freuds ‚Unbewusstes‘ als

Instanz aufrufen. Es genügt, sich klarzumachen, dass wir auf unser Seelenleben nicht immer *Acht haben*, also unaufmerksam sind – wie ein Fernfahrer, der nicht jede Ampel, auf die er korrekt reagiert, aufmerksam wahrnimmt („the distracted driver“). Block hatte uns mit dem Fall des Dampfrahmen-Hörers konfrontiert, der sie auch unaufmerksam gehört hat. (Hier wollen wir aber nicht in den Fehler verfallen, die wir in der zweiten Vorlesung analysiert hatten: Wir wollen Bewusstsein nicht mit Aufmerksamkeit verwechseln.)

Es gibt aber andere Phänomene unseres Seelenlebens, denen wir kein Bewusstsein, wenigstens kein aktuell auftretendes (,okkurrentes‘) Bewusstsein zuschreiben werden: Es handelt sich erstens um das große Feld dessen, was Block *Unconsciousness* genannt hatte: also um die propositionalen Einstellungen, die wir evtl. für reduzierbar halten auf Zustände eines intelligenten Rechners (Chalmers nennt sie ,psychologisch‘ und fügt hinzu, die beiden Varietäten des Bewusstseins, das Phänomenale und das Psychologische, „erschöpfen den Bereich des Geistigen“ (16 u.). Es handelt sich aber insbesondere um „standing attitudes of a person“ wie Überzeugungen, Dispositionen, Einstellungen, intuitives Regelwissen, Gesinnungen, emotionale Grundhaltungen (etwa jähzornig oder eifersüchtig zu sein); die sind ganz oder teilweise unbewusst, d. h. der ,Introspektion‘ ganz oder teilweise unzugänglich. Verweilen wir zur Illustration an den Beispielen des Gehässig- oder Cholerisch-Seins oder des Überzeugungen-Haben. Wenn Sie aufgefordert werden, Ihre Überzeugungen (also die Menge aller Sachverhalte, die Sie für wahr oder richtig halten) herzusagen, werden Sie vielleicht einige über die Lippen bringen; die meisten merken Sie aber erst, wenn sie durch eine Situation provoziert werden, sie sich bewusst zu machen. Viele Leute wissen gar nicht, dass

sie cholertisch oder eifersüchtig oder missgünstig sind (schließlich schreibt sich keiner gern unangenehme Eigenschaften zu); andere sind dann umso heilsichtiger und gnadenloser gegenüber diesen Untugenden. Auch sprechen wir fließend eine Muttersprache, ohne uns der meisten grammatischen regeln bewusst zu sein. Ganz dramatisch ist das Unbewusstsein, das bei gewissen Hirnstörungen wie der Blindsight oder der Prosopagnosie auftritt.

Es gibt sogar noch radikalere Positionen als die von Philosophen, die *einige Bereiche* unseres Seelenlebens für unbewusst erklären. Dennett und andere Reduktionisten dehnen den Zweifel an der epistemischen Zugänglichkeit mancher mentalen Zustände oder Dispositionen auf alle aus: „No activity of mind is ever conscious“ (Lashley, zit. nach Dennett [6]). Dazu sagt Richard Moran: Man kann das zugeben, also □□□ihre Verlässlichkeit epistemologisch offen halten (10), ohne □□□das Auszeichnende dieser Gruppe von mentalen Erlebnissen zu leugnen: Sie sind immer noch von drittpersönlichen Berichten spezifisch verschieden (das heißt, es ist ganz etwas anderes, verliebt zu sein und einen anderen verliebt zu sehen: Moran 2001, 11 f.); und sie kommen zustande, „ohne aus Beobachtungsbeobachtungen erschlossen zu werden [inferred from observational evidence]“ (10, 11).<sup>5</sup>

Wir könnten auch ganz bescheiden sagen: Bis auf Dennett und ganz wenige radikale Eliminativisten [Wort erklären!] geben fast alle Geist-Philosophen zu, dass es Bewusstsein *gibt*, selbst wenn sie sein Auftreten quantitativ stark eingrenzen. Da-

---

<sup>5</sup> Außerdem: Die psychoanalytische Therapie setzt auf die Überführbarkeit von dem Wissen entzogenen psychischen Zuständen in gewusste und damit rational kontrollierbare. Also muss sie eine Kontinuität zwischen beiden Typen psychischer Zustände unterstellen. Wie ist das aber zirkelfrei möglich, wenn a) angeblich unbewusste Zustände sich im Verlauf als ‚wesentlich‘ bewusst heraus stellen können und b) zwischen Fremd- und Ich-Perspektive eine unüberbrückbare Kluft – „asymmetry“ – besteht?

vidson hat einmal gesagt: Wie könnte ich zwei Phänomene miteinander identifizieren, wenn es gar nicht zwei, sondern nur eines gibt? Ich hätte dann statt ‚Geist = Materie‘ nur gesagt: „Materie = Materie“.

Identität – so ist Davidsons Gedanke – darf nicht eliminativ verstanden werden, so, als meine ‚Identität von Natur und Geist‘ dies: ‚Natur, *und nicht* Geist‘ oder: ‚Geist nur, sofern er – unter genauer Analyse – in Natur kollabiert‘. Identität, hat er betont, muss als ein symmetrisches Verhältnis respektiert werden:

I see no good reason for calling identity theories ‚materialist‘: if some mental events are physical events, this makes them no more physical than mental. Identity is a symmetrical relation (Davidson, „Knowing One’s Own Mind“, in: *Proceedings and Adresses of the American Philosophical Association*, LX [1987], 441-458, hier: 453).

(Das gilt natürlich, wie Heinrich Heine auf sehr witzige Weise formuliert hat, auch für den umgekehrten idealistischen Fehlschluss, der Materie auf Geist reduziert. Heine schreibt – und denkt an naturfreundliche Identitätsphilosophen wie Schelling:

„Gott“, welcher [...] von den deutschen Philosophen das Absolute genannt wird, ‚ist alles was da ist‘, er ist sowohl Materie wie Geist, beides ist gleich göttlich, und wer die heilige Materie beleidigt, ist dabei so sündhaft, wie der, welcher sündigt gegen den heiligen Geist [Heine 1997, 565 f.]

Nagel konzentriert sich freilich nicht auf den idealistischen, sondern auf den materialistischen Reduktionismus. Und ihm wirft er – noch einmal – vor, die tiefe Unähnlichkeit zwischen gleichen Materien, die in ungleichen Erscheinungsformen auftreten, zur Relation Bewusstsein-Hirnprozess nicht zu sehen. Es ist für Bewusstsein *wesentlich*, sich irgendwie anzufühlen. Und genau in diesem Zug sieht Thomas Nagel das irreduzible Charakteristikum von Bewusstsein, das bei einer physikalistischen Reduktion über Bord ginge. Darum, sagt er, ist der Physi-

kalismus nicht etwa falsch, sondern nur *unvollständig*. Der Physikalismus ist darin ein Erkenntnis-Hemmnis, dass er seine Unvollständigkeit nicht bemerkt und sich für vollständig hält. Tatsächlich erklärt er nicht diesen wesentlichen Zug unseres Bewusstseins, den wir materiellen oder funktionalen Zuständen nicht zuschreiben würden: Wer bei Bewusstsein ist, dem ist irgendwie zumute. Es ist ihm nicht nur irgendwie zumute, sondern er/sie *könnte* gar nicht bei Bewusstsein sein, wenn ihm/ihr nicht irgendwie zumute wäre. Nagel spricht vom Zumutesein als von einer notwendigen und auch hinreichenden Bedingung für Bewusstsein. Die berühmt gewordenen englischen Ausdrücke sind: ‚It is like something to be in such-and-so state for the organism.‘ ‚It feels somehow.‘

[...] fundamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism – something it is like *for* the organism (166<sub>2</sub>).

Das würden wir aber von Hirnfaserreizungen nicht sagen: Für die ist es nicht auf gleiche Weise *wesentlich/notwendig*, dass sie sich anfühlen. Wir können uns sehr wohl vorstellen, dass *sie* oder unser ganzes Gehirn nichts fühlen und doch alle Funktionen unseres Seelenlebens normal weiterlaufen. Wir können uns sogar roboterartige Wesen vorstellen – Block und Chalmers nennen sie ‚Zombies‘ –, die uns in allen behavioralen Eigenschaften [erklären!] gleichen, denen nur nie irgendwie zumute ist: Sie lachen wie wir über einen Witz, weinen über den Tod ihrer Mutter, werden wütend, wenn man sie beleidigt und schauen nachdenklich, wenn man sie mit einem Problem konfrontiert. Sie sind ganz wie wir, sie funktionieren genau wie wir; es fehlt ihnen nur eines, das wir uns mit erstaunlicher Gewissheit zuschreiben: Bewusstsein. Sie sind keine Subjekte, würde Nagel sagen. Wem nämlich irgendwie beim Schmerzhaben oder Rot-Sehen oder Verliebt-Sein zumute ist, der oder die ist ein Subjekt. Zombies sind *per definitionem* keine Sub-

jekte. Und doch sind sie in keinem physikalisch angebbaren Aspekt von uns verschieden. Darum würde die Hirnforschung gar nichts verlieren, sie würde nicht einmal etwas bemerken, wenn es bei den von ihr untersuchten Gegenständen kein Bewusstsein gäbe:

It [sc.: the subjective character of experience] is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence (166 u.).

Nun *haben* wir aber Bewusstsein, sagt Nagel, und dass wir Bewusstsein haben, dessen sind uns auf die gute alte cartesianische Weise introspektiv *gewiss*; also ist der Physikalismus unvollständig (167). Er kann nicht Rechenschaft geben über ein Merkmal unserer Psychologie, das unter kontingenten Naturgesetzen gar nicht auftreten könnte, weil sein modaler Charakter ‚notwendig‘ ist. Notwendige Eigenschaften nennen wir auch ‚wesentlich‘. Die ‚subjektive Perspektive‘, sagt Nagel, ist gewissen Körpergefühlen und Wahrnehmungszuständen wesentlich. Wir kriegen sie nur los, wenn wir das Bewusstsein selbst loswerden, das mit den entsprechenden seelischen Zuständen verbunden ist. Sie mögen dann physisch sein, aber sie sind uns nicht mehr auf die vertraute Weise bewusst; sie haben, wie wir uns charakteristisch ausdrücken, ihre ‚Innerlichkeit‘ verloren und sich in ‚äußere‘ Zustände der Natur verwandelt. Es sind eben keine bewussten Erlebnisse mehr.

Die letzten Beobachtungen geben uns Gelegenheit, uns grundsätzlich über die Bedeutung der Ausdrücke ‚Reduktion‘ und ‚Physikalismus‘ (bzw. ‚Naturalismus‘) zu verständigen. Sie spielen eine so große Rolle in der Philosophy of Mind.

*Naturalisieren* heißt: Phänomene unseres seelisch-geistigen Lebens auf natürliche Tatsachen zurückzuführen, wie es die sind, von denen die Physik (als Basis-Wissenschaft aller Entitäten des Raum-Zeit-Systems) handelt. Dabei werden zwei

Varietäten unterschiedlicher Radikalität unterschieden: Geistig-seelische Phänomene werden zurückgeführt auf

a) Entitäten des Raum-Zeit-Systems → *Naturalismus*;

b) ausschließlich physische Entitäten, wie sie von einer idealisierten Physik am Ende der Tage („completed physics) erfasst werden: → *Physikalismus*.

(Klarerweise ist Position b) radikaler, da sie nur Einsichten der exakten Naturwissenschaften gelten lässt; in a) kämen allenfalls Entitäten wie geistige Substanzen oder Gespenster unter.)

Die Naturalisierung ist eine Form der *Reduktion*. Und unter ‚Reduktion‘ versteht man die Ersetzung einer Klasse von Phänomenen durch eine andere *salva veritate*, also ohne Verletzung/Veränderung des Wahrheitswerts. Für Reduktionen spricht die ontologische Sparsamkeitsmaxime („Occams Rasiermesser“), nach der Entitäten nicht ohne Grund verfielfältigt werden sollen. Im gegebenen Fall: Kann ich Geistig-Seelisches auf Natürliches reduzieren, so ist meine Ontologie sparsamer als die des Substanz-Dualismus, der für Geist und Leib zwei irreduzible Seinsbezirke annehmen muss.

Naturalisten berufen sich – wie wir gesehen haben – gern auf Analogien zwischen naturalen Gegebenheiten, die sich auf den ersten Augenschein als zweierlei zeigen, aber doch nur eines sind: Wellen und Korpuskeln, Masse und Energie u. dgl. Jeder Reduktionist, spottet Nagel, hat seine Lieblingsanalogie (Nagel 1979, 166 o.). Nun zeigt sich – und Nagel war nur einer der ersten, die das geltend gemacht haben –: Die Analogie, die etwa zwischen Wasser und H<sub>2</sub>O oder zwischen Wärme und mittlerer Molekülbewegung besteht, lässt sich nicht ebenso (etwa) auf das Verhältnis zwischen Schmerz und der entsprechenden C-Faser-Reizung übertragen (Kripke 1980, 152 ff.; darüber mehr in der 10. Vorlesung vom 3. Juli).

Der Dualismus – auf den ich am Beispiel Frank Jacksons schon kommenden Woche einen genaueren Blick werfen werde – ist aber nicht *meine* theoretische Option.

Ich lasse mich lediglich beeindrucken von den genannten *Prima-facie*-Befunden. Sie liefern mir ein Motiv dafür, mich der idealistisch-phänomenologischen Tradition wieder zu besinnen, die Subjektivität für nicht auf geistunabhängige Naturprozesse reduzierbar hielt. Dazu veranlasst uns nicht Wissenschafts- oder Fortschrittsfeindlichkeit. Im Gegenteil: Theorien erstarken, je stärkeren Einwänden sie sich gewachsen zeigen. Die Wissenschaften – Psychologie oder Neurobiologie – tun gut daran, die in der *Philosophy of Mind* gelegentlich beschworene Sorge einer *Erklärungslücke* zunächst einmal ernst zu nehmen.<sup>6</sup> Es bleibt für uns ein unauflösliches Rätsel, dass elektrochemische Abläufe in den Neuronenbahnen in einem intrinsischen Bezug zu dem stehen sollen, was wir subjektives Erleben nennen. Dies Rätsel löst sich auch nicht durch eine neurowissenschaftliche Erklärung, wenn es sie denn gäbe, und selbst nicht durch unsere Überzeugung von ihrer Richtigkeit. Gern zitiert wird in diesem Zusammenhang Emil du Bois-Reymond, der Begründer der experimentellen Physiologie. Er hatte das Rätsel des Bewusstseins nicht für relativ zu unserem Wissensstand unlösbar, sondern für grundsätzlich gehalten (du Bois-Reymond 1974, 65). Ähnlich hatte Julian Huxley das Aufkommen von Bewusstsein aus einer Nervengewebs-Reizung („initiating nerve tissue“) nicht weniger unverständlich genannt als das Erscheinen des Djin, wenn Aladin an seiner Wunderlampe wischt (Huxley 1986, 193).

---

<sup>6</sup> Dazu seither Tobias Schlicht, *Erkenntnistheoretischer Dualismus. Der Problem der Erklärungslücke in Geist-Gehirn-Theorien*, Paderborn: mentis, 2007, bes. Kap. I.

\*

Ich will nun Nagels Fledermaus-Aufsatz genauer in seinem wirklichen Argumentationsablauf nachvollziehen, so wie wir das ja parallel im begleitenden Proseminar tun wollen.

Zunächst einen historischen Hinweis: Nagel ist eigentlich ein geborener physikalischer Reduktionist, In einem eindrucksvollen Anlauf hatte er diese Position 1963 noch selbst verteidigt („Physicalism“, in: *The Philosophical Review* 74 [1965], 339-56). Freilich hatte er im V. Abschnitt des Textes eine unüberwindliche Verlegenheit gegenüber der ‚essentiellen Subjektivität [oder Innerlichkeit] psychologischer Zustände‘ eingeräumt. Diese Verlegenheit hatte er so begründet: Wir haben eine unüberwindliche Intuition, dass wir Bewusstsein haben und dass wir mit ‚Bewusstsein‘ etwas anderes meinen als ‚einfach nur Materie – und sonst nichts‘. Einerseits weiß ich, dass ich eine körperliche Person bin – wie viele andere. Andererseits – gleichsam mit den Augen des Geistes gesehen – scheint solch eine Beschreibung etwas für mich Wesentliches auszulassen: nämlich die Tatsache, dass *ich* es bin, der so beschrieben wird (vgl. Bieri, Einl. zu „Selbstbewusstsein, Privatheit und Subjektivität“, in: ders., *Analytische Philosophie des Geistes*, I. c., 204 f.). Und Bezugnahmen auf uns selbst *als* auf uns selbst haben semantische Besonderheiten, auf die wir in der übernächsten Stunde zu sprechen kommen werden (sie gelingen nur über ‚essentielle Indexikalien‘, Beschreibungen und demonstrative Bezugnahmen können merkwürdigerweise fehlgehen), Diese Intuition kann manchmal so weit gehen, dass wir das Ich für eine ‚transzendente‘ Entität *sui*

*generis* halten, also etwas, das in einem bestimmten Sinne nicht in der Welt ist.

(Erinnern Sie sich an Wittgensteins Sätze aus dem *Tractatus*:

5.632 Das Subjekt gehört nicht zur Welt, sondern es ist eine Grenze der Welt.

5.633 Wo *in* der Welt ist ein metaphysisches Subjekt zu merken?

Du sagst, es verhält sich hier ganz wie mit Auge und Gesichtsfeld. Aber das Auge siehst du wirklich *nicht*.

Und nichts *am Gesichtsfeld* lässt darauf schließen, dass es von einem Auge gesehen wird.)

Darum würde, wer mich als vollständig zur physischen Welt gehörig beschreiben würde, einen Irrtum begehen. Irgendwie passt unsere Vorstellung von uns selbst nicht zu dem Modell eines Subjekts mit Eigenschaften.

Eine Beschreibung der Welt wäre objektiv, wenn sie alle wahren Aussagen über die Welt enthielte. Wahre Aussagen stehen für Tatsachen; die wahre und erschöpfende Theorie referierte also alle Tatsachen. Denken Sie an ein ideales Physiklehrbuch, geschrieben am St.-Nimmerleins-Tag. Dann wäre die Einstellung zu einem der vielen Sätze, in denen diese Tatsachen beschrieben werden, eine propositionale Einstellung. Eine propositionale Einstellung wird sprachlich ausgedrückt durch ein Satzsubjekt und Intentionsverb im Hauptsatz (,x glaubt, denkt, hofft, weiß') und einen eingebetteten Nebensatz: ‚dass dieser Gegenstand so-und-so beschaffen ist'. Propositionales Wissen ist also ein Wissen, dass... Wie aber, wenn es Tatsachen gäbe, die in einer spezifisch subjektiven Erlebnisperspektive bestünden, also den Welttatsachen spezifisch ungleichartig sind – so wie unsere Intuition das von unserer Ichheit angenommen hatte? Diese Erlebnisperspektive würde sich auflösen, wenn ich sie in der Weise konzeptualisieren würde, wie es das Physiklehrbuch tut. Subjektive Erlebnisse *sind* gar nicht (oder nicht erschöpfend) kon-

zeptualisierbar (Nagel sagt: in „objektives Wissen [knowledge]“ auflösbar [170, 172]; darin besteht ihr gleichzeitige Durchsichtigkeit und Vieldeutigkeit. Und wenn es sie trotzdem *gibt* (darin besteht Nagels ‚Realismus‘ hinsichtlich bewusster Zustände [Nagel, *The View from Nowhere*, Oxford University Press, 1986, 27: „Reality is not just objective reality.“]), dann liegt in ihnen mehr vor, als wir in einer vollständigen propositionalen Erfassung der Welt angeben könnten.

There are no comparable imaginative obstacles to the acquisition of *knowledge* about bat neurophysiology by human scientists, and intelligent bats or Martians might learn more about the human brain than we ever will (Nagel 1979, 172<sub>2</sub>[von mir kursiviert]).

Subjektives oder phänomenales Erleben besteht eben nicht in einem deskriptiven *Wissen, dass...*(etwas so-und-so beschaffen und mithin konzeptualisierbar ist), sondern in einem *Wissen, wie...* (es nämlich für den betreffenden Organismus ist, sich in dem betreffenden Zustand zu befinden). Darum kann man einem Lebewesen, dem nicht die Neurophysiologie einer Fledermaus zugrunde liegt, nicht in einer propositionalen Sprache beibringen, wie einer Fledermaus zumute ist. Um die Information zu verstehen, müssten wir zugleich wie menschliche und wie Fledermaus-Organismen konstituiert sein und die eine in die andere Organisation mitnehmen können; das ist aber ein widersprüchlicher und darum nicht zu realisierender Wunsch (169<sub>2</sub>).

Die Uneliminierbarkeit des Subjektiven beim Übergang zu propositionalem (z. B. zu Lehrbuch-) Wissen und umgekehrt: die Unbeeinflussbarkeit des subjektiven durch ein Tatsachen-Wissen ist Gegenstand vieler Texte der Weltliteratur. Nehmen wir ein schlichtes Beispiel, das *Märchen von einem, der auszog, das Fürchten zu lernen* aus den Brüdern Grimm. Der dumme von zwei Söhnen ist furchtlos und geht auf Wanderschaft, um das Gruseln zu lernen: das Vorbild für Wagners *Sieg-*

*fried.* Er gerät in lauter schauerliche Situationen, die ihn das Gruseln aber nicht lehren. Er lernt es aber auch erst recht nicht aus Beschreibungen, die andere ihm geben, also *nicht durch propositionales Wissen*. Nach einer Nacht mit langkralligen Gespensterkatzen und einem fahrenden Bett fragt ihn der Wirt: „Hast du nun gelernt, was gruseln ist?“ „Nein, sagte er, „ich weiß es nicht, wenn mirs nur einer sagen könnte!“ (DKV-Ausgabe, 40,<sub>1</sub>). So geht es ja auch Siegfried, dem Mime schauerliche Situationen schildert. Siegfried *versteht* sie, wie man Lehrbuchsätze versteht; aber er kann sich nicht vorstellen, wie einem sich Gruselnden *zumute* ist. Beide lernen übrigens das Fürchten mit der Liebe (oder eher: der Heirat).

Resümieren wir nun: Die subjektive Komponente alles Erlebens wird durch physikalistische Objektivierung nicht aufgelöst (,physikalistisch reduziert'), sondern zum Verschwinden gebracht. Sie *besteht* aber, und ihr Realität ist uns unmittelbar bekannt, ja sogar gewiss. Sprachlich artikuliert sie sich nicht in der Form eines *Wissens, dass...*, sondern als ein *Wissen, wie* (nämlich, wie es ist, in dem oder jenem psychischen Zustand sich zu befinden). Es kann mir unbekannt sein, in welchem Zustand ich mich befinde (d.h., es kann sein, dass ich den Zustand merkwürdigen Schwindels, der mich beim Gedanken an die Erlebnisse des gestrigen Nachmittags ergreift, nicht genau oder gar nicht durch ein Prädikat oder eine Prädikaten-Menge zu klassifizieren weiß: Habe ich mich verliebt, war ich nur in Hochstimmung, oder war mir nur einfach – wie man so sagt – blümerant zumute, etc.?). Unmöglich aber ist, dass es nicht irgendwie war, in diesem schwer klassifizierbaren Zustand gewesen zu sein (oder jetzt gerade) zu sein. Nagel illustriert das Eigentümliche dieses *Wissens, wie* in Abhebung vom *Wissen, dass* am Beispiel des auf Echolotung beruhenden Wahrnehmungsapparats einer Fledermaus – aber

auch ein Typ von  $\phi$ -Zuständen [Kürzel erklären!], der uns prinzipiell zugänglich ist, wäre zur Veranschaulichung dessen geeignet gewesen, worum es Th. Nagel zu tun ist, nämlich den Nachweis, dass ein (z.B. naturwissenschaftliches) Wissen um die objektiven Qualitäten des sensorischen Apparats eine (unter die Realitäten der Welt gehörige) Information uns niemals, und zwar prinzipiell nicht liefern kann: die, zu wissen, wie das ist (wie sich das anfühlt), in einem bestimmten durch diesen Apparat ermöglichten  $\phi$ -Zustand zu sein. Diese subjektive Seite der Erfahrung kann nur dadurch erworben werden, dass man sich in einem mentalen Zustand befindet, dass man ihn nicht nur aus der Außenperspektive durch ein Wissen, dass wahrheitsgemäß und regelkonform beschreiben kann, sondern ihn erlebt.

Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in a human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them (Nagel 1979, 171).

So if extrapolation from our own case is involved in the idea of what it is like to be a bat, the extrapolation must be incompletable. We cannot form more than a schematic conception of what it is like. For example, we may ascribe general types of experience on the basis of the animal's structure and behavior. Thus we describe bat sonar as a form of three-dimensional forward perception; we believe that bats feel some versions of pain, fear, hunger, and lust, and that they have other, more familiar types / of perception besides sonar. But we believe that these experiences also have in each case a specific subjective character, which it is beyond our ability to conceive. And if there is conscious life elsewhere in the universe, it is likely that some of it will not be describable even in the most general experiential terms available to us. (The problem is not confined zu exotic cases, however, for it exists between one person and another. The subjective character of the experience of a person deaf and blind from birth is not accessible for me, for example, nor presumably is mine to him. This does not prevent us each from believing that the other's experience has such a subjective character.) (l.c. 169 f.).

Man könnte das zugeben, aber einwenden, dieser subjektive Charakter allen Erlebens sei epistemisch irrelevant. Dabei übersieht man, dass er geradezu ein Definiens ist für  $\phi$ -Zustände. Hilary Putnam nennt den *pour-soi*-Bezug – also die Tatsa-

che, dass zum Zeugnis ihres Vorliegens nur das auffassende Subjekt selbst erfordert wird – das ‚engere Kriterium‘ für psychologische Zustände (vgl. Donald Davidson, „Knowing One's Own Mind“, in: *Proceedings and Addresses of the American Philosophical Association*, LX [1987], 441-458, hier: 444; Nagel selbst über die *pour-soi*-Perspektive in Nagel 1979, 168<sub>1</sub>). Etwas ist nach diesem Kriterium überhaupt nur darum und dadurch ein mentales oder psychisches Phänomen, dass es nicht einfach nur (wie der Granitblock auf der Alm) in der Welt vorhanden ist, sondern dass sein Sein sein Maß in dem Bewusstsein hat, das ich von ihm habe. Ein psychischer Zustand (nehmen wir die unbestimmte Angst, die mich seit Tagen verfolgt), ist, was sie ist, überhaupt nur in dem Maße, wie sie Angst-für-mich ist. Hört sie auf, für mich zu sein, so hört sie auf, Angst zu sein (oder überhaupt zu sein). „Was ich nicht weiß, macht mich nicht heiß“, sagt das Sprichwort. Nagel sagt dazu:

Does it make sense, in other words, to ask what my experiences are *really* like, as opposed to how they appear to me (Nagel 1979, 178<sub>2</sub>)?

Searle formuliert dasselbe Problem (populär) wie folgt:

In the case of consciousness, we do have the distinction between the ‚physical‘ processes and the subjective ‚mental‘ experiences, so why can't consciousness be redefined in terms of the neuro-physiological processes in the way that we redefined heat in terms of underlying physical processes? Well, of course, if we insisted on making the redefinition, we could. We could simply define, for example, ‚pain‘ as patterns of neuronal activity that cause subjective sensations of pain. And if such a redefinition took place, we would have achieved the same sort of reduction for pain that we have for heat. But of course, the reduction of pain to its physical reality still leaves the subjective experience of pain unreduced, just as the reduction of heat left the subjective experience of heat unreduced. Part of the point of the reductions was to carve off the subjective experiences and exclude them from the definition of the real phenomena, which are now defined in terms of those features that interest us most. But where the phenomena that interest us most are the subjective experiences themselves, there is no way to carve anything off. Part of the point of the reduction in the case of heat was to distinguish between the subjective appearance on the one hand and the underlying physical reality on the other. Indeed, it is a general feature of such reductions that the phenomenon is defined in terms of the ‚reality‘ and not in terms of the ‚appearance‘. But we can't make that/ sort of appearance-reality distinction for consciousness because consci-

ousness consists in the appearances themselves. *Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is reality. [...] Consciousness is an exception to this pattern [of distinguishability between ,objective physical reality', on the one hand, and mere ,subjective appearance', on the other] for a trivial reason. The reason, to repeat, is that the reductions that leave out the epistemic bases, the appearances, cannot work for the epistemic bases themselves. In such cases, the appearance is the reality (The Rediscovery of the Mind, MIT Press, 1992, 121 f.).*

Subjektives Erleben (eben das, was Nagel ‚Bewusstsein‘ nennt) scheint der einzige epistemische Fall zu sein, in dem Sein und Sich-Erscheinen zusammenfallen. Erinnern Sie sich an Husserls Unterscheidung zwischen innerem und äußerem Wahrnehmen. Während sich die Welt der Gegenstände unabsehbar vor unserem Blick abschattet, ist die ‚innere Welt‘ uns völlig adäquat gegeben, d. h.: Es ist von ihr genau ebenso viel bekannt, wie phänomenal gegeben ist. Es gibt keinen Überhang der (nicht erscheidenden) Wirklichkeit über ihre subjektive Auffassung.

Man könnte das auch so ausdrücken: Während bei unserem Bewusstsein von Katzen, Rauchverbotsschilder und anderen Gegenständen der äußeren Welt diese Gegenstände unabhängig von unserem Bewusstsein sind, gilt das für Wissen über Bewusstsein allein nicht. Darin sehen manche Philosophen (wie etwa Terry Horgan & Uriah Kriegel 2007, „Phenomenal Epistemology: What is Consciousness that We may Know It so Well?“, ungedruckt, S. 12).

Wir werden beim nächsten Treffen einen anderen wirkungsmächtigen Einwand gegen den Naturalismus über phänomenale Zustände durchsprechen: Frank Jackson ebenfalls wirkungsmächtigen Aufsatz „Epiphenomal Qualia“ von 1982.

## 8. (de facto: 6.) Vorlesung vom 5. Juni

Von dem, was Nagel ‚phänomenales Bewusstsein‘ nennt, wird üblicherweise – und schon seit John Locke – als von ‚Qualia‘ (oder auch von ‚qualitativem Bewusstsein‘) gesprochen. Gemeint sind Zustände des Zumuteseins, und sie scheinen ausschlaggebend dafür zu sein, dass überhaupt Bewusstsein vorliegt. Intentionale Zustände mögen, wie Husserl annahm, *auch* eine qualitative Seite haben; aber Block und Chalmers meinen, dass sie sich als Zustände eines probabilistischen Automaten von der Art eines hinreichend komplexen Computers beschreiben lassen – und Robotern und Zombies wollte Nagel – Sie erinnern sich – nicht notwendig Bewusstsein zuschreiben (Chalmers 1996). Qualia sind aber wesentlich mit ihrem eigentlichen Vorliegen bekannt, und diese Bekanntschaft ist das, was ‚Bewusstsein‘ genannt wird. Daher Nagels Seufzer: *Bewusstsein* (nicht Denken) ist das, was Leib-Seele-Theorien so vetrackt macht!

In eine ähnliche Richtung weist auch ein anderer Artikel, der in der Zwischenzeit nicht minder heftige Debatten in der analytischen Philosophie hervorgerufen hat als Nagels Fledermaus-Aufsatz. Ich meine des australischen Philosophen Frank Jacksons Text „Epiphenomenal Qualia“. Er ist zuerst im *Philosophical Quaterly* 32 (1982), pp. 127-36 erschienen und wiederabgedruckt in der Anthologie *Mind and Cognition* von William G. Lycan (Oxford: Basil Blackwell, 1990, 469-477; ich zitiere nach diesem späteren Abdruck, der auch dem Reader des begleitenden Proseminars zugrunde liegt).

Was ein Quale (lat., Plural: Qualia) ist, haben wir eben gesehen. Der zweite dunkle Ausdruck, der im Titel vorkommt, ist ‚Epiphänomenalismus‘. So nennt man in der Philosophie des Geistes eine Position, die Bewusstsein zwar für irreduzibel auf physische Prozesse erklärt, aber um einen hohen Preis: Bewusstsein wird jede kausale Wirksamkeit auf das Naturgeschehen abgesprochen. Es ist bloß noch ein Epi-phänomen, eine wirkungslose Nebenerscheinung von Naturprozessen. Es wäre sozusagen das Endprodukt einer Kausalkette, die selbst keine weitere kausale Kraft mehr hat: wie der Rauch, der nach einem Blitzschlag aus einer verbrannten Holzhütte steigt.

‚Epiphänomenalisten‘ nennt man also Vertreter einer dualistischen Überzeugung. Sie halten bewusste Prozesse nicht für körperlich, glauben aber auf der anderen Seite auch nicht, dass Bewusstsein irgendwelche Veränderungen im Organismus eines Lebewesens hervorbringen könne. Der Epiphänomenalismus scheint durch die vieldiskutierten Experimente empirische Bestätigung gefunden zu haben, die Benjamin Libet an Versuchspersonen durchgeführt hat. Die Probanden sollten ein Zeichen geben, wann sie sich zur Ausführung einer Handlung entschlossen hatten. Nun meint Libet gezeigt zu haben, dass der Entschluss im *Bewusstsein* erst ankam, als schon einige 100 Millisekunden lang das ihn auslösende Bereitschaftspotential durch ein EEG (Elektro-Enzephalo-Gramm) gemessen worden war. Das Bewusstsein würde mithin kausal unwirksam sein und nur nachträglich die physische Verursachung repräsentieren: zu spät, um die gewollte Handlung intentional auszulösen. Beckermann resümiert die Position des Epiphänomenalismus wie folgt:

Zustände im Geist [ich verbessere: im Bewusstsein] einer Person werden zwar durch Zustände in ihrem Körper verursacht, haben aber selbst niemals Wirkungen

auf ihren Körper (Ansgar Beckermann, *Analytische Einführung in die Philosophie des Geistes*, Berlin-New York: de Gruyter, 1999, 48).

In einer relativ populären Einführung in die Philosophie des Geistes hat Frank Jackson den Epiphänomenalismus umständlicher, aber genauer so bestimmt:

*[E]piphenomenalism: the doctrine according to which mental properties (and substances if it comes to that) are causally impotent – they make no difference to what happens. They may themselves have causes, but they do not cause anything. If dualists are committed to epiphenomenalism, then they are committed to saying that the very feature that makes something a desire for coffee is causally irrelevant to movement towards coffee. Similarly, the very feature that makes something an intention to get married or an itch is irrelevant to signing marriage papers or scratching. This is hard to believe (David Braddon Mitchell and Frank Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell, 1996, 6).*

Aber Wahrheiten stoßen sich oft mit unseren Alltagsintuitionen, und in „Epiphenomenal Qualia“ war Jackson noch streng von der Unausweichlichkeit der epiphenomenalistischen Konsequenz überzeugt. Sie werden gleich sehen: Nach Jackson reicht sie immerhin aus, dem Physikalismus Paroli zu bieten – jedenfalls dann, wenn sich der Physikalismus anmaßt, auf *alle* Phänomene zuzutreffen, die es überhaupt in der Welt gibt. Z. B. auch auf Qualia.

Folgen wir erst mal der Argumentation von Jacksons bündig formuliertem Aufsatz.

Wäre der Physikalismus wahr – das ist Jacksons Ausgangsüberlegung –, dann müsste alles, was man überhaupt ‚existierend‘ nennen könnte, physikalisch erklärt werden können. Nennen wir das, was wir durch die Physik als letzte Basiswissenschaft lernen können, ‚physikalische Information‘. (Unter ‚physikalischer Information‘ soll auch alles verstanden werden, was sich funktionalistisch über den Zustand eines intelligenten Organismus ausmachen lässt; z. B., welches die

kausale Rolle dieses Zustands im Gesamt eines Organismus ist [Jackson 1990, 469,1].)

Nun ist Jacksons These, dass der Physikalismus nach dieser Definition eine falsche, eine viel zu starke theoretische Annahme darstellt. Er outet sich als „qualia freak“ und macht geltend, dass alle jene Wahrnehmungs- und körperlichen Erlebnisse, die wir Qualia nennen von einer vollständigen physikalischen Beschreibung nicht einmal erfasst werden. Nun *bestehen* sie aber. Also – das ist der Refrain, der sich durch Jacksons Text durchhält – ist der Physikalismus falsch.

Er ist falsch nicht in dem, was er sagt. Er ist falsch in dem, was er nicht sagt, weil er es nicht in seiner Begrifflichkeit enthalten ist. Der Physikalismus ist falsch, weil er sich für eine *vollständige, eine erschöpfende* Beschreibung unserer psychischen Wirklichkeit hält. Eine physikalische Beschreibung wird nie den Duft einer Rose mit beschreiben können, nie die besondere Art von Schmerz an einer Rippe, die gebrochen ist und ins Rippfell sticht, nie die Art, wie Eifersucht oder Migräne peinigt oder wie es sich anfühlt, in eine frische Zitrone zu beißen.

Das werden wir inzwischen zugeben. Fast hat das, was wir eben formuliert haben, auch schon die Kraft eines Arguments. Zu einem philosophischen Argument fehlt unserer Intuition aber der letzte begriffliche Schliff und die Ausarbeitung. Die soll hergestellt werden durch das, was Jackson sein „Knowledge Argument“ nennt: Wer ein Quale empfunden hat, *weiß* mehr, als wer nur die funktionale Rolle dieses Ereignisses im biologisch-neurologischen Gesamt seines Organismus kennt.

Das „Wissens-Argument“ wird zunächst an zwei Beispielen illustriert und ausgewertet (I). Dann wird es gegen das modale (II) und gegen Nagels „What is it like to be“-Argument abgegrenzt (III). Schließlich wird (IV.) gefragt, ob qualitative Zustände auch bestehen würden, wenn sie keine funktionalen Zustände eines hoch differenzierten Computers wären. Der Vergleich mit II und III wird abgewiesen, IV – wie wir schon vorweggenommen haben – wird bejaht. Unter einer epiphänomenalistischen Deutung kann ein Quale bestehen, ohne kausal etwas auszurichten.

Ad I: Die Optik – und auf Subjektseite ihr entsprechend die Farbwahrnehmungs-Physiologie – hat eine Stufung von Rot-Abschattungen vorgenommen, sagen wir: von  $rot_1$  bis zu  $rot_{182}$ . Ein normalgesichtiger Mensch wird viele dieser Nuancen nicht wahrnehmen, d. h. zwischen dem nächstniedrigeren und dem nächsthöheren nicht unterscheiden können. Nehmen wir an, Fred sei eine Person, die diese unerschwingliche Differenzierungsfähigkeit hat. Er kann die Tomaten in einem Korb nach solchen mit dem einen und solche mit den anderen Rot unterscheiden – uns und allen normalsichtigen Menschen ist das unmöglich.

Nun können wir alle möglichen sinnesphysiologischen Experimente mit Fred anstellen: schlimmstenfalls ihm ein Sehzäpfchen wegoperieren, das wir im Verdacht haben, gewisse Lichtwellen im Rotbereich zu unterscheiden, die wir nicht unterscheiden können. Wir können aus seinem Verhalten auf seine Diskriminationsfähigkeit kluge Schlüsse ziehen; wir könnten fordern (oder besser: Fred könnte in einer Verfügung festsetzen), dass Freds Hirn nach seinem Tode zu wissenschaftli-

chen Experimenten zur Verfügung gestellt oder dass sein optisches System einem lebenden Menschen implantiert werden soll:

Again the fine detail doesn't matter. The important point is that such a happening would create enormous interest. People would say „At last we will know what it is like to see the extra color, at last we will come to know how Fred has differed from us in the way he has struggled to tell us about for so long.“ Then it cannot be that we knew all along all about Fred. But *ex hypothesi* we did know all along everything about Fred that features on the physicalist scheme; hence the physicalist scheme leaves something out.

Put it this way. *After* the operation, we will know *more* about Fred and specially about his color experiences. But beforehand we had all the physical information we could desire about his body and brain, und indeed everything that has ever featured on physicalist accounts of mind and consciousness. Hence there is more to know than all that. Hence Physicalism is incomplete (471).

Sie haben den Punkt bestimmt gleich verstanden: *Wenn* wir – nach der Transplantation – mehr über den Rot<sub>1</sub>- und Rot<sub>2</sub>-Unterschied wissen als vorher, so darum, weil wir ihn unmittelbar aus qualitativem Erleben kennen. Wir kennen ihn *nicht* aus physiologischen Informationen. Also ist einige Kenntnis nicht physisch. (Und im Gegensatz wäre er dem Physikalisten unerlaubt, ihr den Wissens-Charakter abzusprechen.)

Berühmter ist das zweite Gedankenexperiment zugunsten des Wissens-Arguments: Mary sei eine ausgezeichnete Sinnesphysiologin und verfüge über alle derzeit erschwinglichen neurobiologischen Informationen über das Farbsehen. Sie kennt alle physikalischen Tatsachen über uns und unsere Umwelt – in einem weiten Sinn von ‚physikalisch‘, der auch Chemie und Neurophysiologie einschließt. Sie weiß alles über kausale und relationale Tatsachen, die uns beeinflussen, funktionale Rollen natürlich eingeschlossen. Durch einen Zufall (der eher die Genfer Menschenrechts-Kommission zu interessieren hätte..., aber Philosophen sind cool) muss sie aber von ihrer Bewusstwerdung an all ihre Forschungen in einem schwarz-weißen Raum durchführen und auf einem Schwarz-weiß-Bild-

schirm visualisieren. Wird sie nun aus diesem Raum befreit und kann mit einem Farbmonitor arbeiten, so wird sie offenbar etwas über das Seherleben *gelernt* haben. Mithin muss ihre frühere Kenntnis unvollständig gewesen sein. Nun hatte Mary *alle* physikalisch erschwingliche Information. „Also gibt es mehr als all das, und der Physikalismus ist falsch“ (471<sub>4</sub>).

Wieder legt Jackson den Finger auf den entscheidenden Punkt: Wenn Mary zum erstenmal ein Rot-Quale erlebt, *lernt* sie etwas, das sie vorher noch nicht wusste. Sie wusste aber alles über die Tatsachen der Wahrnehmung. Also gehört das, was sie nun neu weiß, nicht zum Gesamt dessen, was ihr an physikalischen Wissen zu gebote stand.

What will happen when Mary is released from her black and white room or is given a color vision monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and Physicalism is false (471<sub>5</sub>).

An anderer Stelle schreibt Jackson:

Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world,  $W(P)$ , for which it is complete, and the difference must be in nonphysical facts; for our world and  $W(P)$  agree in all matters physical. Hence, physicalism would be false at our world [though contingently so, for it would be true at  $W(P)$ ] (Frank Jackson, „What Mary Didn't Know“, in: Rosenthal 1991, 392-94, hier: 392).

Ad II. Eine Variante des *Knowledge argument* ist das *modale* [erklären, was ‚modal‘ heißt]. Es ähnelt dem ‚absent qualia argument‘<sup>7</sup> und besagt, keine, auch nicht die raffinierteste physikalische Information über Marys Neuronenbahnen enthalte

<sup>7</sup> Vgl. Sydney Shoemaker, „Functionalism and Qualia“, in: *Philosophical Studies*, 27 (1975), 291-315; wiederabgedruckt in: ders., *Identity, Cause, and Mind*, Cambridge University Press, 1984, 184-205. (Auch noch als Funktionalist gibt Shoemaker übrigens zu, daß wir einen ‚direkten Zugang zum qualitativen Charakter von Bewusstsein haben, nämlich durch Introspektion‘ [189].)

*logisch* die Information, dass Mary irgend ein Bewusstsein habe (etwas spüre). Also gibt es eine mögliche Welt, in der Organismen, die uns in allen relevanten physischen Aspekten gleichen, dieselben physiologischen Zustände instantiiieren wie wir, aber – anders als wir – nichts dabei erleben: die Welt der Zombies, mit der wir schon bekannt sind. Da dies genau die Welt ist, von der der Physikalismus (in allen seinen Varianten, auch als Funktionalismus) redet und wir aus unmittelbarer Vertrautheit untrüglich *wissen*, dass wir etwas fühlen, muss der Physikalismus falsch (oder besser: unvollständig) sein: Er liefert jedenfalls keine Erklärung unserer aktuellen Welt (Jackson 1990, 472<sub>2/3</sub>).

Aber so ähnlich das modale Argument dem von Jackson vorgetragenen ist, er will es nicht übernehmen. Es gebe Bedenken gegen modale Gedankenexperimente und den Gedanken, es gegen in anderen möglichen Welten Doppelgänger, die uns in allen relevanten Eigenschaften gleich, aber – anders als wir – kein Bewusstsein haben. Wenn auch nur ein Philosoph solche Bedenken hege, wolle er (Jackson) das Modell nicht übernehmen. – Das brauche er aber auch gar nicht. Für ihn genüge es völlig, dass in dieser unserer aktuellen Welt der Physikalismus Fakten unberücksichtigt lassen müssen, die gewiss bestehen, nämlich dass wir Qualia haben. Das Modalargument reagiere auf diese Konsequenz nur mit schwereren Geschützen.

Ad III. Auch mit Nagels What-is-it-likeness-Argument möchte Jackson sein Wissens-Argument nicht verwechselt sehen. Was er darüber sagt, klingt etwas spitzfindig (472 f.).

Zunächst ähneln sich die beiden Argumente ja sehr. Auch Nagel zieht anti-physikalistische Argumente aus der Tatsache, dass wir die Welt aus einer subjektiven Perspektive erleben; auch Nagel nennt den Physikalismus unvollständig.

Jackson scheint aber zu meinen, sein Wissens-Argument arbeite nicht mit einer subjektiven Perspektive, mit dem immer auch ein Ich ins Spiel komme. Es gehe nicht darum, dass ich mich nicht in Freds Ich-Perspektive (oder in die einer Fledermaus) versetzen könne. Vielmehr sei der Punkt, dass noch so reichhaltiges physisch-physiologisches Wissen mir nicht dazu verhelfen kann, was Fred *weiß*, wenn er rot<sub>1</sub> von rot<sub>2</sub> unterscheiden kann (473<sub>2</sub>).

No amount of knowledge about Fred, be it physical or not. Amounts to knowledge „from the inside“ considering Fred. We are not Fred. There is thus a whole set of item of knowledge expressed by forms of words like ‚that is *I myself*, who is ...‘ which Fred has and we simply cannot have because we are not him (l. c.).

Qualitatives Wissen ist also von privilegiertem Selbst- oder Ich-Wissen unterschieden. Besonderheiten der *he, himself locution* haben wir schon kurz im Zusammenhang von Frau Bakers Kritik an Rosenthals *higher-order*-Modell bemerkt (4. Vorlesung vom 15. Mai). Sie werden der Gegenstand der kommenden Vorlesung sein.

Hier darf genügen, dass Jackson das Wissen, das in Form qualitativen Sich-Anfühlers vorliegt, für anonym hält und vom Wissen um ein (ebenfalls physikalistisches) Ich unterschieden wissen will. Ich zweifle sehr, dass dieser Einwand Nagel wirklich trifft, zumal er auch nicht von einem Ich-Wissen sondern von einem phänomenalen *Bewusstsein* handelt. Aber der Punkt bleibt interessant: Phänomenales Bewusstsein scheint im Gegensatz zu Ich-Wissen zu stehen, es ist nicht egologisch strukturiert.

Für Block war die Ichlosigkeit sogar ein Kriterium für *P*-Bewusstsein. Er könnte etwas Ähnliches wie Jackson im Sinn haben, wenn er schreibt:

P-conscious states often seem to have a „me-ishness“ about them, the phenomenal content often represents the state as a state of me. But this fact does not at all suggest that we can reduce P-consciousness to self-consciousness [Block meint: *self-knowledge*], since such „me-ishness“ is the same in states whose P-consciousness is different. For example, the experiences as of red is the same as the experience as

of green in self-orientation, but the two states are different in phenomenal feel (Block 1997, 390, linke Spalte).

Man kann Jackson Einwand vielleicht auch so wiedergeben: Zwei Dinge dürfen nicht verwechselt werden: 1. Marys *Wissen* über den Unterschied, wie es ist, die Welt in Schwarz und Weiß und dieselbe Welt in Farben zu sehen, und 2. zu wissen, dass *sie selbst* es ist, die diese Erfahrung macht. Qualia seien an sich nicht personalisiert: sie bestehen in einem anonymen Spüren, das seinerseits in einer unmittelbaren und untrüglichen Vertrautheit gegeben ist, nicht aber in der Kenntnis eines Subjekts als Träger dieses Spürens. Das für uns Entscheidende ist aber gleichwohl, dass Jackson *beide* Phänome für Instantiierungen von ‚Wissen (knowledge)‘ hält. Darüber hat es bis heute anhaltende Diskussionen in der *Philosophy of Mind* gegeben, mit denen ich meinen Bericht nicht komplizieren will.

Ad IV. Bleibt die Frage, ob dieser Befund den Epiphänomenalismus als theoretische Option unausweichlich mache. Jackson ist vorsichtig. Er sagte: er mache ihn jedenfalls möglich. Wenn wir diese Konsequenz gegen den Physikalismus ziehen, geben wir es auf, Bewusstseins-Zuständen irgendeine Art kausaler Wirksamkeit zuzuschreiben („qualia are causally impotent with respect to the physical world“ [Jackson 1990, 473 u.]). Qualia verändern die physische Welt nicht. Wenn sie vorliegen, mögen sie uns dazu verhelfen, sie von anderen mentalen Zuständen zu unterscheiden, aber von nichts Physischen (473 f.) Dagegen empört sich natürlich unsere vermeinte Erfahrung von der Wirksamkeit wenigstens eines Teils unserer bewussten Zustände.

Jackson geht also abschließend die Standard-Einwände gegen den Epiphänomenalismus durch.

Ein erster besagt: Schmerzen verursachen typischerweise Schmerzverhalten. Habe ich Kopfschmerzen, so halte ich diesen Zustand dafür verantwortlich, das ich (zum Beispiel) ein Schmerzmittel nehme oder wie ein Tiger im Käfig im Zimmer hin- und herrenne und mir die Schläfen halte. Dazu sagt Jackson: Hume habe gezeigt, dass erwartete Abfolgen zwischen Zuständen keine Kausalität implizieren [erklären]; und außerdem könne der Epiphänomenalist leicht sagen, ein bestimmter Hirnprozess habe eben diese beiden Konsequenzen: Schmerzempfinden und Schmerzverhalten. – Diese Auskunft wird uns vermutlich nicht sehr stark beeindrucken.

Die zweite Intuition gegen den Epiphänomenalismus bemüht Darwins Evolutionstheorie. Wie wären Qualia in in der Physiogenese unserer Spezies aufgekommen, wenn sie keinen überlebensnotwendigen Zweck hätten? Jackson reagiert darauf cool: „Epiphenomenal qualia are totally irrelevant to survival“ (475<sub>6</sub>). Wieso haben sie sich dann ausgebildet? Jacksons Antwort ist wieder cool: Auch der dicke Pelz und die Fettschicht drunter sind dem Eisbären überlebensnotwendig – aber das Gewicht dieses Fells ist ein Überlebensnachteil. Um den Eisbären gegen die Kälte zu schützen, musste der Nachteil der Fettleibigkeit als „unavoidable concomitant“, als unvermeidliches Nebenprodukt in Kauf genommen werden. Es hat keine überlebensschützende Funktion – so wenig wie qualitatives Bewusstsein (474 u.).

Es ist drittens eingewandt worden, dass unser Wissen um Qualia bei anderen Personen dem Epiphänomenalismus Kopfzerbrechen bereiten müsse. Denn gilt hier nicht wieder, was für Mary gilt, dass jede beliebige physische (aus Verhaltensbeobachtung gewonnene) Information ungenügend ist zur Erklärung von Qualia? Erst recht muss der Phänomenalist annehmen, dass Qualia das Verhalten einer fremden Person nicht erklären können.

Jackson erwähnt – ich weiß nicht, ob ernsthaft oder ironisch – die klassische Ausflucht der Dualisten: den Interaktionismus. Der nimmt an, nicht-physikalische Eigenschaften beeinflussen das Verhalten kausal. Da sie das aber nicht direkt tun können, muss ein Parallelismus zwischen physischer und nicht-physischer Verursachung angenommen werden. So etwas kann man aber nur annehmen, wenn man nicht daran glaubt, dass unsere physische Welt kausal geschlossen ist [erklären!] – also dass es noch andere Weisen der Verursachung gebe als die physische. Mentale Verursachung – also Verursachung durch Qualia – wäre dann eine Art Überdetermination, weil der Effekt auch ‚von alleine‘, ohne die Mitwirkung von Qualia, eingetreten wäre.

Ein viertes Problem des Epiphänomenalismus erwähnt Jackson nicht in diesem Aufsatz (sondern in Mitchell & Jackson 1996, 7). Wie lässt sich nämlich Gedächtnis erklären, wenn es keine kausale Spur im Bewusstsein gibt, die auf Vergangenes zurückführt? Wenn ich mich an meinen Urlaub im North Cascades National Park erinnere, dann darum, weil, was ich dort sah, hörte und erlebte, kausale Wirkung auf mich hatte. Es hinterlässt in meiner Erinnerung Spuren, die meine Überzeugung, der Urlaub sei oder und so verlaufen, stützen. Epiphänomenale Eigen-

schaften haben aber *per definitionem* keinerlei kausale Kraft, also können sie keinen Spuren hinterlassen und mithin auch nicht erinnert werden.

Sie können sich denken, dass Jacksons provozierende Thesen Widerspruch gefunden haben in der *analytic community*.

Manche, in Wittgensteins Nachfolge, bezweifeln, dass die umstürzend neue Erfahrung, die Mary macht, wenn sie eine reife Tomate bei Tageslicht sieht, den Titel des Wissens verdient. Was wissbar war, das beherrschte sie doch schon vorher. Eine bekannte Neuauflage von Wittgensteins Einwand hat David Lewis formuliert.<sup>8</sup> Er spricht dem Wissen, wie einem zumute ist ('what it is like to'), den Status eines propositionalen (oder auch informationellen) Wissens ('knowing-that') ab, belässt ihm aber den Status eines Fertigkeiten-Wissens ('knowing-how') (Lewis 1988, 516 f.). Das Wissen-wie (es ist, in einem Quale-Zustand zu sein), bestehe in einer Disposition, nämlich in der Fähigkeit ('ability'), 1. sich etwas vorstellen, 2. sich daran erinnern und 3. es gegebenenfalls wiedererkennen zu können (Lewis 1983, 131).

The Ability Hypothesis says that knowing what an experience is like just is the possession of these abilities to remember, imagine, and recognize. It isn't possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities aren't actualized. It isn't knowing-that. It's knowing-how. Therefore it should be no surprise that lessons won't teach you what an experience is like. Lessons impart information; ability is something else. Knowledge-that does not automatically provide know-how.

[...]

---

<sup>8</sup> In: „What Experience Teaches“, in: *Proceedings of the Russellian Society*, Sydney, 1988, wiederabgedruckt in (und zitiert nach): William G. Lycan (Hg.), l.c., 499-519, auch in: „Mad Pain and Martian Pain“, in: D. Lewis, *Philosophical Papers*, Volume I, Oxford University Press, 1983, 122-130, vgl. vor allem das Postscript über "Knowing what it's like" (130-32). Frank Jackson hat auf diese Einwände reagiert in: „What Mary Didn't Know“, in: *The Journal of Philosophy* LXXXIII, 5 (May 1986), 291-95, wiederabgedruckt in (und zitiert nach): David M. Rosenthal (Hg.), *The Nature of Mind*, 392-94.

If the Ability Hypothesis is the correct analysis of knowing what an experience is like, then phenomenal information is an illusion (Lewis 1988, 516 f.).

Man mag diese Unterscheidung *prima facie* triftig finden und bereit sein, die Vertrautheit-mit-einem-Quale vom ersten Wissens-Typ (und damit vielleicht von jederlei kognitiv relevantem Wissen) zu unterscheiden. Gleichwohl muss man Lewis vorwerfen, eine semantische Zweideutigkeit in der Rede vom *Wissen, wie* auszubeden. Auch dispositionelles (oder Regel-)Wissen ist ja ein *Wissen, wie*, nämlich z. B. ein solches, *wie* man es anstellt, eine Knotenschlinge zu legen oder einen Ablativus absolutus aufzulösen. Solches Wissen könnte man ebenfalls durch Implizität, Unmittelbarkeit und Vertrautheit charakterisieren (wenn ich z. B. Deutsch sprechen kann, so bin ich mit der Grammatik irgendwie 'unmittelbar' vertraut und weiß doch keineswegs notwendig explizit die Regeln, denen ich dabei folge; so sprach Monsieur Jourdain Prosa, ohne es zu wissen). Und doch ist solches Regelwissen grundsätzlich verschieden von dem von Nagel und Jackson beschriebenen *Wissen, wie*: Wenn ich Schmerzen habe und mithin weiß, *wie* mir dabei zumute ist, muss ich keine Regel über den Gebrauch des Worts 'Schmerz' (oder über den Typ oder die kausale Rolle von Schmerz-Empfindungen) kennen (oder implizit beherrschen). Wohl aber gilt das umgekehrte: Ich könnte keine 'Fertigkeit' zum Mir-Vorstellen, Erinnern oder gar Voraussagen von Schmerzen ausbilden, wenn ich mit Schmerzen niemals aus der unmittelbaren Vertrautheit eines Spürens bekannt geworden wäre. Und insofern zieht Lewis' Banalisierung des Knowledge-Arguments nicht.

Darum hatte Jackson gute Gründe, die von Lewis zur Abwehr des Knowledge-Arguments aufgebotene Arbeitsteilung von *Wissen, wie* und *Wissen, dass* zu verwerfen (Jackson, 1986, 393). Wenn Mary aus ihrem Schwarz-Weiß-Verließ heraustritt und eine reife Tomate unter Normalbedingungen sieht, dann hat sie nicht eine neue Wissens-Weise erworben. Sie hat nun vielmehr ein Wissen von *etwas, das sie*

früher nicht kannte, also sehr wohl ein *Wissen, dass* (etwas so und so aussieht). Nach Lewis aber besaß sie schon früher jedes denkbare (physikalische) Wissen, dass (und ein anderes Wissen gibt es *ex hypothesi* nicht). Es gebrach Mary lediglich an einer Fähigkeit: einer Art und Weise des Vorstellen-Könnens.<sup>9</sup> Und gegen diese Deutung ist erneut zu bemerken, dass Marys 'Vorstellungs-Vermögen (representational abilities)' vor und nach der Befreiung aus dem Schwarz-Weiß-Raum unverändert geblieben ist. Was sich verändert hat, ist der Umstand, dass Mary ein vorher unbekanntes ‚Tatsachen-Wissen (factual knowledge)‘ dazu erworben hat (l. c., 394 links oben).

Es gibt andere Einwände. Einer der sophisticatedesten ist von Brian Loar vorgetragen worden („Phenomenal States“, in: *Philosophical Perspectives* 4, 81-108). Er lehnt den Wasser-H<sub>2</sub>O-Vergleich als zu grob ab, weil die beiden Ausdrücke verschiedene Wahrheitsbedingungen und mithin verschiedene Frege-Sinne haben („primary intensions“, kurz erklären!). Ein und derselbe Gegenstand kann uns unter verschiedenen Eigenschaften gegeben sein (oder – komplizierter – eine und dieselbe Eigenschaft kann über verschiedene Prädikatsausdrücken ‚eingeführt‘ worden sein) – mit der Folge, dass ihre Selbigkeit nicht *a priori einsichtig* ist. In diesem Sinne erwirbt Mary allerdings ein neues Wissen (Chalmers 1996, 142 f.).

Darauf entgegnet Chalmers: Wie können zwei Frege-Sinne zusammenfallen, ohne dass wir das *a priori* wissen könnten (143<sub>2</sub>)? Wir werden noch in einer kommenden Sitzung (der 11./*de facto* 9. Vorlesung), wenn wir Kripkes berühmte Position

---

<sup>9</sup> „[...] a physicalist can admit that Mary acquires something very significant of a knowledge kind - which can hardly be denied - without admitting that this shows that her earlier factual knowledge is defective. She knew all *that* there was to know about the experiences of others beforehand, but lacked an ability until after her release“ (Jackson, 1986, 393 rechts unten).

über aposteriorische (metaphysische) Notwendigkeit und strikte apriorische Identität kennen lernen werden, auf die Disanalogie zwischen notwendiger Identität zwischen Abendstern und Morgenstern oder Wasser mit H<sub>2</sub>O bei gleichzeitiger logischer Nicht-Notwendigkeit der Identifikation von C-Faser-Reizungen mit Schmerzen diskutieren werden.

Like that objection, Loar's position requires that a conditional from physical facts to phenomenal facts be metaphysically necessary despite being logically contingent, where the gap cannot be explained by a difference in primary intensions. Like that objection, Loar's position requires a brute and arbitrary restriction on possible worlds. Loar offers no argument for this restriction, and his position is subject to precisely the same criticisms (l. c., [die Mehrere-Welten-Theorie und die Semantik andeutungsweise erklären, oder Zitat weglassen!]).

Hören wir noch einen Einwand. Michal hat zwei Grade von Feinkörnigkeit einer physikalischen Untersuchung unterschieden. Nach dem ersten unterscheide ich Wasser von H<sub>2</sub>O und Toms Schlaf von seinem um 2 Uhr morgens an einem Dienstag schlafend angetroffen worden Sein. Das sind zwei verschiedene Repräsentationsweisen und mithin auch zwei verschiedene Konzeptualisierungen desselben. Nachdem ersten ist Wasser eben H<sub>2</sub>O, und Tom schläft eben.

So, meint Tye, kann ich auch mit Marys neuer Erfahrung umgehen. Einmal liegt sie unbegrifflich vor, das zweite Mal wird sie angemessen auf den Begriff gebracht. Hinreichend feinkörnig ausbuchstabiert, scheint Mary nicht zu kennen, was sie erlebt, wenn sie erstmals eine rote Tomate sieht. Aber prinzipiell kann ihr als Wahrnehmungsexpertin diese feinkörnige Konzeptualisierung zugemutet werden (Michael Tye, *Ten Problems of Consciousness. A Representational Theory of the Phenomenal Mind*, Cambridge/MA: The MIT Press, 1995, 171-74). Man kann eben ein und dasselbe Ding unter zwei Darstellungsmodi auffassen, gleichsam physikalisch und mental:

If one knows a FACT under one mode of presentation and one does not know it under another; then one's not knowing that it is the same / FACT that corresponds to each mode of presentation demands that one not know some further substantive FACT, under any mode of presentation (175 f.).

Das hört sich schön und gut an. Die kürzere Version wäre: Es stimmt, dass Mary, wenn sie erstmals Rot sieht, etwas Neues lernt; aber dies Neue ist keine neue *Tatsache*, sondern dieselbe Tatsache unter einem neuen Blickwinkel („under a new mode of presentation“). Aber wir Bewusstseinsfreaks wollen erklärt bekommen, wie wir auch der Selbigkeit der Tatsache unter verschiedenen Präsentationsweisen *gewiss* sein können. Wir sähen gerne einen A-priori-Übergang zwischen beiden Tatsachen. Und das war Jacksons Punkt.

Gewöhnlich wird dieser Wunsch von Physikalisten mit Kopfschütteln quittiert. Warum, so sagen sie, geht es denn in euer Spatzenhirn nicht hinein, dass manche Gegenstände sich eben unter zwei Präsentationsweisen zeigen, unsere bekannte Flüssigkeit etwa als Wasser oder als H<sub>2</sub>O? Aber unsere Intuition sperrt sich dagegen: Wasser und H<sub>2</sub>O gehören beide zur äußeren Welt, die Bedeutungen der beiden Ausdrücke liegen auf derselben Ebene. Vergleichen wir hingegen eine C-Faser-Reizung mit unserem Schmerzempfinden, scheint uns das letztere auf eine andere, intimere Weise gegeben als das erstere. Ja, wir möchten Jackson gern folgen und sagen: Nicht *ein* Faktum zeige sich auf zweierlei Art, sondern: zweierlei Tatsachen werden uns gezeigt. Die Bedeutung ist irgendwie ‚fühlbarer‘ als die der ersten. Darum glauben wir, man führe uns mit einer falschen Analogie aufs Glatt-eis.

Für dieses unser Gefühl hat niemand soviel Verständnis gezeigt wie David Chalmers. In seinem Hauptwerk *The Conscious Mind* beschäftigt er sich ausgiebig mit

den Einwänden, die Jacksons *knowledge argument* von seinen Gegnern erfahren hat (Oxford University Press, 1996, 140 ff.) und schreibt im Blick auf das Unbehagen, das wir gerade artikuliert haben:

We can also put the point a more direct way. Whenever one knows a fact under a mode of representation but not under another, there will always be a *different* fact that one lacks knowledge of – a fact that connects the two modes of presentation (l. c., 141,3).

It follows that if Mary gains any factual knowledge that she previously lacked – even if it is only knowledge of an old fact under a different mode of presentation – then there must be some truly novel fact that she gains knowledge of. In particular, she must come to know a new fact *involving* that mode of presentation. Given that she already knew all the physical facts, it follows that materialism is false. The physical facts are in no sense exhaustive (l. c., 142,3).

Ganz genau dies war auch Colin MacGinnns Punkt. In einem berühmten Aufsatz: „Can We Solve the Mind-Body Problem?“ hat er von einer ‚cognitive closure‘ gesprochen. Sie besteht darin, dass ich das Mittelglied, das Neuronales und Mentales verknüpft, aus Gründen, die prinzipieller Natur sind, nicht selbst wieder in meinem Bewusstsein darstellen kann (*The Problem of Consciousness*, Oxford: Basil Blackwell, 1991, 3).

Let me introduce the idea of *cognitive closure*. A type of mind *M* is cognitively closed with respect to a property *P* (or theory *T*) if and only if the concept-forming procedures at *M*'s disposal cannot extend to grasp of *P* (or an understanding of *T*).

Das aber erhoffen sich unsere kühnsten Intuitionen von einer dieses Namens würdigen Bewusstseins-Theorie. Wir sind intuitiv nicht zufrieden mit einer Erklärung, die sich als Schluss auf die beste Theorie darstellt und uns den mentalen Zugang zu uns selbst nur als einen möglichen ‚mode of representation‘ darstellt neben dem anderen: dem neurobiologischen, der uns nicht in gleicher Weise einsichtig ist. Wir würden gern den neurobiologischen Zugang ebenso evident finden wie den ‚inneren‘.

David Chalmers kommt seinem Kollegen und Auch-Bewusstseinfreak Jackson jedenfalls argumentativ zu Hilfe:

There have been many replies to the [knowldege] argument; I will discuss them in the next chapter, where materialism rather than reductive explanation will be at issue. But for now it is interesting to note that most of the objections to the argument against materialism have *conceded* the point that is relevant to the argument against reductive explanation: that knowledge of what red is like is factual knowledge that is not entailed *a priori* by knowledge of the psycical facts. The only way that the conclusion can be evaded is to deny that knowing what red experience is like gives knowledge of a *fact* at all. This is the strategy taken by Lewis (1990) and Nemirow (1990), who argue that all Mary is lacking is an *ability*, such as the ability to recognize red things. I discuss this suggestion in the next chapter; here, I simply note that insofar as it seems clear that when she sees red for the first time, Mary is *discovering* something about the way the world is, it seems clear, that the knowledge she is gaining is knowledge of a fact (*The Conscious Mind. In Search of a Fundamental Theory*, Oxford University Press, 1996, 104).

Dies nächste Kapitel lohnt es sich absolut zu lesen, zumal es weiteres Wasser auf Jacksons Mühlen leitet. Aber in dieser Vorlesung müssen wir das nicht tun.

## 9. (de facto: 7.) Vorlesung vom 12. Juni

Roderick Chisholms *The First Person* ist drei Denkanstößen besonders verpflichtet: 1. der Brentanoschen Theorie unmittelbarer Evidenz und 2., eng mit ihr verbunden, der des inneren Bewusstseins, 3. einer Ende der 70er Jahre von David Lewis vorgestellten Theorie, derzufolge epistemische Selbstzuschreibungen nicht als propositional gelten dürfen (auch diese Theorie geht letztlich auf Brentano zurück).

1. Das Interesse am Phänomen der unmittelbaren Evidenz hat zu tun mit Chisholms fundamentalistischen Neigungen [Wort erklären!]. Gäbe es kein Wissen, das unabhängig von seiner Verwiesenheit auf anderes bestünde, so verlöre sich jede Begründung von Wissens-Ansprüchen in einem infiniten Regress. So hatte schon Sextus Empiricus unterschieden zwischen Gegenständen, die durch sich selbst, und solchen, die durch andere (Gegenstände) wahrgenommen werden.<sup>10</sup> Solche, „die durch sich selbst wahrgenommen erscheinen“, können, falls es sie gibt, als Ausgangspunkt des Philosophierens angenommen werden. Sie sind 'durch sich selbst bekannt (oder von selbst einsichtig)' - und das meint ja ungefähr der lateinische Ausdruck 'evident'. Für eine in genau diesem Sinne unmittelbar evidente Erfahrung hielt Leibniz das Selbstbewusstsein: Bei ihm handelt sich's nicht um eine apriorische Vernunft-Wahrheit, sondern um eine Wahrnehmung a posteriori, die aber ihrer Unmittelbarkeit wegen weder beweisbar noch bezwei-

---

<sup>10</sup> *Grundriß der Pyrrhonischen Skepsis*. Eingel. und übersetzt von Martin Hossenfelder, Frankfurt/M.: Suhrkamp (stw 499), 1985, I. Buch, 6. Kap.

felbar (also "evident") sei; denn die innere Erfahrung ('aperception') besteht in der "immédiation entre l'entendement et son objet".<sup>11</sup>

In Brentanos *Kategorienlehre* (hg. von Alfred Kastil, Leipzig 1933, verbesserter Nachdruck Hamburg 1968, 165) ist im gleichen Sinne die alte aristotelische Unterscheidung zwischen einer Kenntnis per se und einer solchen per accidens aufgegriffen; und es wird dann gesagt (hier erfolgt der Übergang von 1. zu 2.), dass Selbstbewusstsein (und *nur* Selbstbewusstsein) ein durch und von sich selbst Bekanntes sei. (Wenn ich mich evident auf mich beziehe, muss ich meine Akzidentien: wann geboren, wie beschaffen, wie aus einer Fremd-Perspektive – perzeptiv – mich anderen präsentierend, nicht kennen. Das bedeutet nicht, wie etwa die Strawsonianer einwenden würden, dass sich die denkende Substanz in der inneren Erfahrung unspezifiziert gegeben sei: sie erfährt *etwas*, und zwar nicht irgend etwas. Aber das Erfahrene ist doch nicht nach Zeit und Ort, Gestalt und Ausdehnung individuiert: Es bildet buchstäblich ein Subjekt ohne Eigenschaften. Wenn Descartes *ego cogito, ego sum* dachte, meinte er nicht: „René Descartes denkt gerade“ [l. c., 158; vgl. *Psychologie vom empirischen Standpunkt*, Bd. III, Hamburg: Meiner, 2<sup>1968</sup>, 98].) So ist, was mir evident – Brentano sagt auch 'modo recto' – präsentiert wird, nur das auffassende Bewusstsein selbst; an dem transzendenten Objekt kann ich grundsätzlich zweifeln: es ist mir nicht per se, sondern nur 'modo obliquo' gegeben. Das sind terminologische Unterscheidungen aus Brentanos späterer Schaffenszeit, wie sie sich zumal im Zweiten (Leipzig 1925, Nachdruck Hamburg 1971, "Anhang" 133 ff.) und Dritten Band (37 ff.) der *Psychologie vom empirischen Standpunkt* und in den Nachlassnotaten der sogenannten *Deskripti-*

---

<sup>11</sup> *Nouveaux Essais sur l'Entedement Humain*, Paris (Garnier-Flammarion) 1966, Livre IV, Chapitre IX, 383 f.

*ve[n] Psychologie* (hg. von Roderick Chisholm und Wilhelm Baumgartner, Hamburg 1982) niederschlagen. Dort etwa sagt Brentano zur "Erläuterung des Ausdrucks Objekt: etwas innerlich Gegenständliches ist gemeint. Draußen braucht ihm nichts zu entsprechen" (22). Im 3. Band der *Psychologie [...]* ist Brentano am deutlichsten:

Über uns selbst als psychisch Tätige hinaus haben wir „[...] keine unmittelbar evidente Erkenntnis einer Tatsache“ (4, im Orig. gesperrt). Und: „Es ist sicher, daß weder wir noch irgendein anderes Wesen, welches etwas unmittelbar mit Evidenz als Tatsache erfaßt, dabei etwas anderes als sich selbst zum Objekt der Erkenntnis haben kann. Denn wie immer die Erkenntnis eine bloß tatsächliche und keine Erkenntnis als notwendig sein mag, so muß doch die relative Unmöglichkeit *g e s i - c h e r t s e i n*, daß der psychisch Tätige, so wie er ist, sei und das Objekt der Tätigkeit nicht sei. Und so ist es bei der Selbsterkenntnis, nicht aber bei der Erkenntnis eines anderen, als einfach Tatsächlichen“ (l. c., 6 f.).

Die Nicht-Evidenz der Objekt-Gegebenheit betrifft seine Existenz, komparativ (durch ein 'Relativ' oder in prädikativer Aussageform) ausgemachte Eigenschaften, blinde Urteile und dergleichen. In evidenten Urteilen wird aber, nach Brentanos Überzeugung, nicht etwas (ein Subjektterm) auf etwas (ein Prädikat) bezogen, sondern eine Vorstellung wird geradehin 'anerkannt' oder 'gesetzt'. Darum kann sich in solche Urteile auch kein Irrtum einschleichen. Sie haben einfach nicht die Propositional-Struktur des 'etwas als etwas'. So erklärt sich, wie der Begründungsregress (der bestehende Sachverhalte auf andere bestehende Sachverhalte – Propositionen – als ihre Ursachen zurückführt) in einer Evidenz seinen Abschluss finden kann, in dem ein Bewusstsein sich 'basal' – ohne Verweis auf ein anderes – durchsichtig ist. Die vollständige Struktur der Intentionalität wäre dann (am Beispiel des Hörens), dass eine Vorstellung – Brentano nennt sie primär – sich aufs transzendente Objekt richtet und ihrerseits sich selbst durchsichtig ist (Brentano spricht ungeschickterweise vom Gegenstand dieses in Wahrheit ungegenständlichen, der Intention einwohnenden Bewusstseins als vom 'sekundären

Objekt'). Nur dies sekundäre Bewusstsein ist 'modo recto' (oder 'evident') gegeben, das bewusste primäre Objekt (dessen ontologischen Zustand als innermental oder transzendent Brentano in merkwürdiger Zweideutigkeit belässt) ist 'modo obliquo' gegeben und in einigen Rücksichten nur wahrscheinlich.

2. Das ist auch Chisholms Überzeugung. Wie Brentano glaubt er, Wissen ultimativ nur fundieren zu können in einem Bewusstsein, das sich 'per se' oder 'modo recto' (in 'innerer Wahrnehmung') bekannt ist und nicht die relationale Struktur einer Proposition hat. Chisholm führt den Begründungs-Regress in folgender Variante vor: Gewöhnlich halten wir Meinungen (z.B. die, dass mein Kind Fieber hat) für gerechtfertigt, wenn wir als Grund dafür eine „evidence“ – deutsch würden wir sagen: einen Beleg – anführen können – sagen wir, dass das Fieberthermometer 39° anzeigt.<sup>12</sup> Auf die Weise verlagern wir aber nur "die Last der Rechtfertigung von einem Anspruch auf einen anderen" - und könnten auf die Weise endlos fortfahren, ohne an einen Schlusspunkt zu gelangen. (Woher weiß ich, dass der Thermometer funktioniert? Weil ich ihn vorher getestet habe. Woher weiß ich, daß die Meßskala korrekt ist? usw.) Der Regress wäre nur zu stoppen, wenn irgendwo eine Proposition auftauchte, die unmittelbar-sich-selbst-verbürgend wäre, also folgende Form hätte:

What justifies me in thinking that I know that *a* is *F* is simply the fact that *a* is *F* (Roderick M. Chisholm, *Theory of Knowledge*, Second Edition, Englewood Cliffs/NJ: Prentice Hall, 1977, 21).

Da wir aber bestehende Sachverhalte nicht schon für Instantiierungen von Wissen halten (es gibt so viele Tatsachen, die wir nicht kennen), muss dieser sich-selbst-evidenzierende Sachverhalt *ins Wissen selbst* fallen, anders gesagt: unmittelbar

evident sein („directly evident“). Chisholm nennt Zustände, die nur ineins mit ihrer Bekanntheit vorliegen, mit Brentanos Schüler Alexius Meinong "selbst-präsentierend" (l. c., 20 ff.). Und dann ist klar, dass es sich bei ihnen nur um mentale Zustände handeln kann, z. B. um Denken oder Glauben: um solche, die Descartes 'cogitationes', Leibniz 'perceptions' und Kant und Brentano 'Vorstellungen' genannt hatten. Denn nur für sie gilt, dass sie nur dann bestehen können, wenn dieses Bestehen – wie auch Meinong sagt – "durch sich selbst erfaßt" wird.<sup>13</sup> Das war mit der Formulierung gemeint, die Evidenz-Regel könne den Begründungsregress nur stoppen, wenn der Sachverhalt, der als Begründung dient, selbst ins Wissen fällt: "Was mich berechtigt zu denken, ich wüsste, dass *a F* ist, ist einfach die Tatsache, daß *a F* ist", das heißt: ein Fall sich selbst präsentierenden Wissen ist. Man wird sagen, dass hier die Begründung einfach nur die zu begründende Proposition wiederholt, also unangemessen sei. Tatsächlich entspricht diese Zirkularität genau der geforderten *Unbedingtheit* und *Unmittelbarkeit* der Selbst-Evidenz, die sich auflösen würde, wenn Analysans und Analysandum auf zwei verschiedene Propositionen verteilt wären.

So können wir nun die Natur der Selbstpräsentation genauer charakterisieren. Sie besteht darin, dass das Bestehen (das Esse) mentaler Zustände (etwa "verliebt zu sein scheinen") eine notwendige und zureichende Bedingung darin hat, dass diese Zustände dem Bewusstsein bekannt (oder evident) sind ("esse est percipi") – und das gilt *nur* für das Sein des Bewusstseins selbst; für alles andere Sein würden wir

---

<sup>12</sup> Roderick M. Chisholm, „Das unmittelbar Evidente“, in: Peter Bieri (Hg.), *Analytische Philosophie des Geistes*, Königstein/Ts.: Hain, 1981, 191. Ich zitiere im Folgenden nach der zweiten, stark bearbeiteten Auflage von Chisholms *Theory of Knowledge*.

<sup>13</sup> Alexius Meinong, „Über emotionale Präsentation“, in: A.M., *Gesamtausgabe*, hg. von Rudolf Haller und Rudolf Kindinger, Bd. 3, Graz 1968, S. 285-465, § 1. Ferner Thomas J. Steel, „Knowledge and the Self-Presenting“, in: Keith Lehrer (Hg.), *Analysis and Metaphysics*, Dordrecht 1975, 145-150.

nicht sagen, dass es sein Maß am auffassenden Bewusstsein hat. Chisholm gibt zwei Definitionsvorschläge für selbst-präsentierende Kenntnisse:

Let us try now to characterize self-presentation more exactly. If seeming to have a headache is a state of affairs that is self-presenting for *S* at the present moment, then *S* does now seem to have a headache [*h*] and, moreover, it is evident to him that he seems to have a headache [hier ein Fußnotenverweis auf Meinong]. And so we may formulate the definition that way:

D2.1 *h* is self-presenting for *S* at *t* = Df *h* occurs at *t*; and, necessarily, if *h* occurs at *t*, then *h* is evident for *S* at *t*.

An alternative formulation would be this:

*h* is self-presenting for *S* at *t* = Df *h* is true at *t*; and necessarily, if *h* is true at *t*, then *h* is evident for *S* at *t* (l. c., 22).

3. So führt die Suche nach einem in sich evidenten Grundsatz zur Fundierung des Systems des Wissens auf ein Phänomen, das für den geeigneten Kandidaten auf den Posten des 'obersten Grundsatzes' gelten kann. Bei Descartes, Kant und Fichte war es das Selbstbewusstsein. Chisholm spricht vom 'self-presenting knowledge'. Zwar hält auch er es für eine notwendige Voraussetzung für Wissen (und Wissenschaft) . Dagegen scheint er den vermessenen Anspruch aufzugeben, Selbstbewusstsein sei ein Deduktionsprinzip aller wahren Sätze (*propositions*). Chisholm hält Selbstbewusstsein nämlich – wie Brentano – überhaupt nicht für ein propositionales Phänomen. Und da nur aus Sätzen Sätze abgeleitet werden können, entsteht nun die Frage, ob dem Selbstbewusstsein dann noch eine fundamentale Rolle bei der Bildung von Theorien verbleibt.<sup>14</sup>

---

<sup>14</sup> Schon Fichtes Selbstbewusstseins-Theorie war an diesem Punkt merkwürdig inkonsistent. Einerseits charakterisiert er es als 'un-mittelbar' (ohne die Unmittelbarkeit wäre es nicht 'aus sich selbst heraus gewiss', andererseits schreibt er ihm die Satz-Form zu, ohne die es nicht als Deduktions-Prinzip des Wissens in Betracht käme. Wie aber könnte ein Deduktions-Prinzip der Philosophie zugleich Grundsatz und unmittelbare (intellektuelle) *Anschauung* sein?

Nun ist bekannt, dass in den letzten 40 Jahren mehrere analytische *Philosophers of Mind* die Fichtesche These von der Unmittelbarkeit des Selbstbewusstseins unterschrieben und sogar mit neuen Argumenten befestigt haben. Sidney Shoemaker hat gezeigt, daß der Subjekt-Gebrauch von 'ich' grundlegend ist für den Objekt-Gebrauch und dass im Subjekt-Gebrauch das Bewusstsein durch kein zweites Element (etwa einen Spiegel, eine Wahrnehmung, eine Körper-Information) mit sich vermittelt ist.<sup>15</sup> Thomas Nagel hatte – wie wir wissen – daraus schon die Konsequenz gezogen, dass die subjektive Komponente mentaler Zustände nicht als Proposition ausgelegt werden könne (Nagel 1979, 171). Wenn ich nämlich in einem mentalen Zustand bin, ist mir bekannt, *wie* es ist ('what it is like'), in ihm zu sein, und zwar auch dann, wenn ich keine Proposition finde, die ich für wahr von diesem Zustand halte. (Wir werden andere Ansichten über dies Phänomen in der übernächsten Vorlesung kennen lernen, wenn wir uns mit Kripkes wahrhaft bahnbrechendem Argument für die cartesianische Position beschäftigen.)

Aufgrund dieser und einiger weiterer Befunde waren seit Anfang der 80er Jahre mehrere Philosophen zu der radikalen Überzeugung gekommen, dass Selbstbewusstsein gar kein Fall von propositionalem Wissen ist. Es stellt sich nicht in Sätzen (der Form: 'Ich weiß, dass ich  $\phi$ ') dar, und in ihm gibt es auch keine zwei getrennten Elemente, die zu identifizieren wären. Ich denke in diesem Zusammenhang vor allem an David Lewis, Hector-Neri Castañeda, John Perry und Roderick Chisholm. Chisholm konzentriert sich auf Lewis; ich werde einen Bericht über Lewis nachtragen, fairnesshalber aber zunächst Castañedas Vorläuferschaft in den

---

<sup>15</sup> „Self-reference and self-awareness“, in: S.S., *Identity, Cause, and Mind. Philosophical Essays*, Cambridge 1984, 6-18 und *Unity of Consciousness and Self-Consciousness*, in: ders. und Richard Swinburne: *Personal Identity*, Oxford 1984, 102-105.

Blick bringen.<sup>16</sup> Von Castañeda lässt sich besonders viel lernen über Eintümlichkeiten der von Chisholm so genannten *he, himself locutions*.

Castañeda hatte seit der Mitte der 60er Jahre die starke These vertreten, dass das ‚ich‘ des Selbstwissens nicht reduzierbar sei auf das, worauf wir mit Namenwörtern, mit Namenwort-Stellvertretern (indexikalischen Ausdrücken aller Art) oder mit Kennzeichnungen Bezug nehmen. (Daraus folgt, dass *he\** – wider den Augenschein – *kein* Pronomen oder anderer indexikalischer und damit kein gegenstandsidentifizierender Ausdruck ist: Castañeda nennt ‚*he\**‘, ‚hier‘ und ‚jetzt‘ *Quasi-Indikatoren*.) mit der *he himself-locution* verweist eine Person auf sich *als* auf sich. Castañeda (und seine Nachfolger) gehen aber weiter und zeigen, dass quasi-indikatorischer Selbstverweis auch nicht auf eine Bezugnahme *de dicto* (also auf ein sprachlich geäußertes Tatsachenwissen) reduziert werden kann (wie es die These des Faktualismus unterstellt).

Beispielsätze: Sie formulieren grundsätzlich intentionale Aussagen, also solche mit einem Hauptsatz, der einer Person ein Intentionsverb zulegt, z. B. : ‚Paul glaubt (denkt, weiß, hofft, wünscht usw.)‘. Der davon abhängige Nebensatz drückt

---

<sup>16</sup> Chisholm hatte der Theorie, mit der wir uns hier beschäftigen, zwei Vorläuferinnen vorausgeschickt, von deren Unhaltbarkeit er sich (oder Dieter Henrich ihn) überzeugt hatte (*The First Person. An Essay on Reference and Intentionality*, Brighton, Sussex: The Harvester Press, 26, Anm. 12: "For a critical account of the relations between the two theories I had formerly held, compare Dieter Henrich, „Zwei Theorien zur Verteidigung von Selbstbewußtsein“, in: *Grazer philosophisches Studien* 7/8, 1979, 77-99."). – Chisholm selbst gesteht die Priorität der Entdeckung von ‚de se-Attributionen‘ David Lewis zu (*FP* 32 und 40, Anm. 2). Vgl. D.L., „Attitudes *De Dicto* und *De Se*, in: *Philosophical Papers*, Vol. I, Oxford 1983, 133-156- Castañeda nimmt beide Positionen (die er unter dem Titel der ‚Selbstzuschreibungs-Ansicht‘ zusammenfaßt) kritisch ins Visier in „Self-Consciousness, Demonstrative Reference, and the Self-Ascription View of Believing“, in: J. Tomberlin (ed.), *Philosophical Perspectives*, 1, *Metaphysics*, Atascadero: Ridgeview 1987, 405-454. John Perry hat seinen anti-propositionalistischen Standpunkt übers Selbstbewusstsein mit ausdrücklicher Referenz auf die forrunner-Rolle von Castañeda entwickelt in: „The Problem of the Essential Indexical“, in: *Nous* XIII (1979), 3-21. Vgl. auch ders.: „Castañeda on He and I“ und Castañedas „Reply to John Perry: Meaning, Belief, and Reference“ in: James E. Tomberlin (ed.), *Agent, Language, and the Structure of the World*, Indianapolis (Hackett) 1983, 15-42 und 313-327. – Ich kann in die *Détails* dieser Texte hier nicht eindringen, habe das aber anders-

dann den ‚propositionalen Gehalt‘ des Intendierten aus. Also etwa: ‚Paul glaubt, dass Marie glücklich ist.‘ Hier haben wir einen so genannten Glauben *de dicto*. Paul glaubt nämlich das *dictum* (das Gesagte, die Proposition), ‚dass Mary glücklich ist‘. Aber Paul kann auch einen Sach-Glauben haben, also (wie die Scholastiker das nannten) einen Glauben *de re*. Dann hat er eben eine Überzeugung über etwas (oder jemanden), wie sie sich etwa in dem Satz artikuliert: ‚Paul glaubt von Marie, dass sie glücklich ist.‘ Schließlich kann Marie aber auch einen Glauben über sich selbst haben und das auch wissen. Dann hat sie einen Glauben *de se* oder ein Selbstwissen: ‚Marie glaubt, dass sie\* (*selbst*) soundso.‘

Ich wechsele die Beispielsätze:

a) Die jüngste Abiturientin von Schriesheim glaubt, dass die jüngste Abiturientin von Schriesheim die nächste Schriesheimer Weinkönigin wird (*de dicto*);

(b) Es gibt ein *x* so, dass *x* mit Marie (nämlich mit der jüngsten Abiturientin von Schriesheim) identisch ist, und *x* wird von *x* für die künftige Weinkönigin gehalten (*de re*);

(c) Die jüngste Abiturientin von Schriesheim glaubt, dass sie selbst die nächste Weinkönigin wird (*de se*).

Um den Beweis der Irreduzibilität von Einstellungen *de se* über die Evidenz des ersten Augenscheins hinaus zu plausibilisieren, ist es nötig, die Implikationsverhältnisse zwischen den einzelnen Formulierungen ausbuchstabieren (Chisholm 1981, 19):

(1) (a) impliziert (b) (also *de dicto* impliziert *de re*): Wenn eine Person, welche die jüngste Abiturientin von Schriesheim ist, das *dictum* glaubt, dass die jüngste Abiturientin von Schriesheim glücklich ist, so gibt es eben jemanden, der mit der jüngsten Abiturientin von Schriesheim identisch ist und von eben dieser für die künftige Weinkönigin gehalten wird. Man könnte befürchten, hier würde unerlaubt aus einem ungeraden (intensionalen) in einen geraden (extensionalen) Kontext übergegangen. Das ist aber nicht der Fall, denn (a) identifiziert ‚die jüngste

---

wo getan, nämlich in: *Selbstbewußtsein und Selbsterkenntnis. Essays zur analytischen Philosophie der Subjektivität*, Stuttgart: Reclam, 1991.

Abiturientin' außerhalb des intensionalen Kontextes (,die jüngste Abiturientin' steht außerhalb des Skopus des Existenz-Quantors). Anders gesagt: Die Existenz der jüngsten Abiturientin ist unabhängig vom Inhalt des Glaubens gesichert.

(2) (b) impliziert nicht (a) (also *de re* impliziert nicht *de dicto*): Wenn x von x für glücklich gehalten wird und mit der jüngsten Abiturientin von Schriesheim (nämlich mit Marie) identisch ist, muss sie weder wissen, dass ihr selbst der Status der Weinkönigin zugesprochen noch dass es Marie, die jüngste Abiturientin am Ort (die sie ist), ist, die von x für glücklich gehalten wird.

(3) (c) impliziert nicht (a) (also *de se* impliziert nicht *de dicto*): Wenn die jüngste Abiturientin von sich glaubt, sie werde zur Weinkönigin gewählt, muss sie nicht glauben, es sei die jüngste Abiturientin, die Weinkönigin wird (sie identifiziert sich in der *de-se*-Lesart nicht als Marie, die tatsächlich jüngste Abiturientin vor Ort).

(4) (a) impliziert nicht (c) (also *de dicto* impliziert nicht *de se*): Glaubte die jüngste Abiturientin, die jüngste Abiturientin werde die diesjährige Weinkönigin, so muss sie nicht wissen, dass sie mit der Kennzeichnung ,die jüngste Abiturientin von Schriesheim' auf sich selbst Bezug nimmt.

(5) (c) impliziert (b) (also *de se* impliziert *de re*): Wenn jemand was über sich selbst glaubt, so glaubt er es von jemandem.

(6) (b) impliziert nicht (c) (also *de re* impliziert nicht *de se*): Wird x von x für die kommende Weinkönigin gehalten (und ist außerdem die jüngste Abiturientin von Schriesheim), so muss beides nicht Gegenstand ihres Selbst-Glaubens sein (x mag identisch sein mit der jüngsten Abiturientin, ohne dass sie das weiß; und sie mag x die Weinköniginwürde zuschreiben, ohne zu wissen, dass sie selbst x ist).

Das ist ganz schön kompliziert – aber die Konsequenz ist doch leicht zu verstehen.

Und Sie können ja in größerer Ruhe einmal durchrechnen, ob Chisholms (auf Castañeda gestützte) Begründung der Irreduzibilität von Selbstwissen auf Wissen *de dicto* und *de re* logisch in Ordnung ist.

Versprochen hatte ich Ihnen aber (obwohl ich das in der Vorlesung nicht vortragen werde), David Lewis' Argument für die Nicht-Propositionalität von Selbstwissen (Wissen *de se*) nachzuliefern – einfach darum, weil Chisholm explizit Lewis' Bedeutung hervorhebt (Chisholm 1981, 3, 32, 40), auch wenn er sich mehrfach auch auf Castañeda beruft (l. c., 18, 22, 26, 35, 53, 122)

Man hat die These, wonach die Welt sich auf Sachverhalte reduziere (Wittgenstein, David Armstrong) gelegentlich ‚Faktualismus‘ genannt. David Lewis ist damit beschäftigt, diese Position zu widerlegen. Auf Fakten beziehen wir uns in propositionalen Einstellungen, also in Einstellungen *de dicto*. Dass wir auch so das Eigene von Selbst-Wissen nicht erfassen, wohl aber in Einstellungen zu gewissen Eigenschaften – das ist Gegenstand von Lewis‘ zu Recht berühmtem Aufsatz „Attitudes *De Dicto* and *De Se*“ (zuerst 1979 erschienen, wiederabgedruckt in Lewis 1983)

Das wird so begründet:

- a. Propositionen sind Mengen möglicher Welten, einer ‚Region des logischen Raums‘.
- b. Eigenschaften (Relationen eingeschlossen) sind die entsprechenden Mengen von Wesen, die diese Welt bewohnen (bzw. in der entsprechenden Relation zu ihr stehen). Dabei gilt die *Annahme*, dass jedes Subjekt genau *eine* Welt bewohnt, die es indexikalisch zueignet.
- c. Kraft a. allein (propositionalem Wissen) wissen wir nicht, welche Welt *wir* bewohnen (wie – in Lewis‘ Beispiel – die allwissenden Götter auf getrennten Bergen, deren einer Manna spendet, während der andere Blitze schleudert; aus dieser Beschreibung allein wissen sie nicht, wer welcher Gott ist).
- d. Aber wir wissen das durch Selbstzuschreibung der relationalen Eigenschaft, Bewohner genau einer (unserer) Welt zu sein. Es gibt einen Weg von *einigen* Eigenschaften zu *allen* Welten, aber nicht umgekehrt.

e. Also ist nicht alles Wissen propositional. Wohl aber hat Selbstwissen einen klar kognitiven Charakter.

Für dieses Thesen-Bündel sind viele Gedankenexperimente bemüht worden. Das berühmteste ist John Perry's in der UB Stanford verirrter Rudolph Lingens, der sein Gedächtnis verloren hat (Perry 1977; Lewis 1983, 138 f.). Selbst eine Karte mit einem roten Pfeil und der Angabe „Sie stehen jetzt hier“ kann er so wenig auf sich beziehen wie die Lektüre einer Lebensgeschichte von Rudolph Lingens, die über der Auftritt seiner Amnesie hinaus bis zu dem Zeitpunkt fortgezählt ist, an dem er jetzt gerade steht und dies erfährt: Er kann diese Informationen, ohne vorherige nicht-propositionale Selbstkenntnis, nicht auf sich selbst beziehen. Lediglich sein gegenwärtiges Wahrnehmungs- und Verständnis-Bewusstsein schreibt er sich unmittelbar und nicht-propositional zu; das reicht nicht zur Stiftung einer selbstbewussten Identität.

Fazit: Subjektivität sperrt sich in ihren beiden Varietäten (*de re* und *de dicto*) der Reduktion: in der ersten auf Gegenständliches, in der zweiten auf Tatsächliches.

Die leichte Präferenz, die Chisholm für Lewis' Erklärung zeigt, ist zweifellos darin begründet, dass Lewis' Ontologie mit viel sparsameren Mitteln zum Ziel kommt als der ontologisch hoch komplizierte, fast phänomenalistische Ansatz Castañedas. Wie Lewis ersetzt Chisholm eine auf Propositionen (oder Tatsachen) basierte Ontologie (à la David Armstrong: *A World of States of Affairs*, Cambridge University Press, 1997) durch eine Eigenschafts-Ontologie. Intentionen sind danach relationale Eigenschaften, die sich ein denkendes Wesen direkt (und evidenterweise) zuschreibt.

Wir kennen Chisholms Motivation für diese ontologische Option (ich wiederhole Bekanntes in anderen Worten): Sogenannte *de-se*-Attributionen, also Wendungen mit 'er/sie', einem Intentionsverb und 'er/sie selbst' (oder 'sich selbst' – Beispiel: 'Sie glaubt, sie sei selbst die Weinkönigin von Schriesheim'), schöpfen die ihnen eigene Sicherheit aus der logischen Unersetzlichkeit der 'ich'-Perspektive. Chisholm schließt daraus auf die 'Unplausibilität der Propositional-Theorie von (z. B.) Glaubens-Kontexten' (Chisholm 1981, 15). Die Normalfunktion von Sätzen mit 'ich' (oder anderen deiktischen Ausdrücken) sei es nicht, Propositionen auszudrücken (l. c., 17,5). Das ließ sich, wie wir sahen, an einer Eigentümlichkeit von Glaubenssätzen mit 'er' und 'sich selbst' illustrieren. Wenn ich etwas von mir selbst glaube, muss ich mir keineswegs irgendwelche in Sätzen ausgedrückten Kenntnisse zuschreiben, wie Dritte das täten; ich kann selbst meinem Spiegelbild misstrauen, meine Biographie für einen Traum halten und meinen Charakter verleugnen: Immer noch werde ich mit unfehlbarer Gewissheit mit mir (und evtl. meinen Zweifeln an meiner Biographie) vertraut sein. Welcher Art ist dann dieses Ich? Keineswegs von der Person, wenn 'Person' meint: ein raum-zeitliches Objekt mit einer kausal rückverfolgbaren Geschichte, deren Elemente in Propositionen aufgereiht werden können. Elizabeth Anscombe hatte gezeigt, dass, wenn Descartes zweifelt, ob er selbst Descartes ist, er keineswegs die Gewissheit seines 'cartesianischen cogito' in Frage stellt („The First Person“, in: S. Guttenplan, ed., *Mind and Language: Wolfson College Lectures 1974*, Oxford 1975, 45-65; wiederabgedruckt in *Collected Papers II, Metaphysics and the Philosophy of Mind*, Oxford: Blackwell, 1981, 21-36, hier: 22). So ging es John Perry, der in einem Stanforder Supermarkt den Besitzer des Wagens suchte, aus dem eine feine Zuckerspur rann

– nicht wissend, dass er es selber war. Epistemischer Selbstverweis ('x glaubt, dass *er selbst* soundso') ist also nicht zu reduzieren auf Wendungen de dicto ('x glaubt [die Proposition], daß x soundso') oder Wendungen de re ('Es gibt ein x so, daß x mit mit y identisch ist, und x hält x für soundso') (Chisholm 1981, 18 f., 32). Die Pointe von Chisholms Argument besteht nun darin, daß die objektivistische (*de re-*) Formulierung die 'emphatische' Selbstkenntnis *nicht* impliziert, dass also 'emphatisches Selbstbewusstsein' aus einer objektivistischen, in Propositionen niedergelegten Beschreibung des gleichen Sachverhalts ohne Verwendung 'emphatischer Reflexiva' ('er/sie ... *sich selbst*') nicht abzuleiten ist. So zeigt sich, dass cartesianisches Selbstbewusstsein an einen *pour-soi*-Standpunkt gebunden ist, der in (fallibles) propositionales Wissen über die real existierende Person in der Welt nicht aufgelöst werden kann. Chisholm spricht von nicht-propositionaler oder direkter Selbstzuschreibung einer *Eigenschaft*. Eigenschaften sind also in Chisholms Ontologie die Nachfolger von Propositionen. Beim Selbstbewusstsein ist es die *Eigenschaft, in einem intentionalen Bezug auf einen Gegenstand zu stehen*. Schreibe ich mir diese Eigenschaft direkt zu, so durchlaufe ich nicht den Prozess der Deutung von etwas als etwas. Ich forme keine Aussage, die falsch sein könnte. Ich beziehe mich auf einen Gegenstand in der Welt (und hinsichtlich der Interpretation desselben kann ich mich irren). Aber die Eigenschaft, in einer epistemischen Beziehung zu dem Gegenstand zu stehen, die schreibe ich mir *direkt* und unfehlbar zu (Brentanos 'modo recto' oder 'inneres Bewusstsein'). Denn sie betrifft ja 'mich selbst', sie ist ein Urteil 'per me', nicht 'per accidens', also nicht ein Urteil über meine innerweltliche Person, keine Beschreibung, die wahr oder falsch sein könn-

te, keine Einsicht, die intersubjektiver Bestätigung oder Kommunikabilität bedürfte, und keine Identifikation, die treffen oder fehlgehen könnte.<sup>17</sup>

Nun können wir uns zum Abschluss fragen, ob Chisholm wirklich gelungen ist, uns eine ungegenständliche Theorie des Selbstwissens zu liefern. Ein Verdacht drängt sich dagegen auf. Chisholm hält nämlich die direkte Selbstzuschreibung von Intentionen für gegenständlich (oder für eine Beziehung).<sup>18</sup> Wie aber könnten Relationen, die etwas auf etwas beziehen, *direkt* heißen? Und wie sollten sie den Träger der Eigenschaft, die sie direkt zuschreiben, nicht als Objekt bestimmen? In der Tat spricht Chisholm vom "*object of direct attribution*", welches 'immer ich selbst bin' (Chisholm 1981, 35 f.).<sup>19</sup> Ist das Ich einmal als *Objekt* in Anschlag gebracht, sieht man schwer, wie es hernach in seiner lauterer Subjektivität (d.h. in seiner Ungegenständlichkeit) bekannt werden könnte, ja worauf die cartesianische *Gewissheit* beruhen könnte, die ich davon habe, daß ich mich nicht auf irgendwas, sondern auf mich beziehe. Chisholms Grundeinsicht in den epistemischen Sonderstatus der de-se-Attribution scheint getrübt durch seine objektivistische Auslegung des 'Inhalts' der Selbstzuschreibung.<sup>20</sup>

<sup>17</sup> In der Abwehr der 'altmodischen Propositionalitäts-Doktrin' und einer intersubjektivistischen Deutung von Selbstbewußtsein (à la Taylor) ist Castañeda mit Chisholm (und Lewis) übrigens einig. Vgl. H.-N. Castañeda, „Self-Consciousness, Demonstrative Reference, and the Self-Ascription of Believing“, in: J.E. Tomberlin (ed.), *Philosophical Perspectives: 1. Metaphysics*, Atascadero 1987, 405-454, hier: 409. Auch Castañeda gründet die 'Selbstdurchsichtigkeit' von Bewusstseins-Episoden auf ihre (von fremdem Verständnis, intersubjektiver Zugänglichkeit und Wahrheitswerten unabhängige) 'unmittelbare Einsichtigkeit' (412 f., 414 f., 418).

<sup>18</sup> "A believer can take himself als his intentional object"; "I will be the *object* of such consideration [myself as having a certain property] and the property I consider myself as having will be the *content*" (Chisholm 1981, 28 f., vgl. 32, 35 ff.).

<sup>19</sup> Anderswo: "I am the *object* of these [mental] attributions." "The *object* of (...) direct attribution will be the subject himself" (Chisholm 1981, 75, 86 o., 89 f.).

<sup>20</sup> Ich habe diesen Einwand ausgeführt in *Selbstbewußtsein und Selbsterkenntnis. Essays zur analytischen Philosophie der Subjektivität*, Stuttgart: Reclam 1991, 320 ff.

Unbefriedigend ist aber auch, was Chisholm zum erkenntnistheoretischen Status des Selbstbewusstseins selbst (vor allem im 7. Kapitel über "Certainty and the Unity of Consciousness") anführt. Selbstbewusstsein, sagt er, sei kein Ich-Bewusstsein; die unmittelbare Kenntnis des *Inhalts* eines direkt-attribuierten Prädikats sei an ihr selbst "subjectless" (l. c., 75, 86 u.). Diese Ansicht steht in der Tradition des frühen Brentano, aber auch des frühen Husserl und Sartres. Wir haben sie auch bei Ned Block und Frank Jackson wiedergefunden. Wie aber erwirbt die Person Kenntnis nicht nur ihrer psychischen Zustände, sondern auch ihres Selbst (als des Objekts der Attributionen, das nicht selbst zu den Eigenschaften dazugehört [l. c., 75,5])? Chisholms Antwort lautet: durch Reflexion. Ich kann mir nicht eine *Eigenschaft* direkt-zuschreiben, ohne auch zu wissen, dass *ich* es bin, dem ich sie zuschreibe (l. c., 86 f.). Aber 'ich' steht hier nicht für den Gegenstand einer individuellen Wesenheit, die nur mich individuiert, sondern für den Träger derjenigen Eigenschaft, die prinzipiell auch andere haben könnte, die ich aber gerade *mir* direkt-zuschreibe. Die mentale Selbstzuschreibung verschafft mir keinen Begriff von *mir* ("concept of *himself*"), sondern von dem *Selbst*, dem ich die Eigenschaft zulege ("concept of *a self*"). Schreibe ich mir etwa den Zustand der Traurigkeit zu, so schreibe ich sie mir als dem Selbst zu, an dem diese Eigenschaft mit cartesianischer Gewissheit gerade exempliziert ist – und zwar auf dem Wege der Reflexion. Ich merke auf den betreffenden psychischen Akt oder Zustand auf. Dies Aufmerken nennt Chisholm "considering that". Gelangt nun aber der Gegenstand des Aufmerkens erst durch die Reflexion zu der Eigenschaft der epistemischen Selbst-präsentiertheit, so wäre diese 1. keine ursprüngliche Gegebenheit des Bewusstseins, und 2. hätte das aufmerkende Bewusstsein auch kein Kriterium dafür, sich diesen nicht-selbst-präsentierten Zustand *als* sich selbst zuzu-

schreiben (vgl. l. c., 80, 82, 84, 89 f.), So dreht sich Chisholms Beschreibung – auch gegen seinen Widerstand (vgl. l. c., 81) – in einem offenkundigen Zirkel. Wir kennen ihn von Rosenthals *second-order*-Modell, dem gegenüber Chisholms Ansatz natürlich einen ganz entscheidenden Fortschritt darstellt.

Zirkulär wird Chisholms Theorie aber auch im Blick auf seine Ansicht, selbstgebende (direkt-zuschreibbare) Eigenschaften seien nicht-propositional. Chisholm spricht in bezug auf Selbstbewusstsein von 'nicht-propositionaler Gewissheit' (l. c., 82 u.). Soll diese 'supervenient sein über einer psychischen Eigenschaft und deren epistemischer Reflexion (and such that one considers one's being  $\phi$ )', dann wird hier das 'epistemische level' der Gewissheit erstiegen über einen kombinierten Zustand (von Haben und Aufs-Haben-Aufmerken), in dem solche noch nicht vorlag. Was soll das aber bedeuten? Entweder dies, dass Gewissheit in Analogie zur (propositionalen) Wahrheit gedacht wird – und dann müsste ihr ein Zustand möglicher Nicht-Gewissheit entsprechen, was angesichts der Alternativlosigkeit cartesianischer Gewissheit gegenstandslos ist –; oder dies, dass die Gewissheit einen epistemischen Zustand nur explizit macht, der auch schon vorher vorlag – dann aber macht die Rede von 'Supervenienz' keinen Sinn mehr.<sup>21</sup>

## Anhang (kritische Frage an mich selbst)

---

<sup>21</sup> Chisholm definiert sie so: Supervenient ist eine Eigenschaft  $G$  über einer Eigenschaft  $H$ , wenn  $H$  zwar nur im Verein mit  $G$  auftreten kann, eine Person, die sich  $G$  zuschreibt, damit aber nicht notwendig auch schon  $H$  zuschreibt. Soll  $G$  epistemische Gewissheit und  $H$  das Haben eines psychischen Zustands ausdrücken, dann sollte die Situation eintreten können, daß ich den psychischen Zustand zwar habe, ihn aber nicht für gewiss halte – was absurd wäre. Denn die explizit selbst-attribuierte Gewissheit kann nur an einen Zustand anknüpfen, in dem sie selbst schon ein - meinetwegen impliziter – Bewusstseinsbestand war; andernfalls wäre ihre Zuschreibung einfach unbegründet.

„So weit, so gut“, werden Sie sagen. Die für meine Argumentation zentrale Überzeugung von der Präreflexivität sowohl des Selbstbewusstseins als auch des Selbstwissens zieht ihre Kraft bisher aus dem Nachweis, dass alternative Erklärungsmodelle scheitern, indem sie in die bekannten Schwierigkeiten geraten (Zirkel und Regresse). Es wäre aber – so werden Sie verlangen – etwas Positives zu dem Phänomen zu sagen: Welches sind die Wahrheitsbedingungen von ‚ich‘-Äußerungen?

Ferner: Die Präreflexivitäts-These arbeitet mit der Annahme, es sei ein *nicht relationales Bewusstsein seiner selbst* (im Doppelsinn: des Bewusstseins und seines Trägers, des Ich) denkbar, obwohl uns die Sprache bei der Ausbuchstabierung dieses Gedankens zum Gebrauch von Reflexivpronomina zwingt (‚sich‘, ‚selbst‘). Ist das ein Zufall, und können wir uns über die sprachlich gespiegelte Reflexivität des Phänomens mit einem Machtspruch hinwegsetzen? Und können wir Zeitbewusstsein, durch das wir unser Selbst in geordneter Weise zerlegen (in eine Biografie), aus der Unmittelbarkeit eines einstelligen (undifferenzierten) Selbst verständlich machen? Betreffen Persönlichkeitsstörungen wie die Schizophrenie (mit Erlebnissen wie dem der Gedankeneingebung) unser Selbstwissen selbst (wie Christopher Frith und andere Psychiater annehmen), oder werden sie erlebt als Störungen, die einem Ich widerfahren, das insofern intakt sein muss, dass ihm dies Leiden als das seine bewusst werden kann (Paul Schilder, Shaun Gallagher und andere).

Mit den letzten beiden Fragen hängt die andere eng zusammen, die, soviel ich weiß, in dieser Form nur Robert Nozick aufgeworfen, aber nicht weiter bearbeitet

hat: Ist eine irrelationale oder nicht-duale epistemische Selbstbeziehung logisch überhaupt vorstellbar („The Identity of the Self“, in: *Philosophical Explanations*, Cambridge/MA: Harvard Univ. Press, 1981)? Nozick bezieht sich beiläufig auf Fichte, über den er in Harvard ein Seminar gegeben hat und dessen Rede vom ‚sich setzenden‘ Ich ihm gerade wegen der darin sich ausdrückenden Ratlosigkeit („perplexity“) einleuchtet (S. 76, 89). Natürlich bezeichnet die Formel ein Problem, statt es zu lösen. Nozick erwähnt auch nicht den (ebenfalls ratlosen) Versuch, Selbstbewusstsein als prä-reflexiv zu verstehen, sondern unterstellt schon (ohne Begründung), dass „reflexiver Selbstbezug“ Selbstwissen impliziert; und von Selbstwissen sagt er, dass es nicht verständlich wird in Begriffen von irgend etwas anderem, das es konstituiert oder hervorbringt, ausgenommen seine ebenfalls nicht weiter explizierbare wesentliche Eigenschaft, mit sich identisch zu sein (82). In diese Situation gerät er, weil er (wie Fichte) reflexives Selbstwissen durch etwas Unmittelbares fundiert denkt, d. h. als nicht erklärbar aus nicht-reflexiven Selbstbezügen, in denen ohne Selbstwissen eines („mittelbar“) in einer Beziehung auf sich selbst steht („No reflexive self-referring statement is derivable from only non-reflexively self-referring ones“ [74]). Nicht-reflexiv sind typischerweise Selbstbezüge *de re*, die sich in Beispielsätzen wie diesem ausdrücken: „Es gibt ein  $x$ , so dass  $x$  mit der jüngsten Schriesheimer Abiturientin des Jahrgangs 1971 identisch ist, und  $x$  wird von  $x$  für die Weinkönigin von 1971 gehalten.“  $X$  referiert richtig auf sich selbst,  $x$  ist bei Bewusstsein, sie referiert also bewusst richtig auf sich selbst, und  $x$  erfüllt nicht nur eine oder mehrere, sondern schreibt sich auch richtig eine oder mehrere Eigenschaften/Kennzeichnungen zu; aber  $x$  muss in diesem so konstruierten Falle nicht wissen, dass *sie selbst*  $x$  ist.