

In vier Schritten gegen die cartesische Skepsis

Mit Putnam, Wright, Davidson und Moore gegen
die Hypothese vom Gehirn im Tank

Olaf L. Müller

Kurze Zusammenfassung:

Ich möchte philosophische Ideen von Hilary Putnam, Crispin Wright, Donald Davidson und George Eduard Moore zusammenbringen, um hieb- und stichfest zu beweisen, dass die beste skeptische Hypothese (gegen unser Wissen um die Beschaffenheit der Aussenwelt) nicht zutreffen kann. Putnams Externalismus, Wrights zusätzlicher Appell an Disquotationsprinzipien, Davidsons wahrheitskonditionale Semantik und Moores Verweis auf seine eigenen Hände lassen sich zu einem vierzeiligen Beweis verschmelzen, dessen Konklusion besagt, dass wir nicht von Anbeginn Gehirne im Tank sein können.

I. Cartesische Skepsis in modernen Kleidern

Wäre Descartes noch am Leben, so könnte er sein Projekt des umfassenden Zweifels (an unserem Wissen um die Beschaffenheit der Aussenwelt) am besten durch die skeptische Hypothese vom Gehirn im Tank ingangbringen. Dieser Hypothese zufolge stammen die Sinneserfahrungen des vermeintlich wissenden Subjekts nicht aus externen Reizungen seiner Sinnesorgane; das Subjekt verfügt über keine Sinnesorgane, sondern nur über ein Gehirn, in dessen Innern sich genau dasselbe abspielt wie z.B. in meinem Gehirn, dessen Nervenbahnen aber nicht in Sinneszellen enden, sondern an einen Simulationscomputer angeschlossen sind, der dem Gehirn das Dasein einer äusseren Welt, seines Körpers und seiner fünf Sinne nur vorgaukelt.¹

Diese skeptische Hypothese ist so radikal wie die traditionellen cartesianischen Gedankenspiele vom ewigen Traum und vom bösen Dämon. (Unter allen drei Hypothesen wäre — fast — jede Meinung des betroffenen Subjekts über seine Umwelt falsch). Aber die Hypothese vom Gehirn im Tank lässt sich besser verstehen als ewige Träume oder böse Dämonen. Die Hypothese vom Gehirn im Tank ist das beste Pferd im Stall einer modernen cartesianischen Skeptikerin. Um die Skeptikerin zu bekehren, sollte es daher genügen, die Geschichte vom Gehirn im Tank zu widerlegen.

II. Putnams Externalismus

Von Hilary Putnam stammt die erste philosophische Idee, mit deren Hilfe sich die Hypothese vom Gehirn im Tank widerlegen lässt: Ohne kausalen Kontakt zu mindestens einigen Exemplaren einer natürlichen Art von materiellen Gegenständen kann man über diese Gegenstände nicht sprechen.

Dies ist eine Bedingung der Möglichkeit für die Rede über natürliche Arten, die sich durch apriorische Gedankenexperimente rechtfertigen lässt — etwa durch die berühmten Reisen zur Zwillingerde, in deren Flüssen, Regengüssen und Teekesseln anstelle von Wasser (H₂O) eine oberflächlich gleiche, aber strukturell unterschiedliche Substanz namens Zwasser (der Struktur XYZ) vorkommt.²

Putnam hat diese externalistische Idee im ersten Kapitel seines Buchs *Reason, Truth and History* gegen das Szenario vom Gehirn im Tank ins Feld geführt. Putnam behauptet, dass sich die Hypothese vom Gehirn im Tank selbst widerlegt, weil eingetankte Gehirne — mangels kausalem Kontakt — keine eingetankten Gehirne bezeichnen können und weil daher die Hypothese

(0) Ich bin ein Gehirn im Tank,

wenn sie wahr ist (und also von einem Gehirn im Tank ausgesprochen wird), *nicht* von einem Gehirn im Tank handeln, also auch nicht wahr sein kann.³

Diese kühne Überlegung hat einen guten Kern, provoziert aber in der von Putnam vorgelegten Fassung mindestens drei Einwände, die uns nach einer Verbesserung des Beweises Ausschau halten lassen sollten.

Erstens steht das Gehirn im Tank sehr wohl mit einem eingetankten Gehirn in kausalem Kontakt (mit sich selbst). Dieser Punkt scheint eine aufwendige Verfeinerung der externalistischen Bedingung für Erfolg beim Bezeichnen zu verlangen.

Zweitens weckt Putnams Beweis den Verdacht, dass wir Gehirne im Tank sein könnten, die einfach nur ausserstande sind zu *sagen*, dass sie im Tank stecken. Dieser Punkt scheint die antiskeptische Kraft des Beweises von Putnam zu erschüttern.

Und drittens geben die externalistischen Gedankenexperimente offenbar weniger her, als Putnam braucht: Sie zeigen nur, dass man ohne kausalen Kontakt *mithilfe eines Artnamens* keine Exemplare der fraglichen Art bezeichnen kann; ob man vielleicht durch geschickte *Kombination* verschiedener Wörter — durch Beschreibung oder Kennzeichnung — doch erfolgreich Dinge bezeichnen kann, mit denen man nicht kausal verbunden ist, lässt der Externalismus offen. Dieser Punkt provoziert die Suche nach geschickten Beschreibungen, mit deren Hilfe Gehirne im Tank ihre prekäre Lage vielleicht doch erfassen könnten; ob man bei dieser Suche notwendigerweise in magische Theorien des Bezeichnens zurückfallen muss, scheint bis auf weiteres nicht entschieden.⁴

III. Wrights Vierzeiler

Die drei im letzten Abschnitt angeführten Ansatzpunkte für Kritik des Putnam-Beweises lassen sich neutralisieren. Crispin Wright hat eine Rekonstruktion des Beweises vorgelegt, durch die der Beweis den drei Kritikpunkten enttrinnen kann.⁵

Neben Putnams Externalismus (siehe Prämisse (2)) braucht Crispin Wright einen weiteren sprachphilosophischen Grundsatz, um seinen von Putnam inspirierten Beweis ingangzubringen: Disquotation (siehe Prämisse (1)). Der komplette Beweis läuft so:

- (1) In meiner Sprache bezeichnet das Wort "Tiger" die Tiger.
- (2) In der Sprache eingetankter Gehirne bezeichnet das Wort "Tiger" nicht die Tiger.
- (3) Also unterscheidet sich meine Sprache von der Sprache eingetankter Gehirne. (Aus (1) und (2)).
- (4) Also bin ich kein Gehirn im Tank. (Aus (3)).

Die Prämissen (1) und (2) gelten beide apriori. Im Gegensatz zu (2) ist (1) sogar ein analytischer Satz. Der Zwischenschritt (3) ergibt sich logisch aus den Prämissen. (Zwei Sprachen können nicht identisch sein, wenn sie unterschiedliche Eigenschaften haben). Und die Konklusion (4) folgt logisch aus (3). (Wäre ich ein eingetanktes Gehirn, so müsste ich auch dessen Sprache sprechen, im Widerspruch zu (3)).

In der Fassung des Beweises von Crispin Wright verschwinden die drei Probleme, denen der Beweis in Putnams ursprünglicher Fassung ausgesetzt war: Der Beweis wird mit einer natürlichen Art geführt ("Tiger" anstelle von "Gehirn"), mit deren Exemplaren das eingetankte Gehirn wirklich überhaupt nicht in kausalem Kontakt steht. Damit verschwindet das erste Problem.

Zweitens liefert der Beweis von Crispin Wright explizit die Konklusion:

- (4) Also bin ich kein Gehirn im Tank.

Mehr kann die Skeptikerin nicht verlangen. Ihr berechtigtes Unbehagen mit Ergebnissen wie:

Wenn ich die Hypothese "Ich bin ein Gehirn im Tank" formulieren kann, ist sie nicht wahr,

kann sich an (4) nicht entzünden.

Und drittens schliesst Crispin Wrights äusserst schwache Prämisse

- (2) In der Sprache eingetankter Gehirne bezeichnet das Wort "Tiger" nicht die Tiger,

nicht aus, dass das Gehirn im Tank durch geschickte Kombination von Wörtern über Tiger reden kann; die Prämisse sagt, dass der eingetankte Artname "Tiger" für sich allein nicht reicht, um dem Gehirn im Tank zur Bezeichnung von Tigern zu dienen. Diese schwache Behauptung springt in der Tat aus den externalistischen Gedankenexperimenten zur Zwillingserde heraus.

Trotz der Unverwundbarkeit des Vierzeilers von Crispin Wright gegen die drei Kritikpunkte, die Putnams eigene Fassung des Beweises zweifelhaft erschienen liessen, müssen wir nach besseren Fassungen des antiskeptischen Beweises suchen. Denn Wrights Beweis steht und fällt mit der philosophischen Respektabilität des Begriffs des Bezeichnens. Quine, Davidson und ausgerechnet Putnam selber haben aber mithilfe ausgefeilter Permutationsargumente nachgewiesen, dass man Prädikaten jede beliebige Extension zuweisen kann.⁶ Das ist der Kern von Quines "Unerforschlichkeit des Bezeichnens": Wer einem Prädikat unstandardgemässe Extensionen zuweisen will, kann unplausible Konsequenzen vermeiden, indem er auch allen anderen Wörtern der Sprache geeignete unstandardgemässe Extensionen zuweist; welche Gegenstände die Wörter einer Sprache bezeichnen, ist nicht eindeutig bestimmt. (Wie wir sehen werden, bedroht dieser Punkt nicht die Eindeutigkeit des Gebrauchs *ganzer Sätze*).

Betrachten wir dazu ein Beispiel. Als kosmisches Komplement des Katers Murr bezeichnen wir die (mereologische) Summe aller im Universum existierenden Materiebausteine, aus denen der Kater Murr *nicht* besteht. Murrs kosmisches Komplement ist also ein Objekt von gigantischen Ausmassen, in dessen Mitte ein katerförmiges Loch klafft. Auf gleiche Weise bestimmen wir das kosmische Komplement jedes beliebigen materiellen Gegenstands als Summe aller Materiebausteine, aus denen der fragliche Gegenstand nicht besteht; das kosmische Komplement jedes einzelnen materiellen Gegenstands ist also ein wohlbestimmter materieller Gegenstand.

Mithilfe der kosmischen Komplementbildung können wir uns (im Gefolge von Quine (1995), 71) eine alternative Interpretation für die Wörter unserer Sprache ausdenken. Anstelle der standardgemässen Interpretation:

- (1) In meiner Sprache bezeichnet das Wort "Tiger" die Tiger,

setzen wir:

- (1') In meiner Sprache bezeichnet das Wort "Tiger" die kosmischen Komplemente von Tigern.

(Entsprechend für alle anderen Wörter, die materielle Gegenstände bezeichnen).

Quine behauptet, es gebe nicht den geringsten Grund für die Annahme, dass die Standardinterpretation die *wahre* Extension des Wortes "Tiger" besser trifft als die komplementäre Alternativinterpretation. Beide Interpretationen sind laut Quine genau gleich gut; und hier hat Quine offenbar recht. Denn wenn wir sprachliche Ausdrücke mit der Welt konfrontieren, dann findet der Kontakt zwischen Sprache und Welt immer auf der Ebene *ganzer Sätze* statt: Angesichts der Beobachtung irgendwelcher Sachverhalte behaupten und rechtfertigen wir — oder verneinen und kritisieren wir — immer nur ganze Sätze. Welche Teile dieser Sätze wir auf welche Teile der beobachteten Szenerie beziehen, steht dagegen nicht von vornherein fest.

So sprechen ein und dieselben Evidenzen sowohl für den Satz:

(5) Alle Katzen sind Tiere,

als auch für den Satz:

(5') Alle kosmischen Komplemente von Katzen sind kosmische Komplemente von Tieren.

Die beiden Sätze scheinen zwar von unterschiedlichen Dingen zu handeln; der erste Satz scheint von echten Tieren zu handeln, der zweite von gigantischen Gegenständen mit tierartigen Löchern. Aber diese Unterschiede gleichen sich auf Ganzsatzebene aus; als Ganzes stehen die beiden Sätze in jeder Hinsicht genau gleich da. Ist der eine Satz wahr, so auch der andere; gilt der eine als gut begründet, so auch der andere; Einwände gegen den einen Satz sprechen auch gegen den anderen; usw.

Kurz, der Gebrauch beider Sätze ist gleichwertig. Und das bedeutet, dass sich schon beim Sprechenlernen nicht kontrollieren und auch nicht sicherstellen lässt, ob der Sprechende das Wort "Katze" so versteht wie ich — oder ob er sich das Wort so zurechtlegt, dass es in seiner Sprache das bezeichnet, was ich als "kosmisches Komplement von Katzen" bezeichnen würde.

Aus Platzgründen habe ich Quines Argument nur grob wiedergeben können. Ob Quine am Ende mit seiner Behauptung der Unerforschlichkeit des Bezeichnens recht hat, brauche ich hier nicht zu entscheiden. Für unsere Zwecke genügt es zu sehen, dass es im Streit mit der cartesischen Skeptikerin besser wäre, ohne die Behauptung auszukommen, dass sich — entgegen Quine — eindeutig fixieren lässt, was Wörter bezeichnen.

Wenn sich also die Extensionen von Wörtern nicht eindeutig fixieren lassen, dann weckt dies Zweifel an Prämissen wie (1) und (2), in denen so getan wird, als wären Extensionen eindeutig bestimmt. Daher empfiehlt es sich, Wrights Vierzeiler mithilfe eines anderen semantischen Begriffs als

dem Begriff des Bezeichnens zu wiederholen. Der Begriff sollte auf der Ebene ganzer Sätze greifen (da das vorgeführte Permutationsargument im wesentlichen davon lebt, dass Terme unterhalb der Satzebene erst durch den Kontext des ganzen Satzes zu ihrer Bedeutung kommen). Und der Begriff sollte philosophisch unverdächtig sein. Anders als der Begriff der Satzbedeutung (dessen Sinn von Quine mit machtvollen Argumenten angefochten worden ist) bietet sich für unsere Zwecke der Begriff der Wahrheit an. Davidson hat dafür plädiert, den Wahrheitsbegriff für die Sprachphilosophie als grundlegend (und unproblematisch) anzusehen.⁷ Zumindest die cartesische Skeptikerin sollte dem nicht widersprechen. Sie macht selber eifrigen Gebrauch vom Wahrheitsbegriff; sie will ja behaupten, dass es möglich ist, dass (fast) keiner unserer Sätze über die Aussenwelt *wahr* ist.

IV. Davidsons wahrheitskonditionale Semantik

Wenn man unseren Vierzeiler (anstelle von Bezeichnen) mit Wahrsein durchführt, dann lautet eine gute Kandidatin für unsere disquotative Prämisse folgendermassen:

- (1") In meiner Sprache ist der Satz (*) "Hier sind Tiger" genau dann wahr, wenn in der Nähe des Sprechers Tiger sind.

Dies Bikonditional gilt apriori; im Gefolge von Davidson wollen wir seinen Junktor "genau dann, wenn" wahrheitsfunktional lesen, also ohne jede Implikation eines naturgesetzlichen — oder sonstwie notwendigen — Zusammenhangs zwischen Vorder- und Nachsatz.⁸

Auch die zweite — externalistische — Prämisse scheint sich mithilfe von Wahrheit anstelle von Bezeichnen formulieren zu lassen:

- (2") Für alle eingetankten Gehirne a: In a's Sprache ist der Satz (*) "Hier sind Tiger" nicht genau dann wahr, wenn in der Nähe des Sprechers Tiger sind.

Dies ergibt sich offenbar apriori aus Überlegungen zur radikalen Interpretation eingetankter Gehirne. Da eingetankte Gehirne einen Beobachtungssatz wie (*) nicht etwa dann äussern, wenn sich in ihrer Nähe echte Tiger aufhalten, sondern wenn im Simulationscomputer gewisse Bits und Bytes vorkommen (die sogenannten Bit-Tiger aus Nullen und Einsen), verstiesse es gegen das interpretationstheoretische *Prinzip des Wohlwollens*⁹ den Satz (*) aus dem Munde eingetankter Gehirne wortwörtlich zu verstehen; er muss uminterpretiert werden. (Und für den geplanten antiskeptischen Beweis brauchen wir uns nicht auf die Interpretation des Satzes festzulegen; es genügt die negative Behauptung aus (2")).

Leider spricht diese Überlegung nur gegen die syntaktisch unveränderte *Übersetzung* des Satzes (*) in unsere Sprache. Sie spricht nicht für das verneinte Bikonditional in Prämisse (2"). Denn wenn wir den Junktor wiederum wahrheitsfunktional lesen, dann könnte die Prämisse unter gewissen ungünstigen Umständen falsch werden. Wenn der böse Zufall es so will, könnte es vorkommen, dass dem eingetankten Gehirn immer dann aus dem Computer Tigersimulationen eingebläst werden, wenn sich *de facto* in seiner Nähe echte Tiger aufhalten. In diesem Fall träge (2") nicht zu.

Dies ist ein notorisches Problem für wahrheitskonditionale Semantiken im Stile Davidsons. Seit Davidson die raffinierte Idee in die Welt gebracht hat, man könne die Semantik von Sprachen einzig und allein mithilfe wahrheitsfunktionaler Bikonditionale charakterisieren, musste er sich mit widerspenstigen Beispielen herumschlagen, die unserem augenblicklichen Problemfall verdächtig ähneln. Zum Beispiel lässt sich (bei rein wahrheitsfunktionaler Betrachtung) nichts gegen das folgende Bikonditional einwenden:

(6) Der Satz "Snow is white" ist genau dann wahr, wenn Gras grün ist.

Und trotzdem wären wir nicht bereit zuzugeben, dass uns das Bikonditional über die Bedeutung des Satzes "Snow is white" informiert.¹⁰

Die Standardantwort auf solche Probleme lautet, dass Bikonditionale wie (6) nicht isoliert zu betrachten sind, sondern im Zusammenhang einer systematischen Ableitung *vieler* solcher Bikonditionale — für alle Sätze der betrachteten Sprache. Diese Ableitung muss unendlich viele Bikonditionale liefern, die sich nicht alle einzeln aufzählen lassen und die also rekursiv zu gewinnen sind; unter Ausnutzung der Tatsache, dass sich die Wahrheitsbedingungen längerer Sätze systematisch aus den Wahrheitsbedingungen kürzerer Sätze ergeben.¹¹

Um die Grundidee dieser komplizierten Idee zumindest plausibel zu machen, möchte ich zu unserem widerspenstigen Beispiel

(6) Der Satz "Snow is white" ist genau dann wahr, wenn Gras grün ist,

zurückkehren und hervorheben, dass es keinen einfachen systematischen Zusammenhang zwischen diesem Bikonditional und den folgenden beiden Bikonditionalen gibt:

(7) Der Satz "It is snowing" ist genau dann wahr, wenn in der Nähe des Sprechers Schnee fällt;

(8) Der Satz "This is white" ist genau dann wahr, wenn der Sprecher auf etwas Weisses weist.

Diese beiden Bikonditionale passen dagegen auf logisch einfache Weise zu dem erwünschten Bikonditional:

- (9) Der Satz "Snow is white" ist genau dann wahr, wenn Schnee weiss ist.

Man sieht, dass sich (9) auf geschickte Weise aus (7) und (8) gewinnen lässt; wie (6) aus (7) und (8) entstanden sein soll, lässt sich dagegen nicht leicht ausmachen.

Natürlich können wir auch für (6) zwei Bikonditionale angeben, die sich zu (6) so verhalten wie (7) und (8) zu (9):

- (10) Der Satz "It is snowing" ist genau dann wahr, wenn in der Nähe des Sprechers *Gras* ist;
- (11) Der Satz "This is white" ist genau dann wahr, wenn der Sprecher auf etwas *Grünes* weist.

Aber diese Bikonditionale sind falsch; der betrachtete Sprecher wird dem Satz "It is snowing" nicht immer zustimmen, wenn sich in seiner Nähe *Gras* befindet. Bei *Abwesenheit von Gras* und Anwesenheit von Schnee wird er dem Satz "It is snowing" zustimmen; und das spricht gegen das Bikonditional (10). Eine parallele Überlegung lässt sich gegen (11) ins Feld führen.

Kurz und gut, Davidson kann seiner wahrheitskonditionalen Semantik zuliebe unerwünschte (aber wahre) Bikonditionale wie

- (6) Der Satz "Snow is white" ist genau dann wahr, wenn *Gras grün* ist,

aus dem Spiel werfen, indem er darauf verweist, dass sich solche Bikonditionale nicht auf systematische Weise mit anderen wahren Bikonditionalen in Zusammenhang bringen lassen.

Diese Antwort auf das Standardproblem wahrheitskonditionaler Semantiken hat viele Philosophen überzeugt. Trotzdem hinterlässt sie ein mulmiges Gefühl. Denn Davidson hat uns einfach nur glaubwürdig versichert, dass sich irgendwelche unerwünschten Bikonditionale *niemals* auf systematische Weise mit anderen wahren Bikonditionalen in Zusammenhang bringen lassen; begründet hat er diesen Optimismus nicht. Andererseits ist es Davidsons Gegnern nicht gelungen, ein ganzes, rekursiv gut zusammenhängendes System unerwünschter (und doch wahrer) Bikonditionale aufzuweisen. Es war zu schwierig, den standardgemässen Bikonditionalen eine echte, systematische Alternative gegenüberzustellen, die hinreichend grosse Bereiche der betrachteten Sprache umfasst. *Und ausgerechnet solch eine unerwünschte umfassende Alternative scheint sich in unserer Geschichte vom Gehirn im Tank anzubahnen.*

Um diesen Verdacht zu erhärten, machen wir uns zunächst klar, dass externalistische Überlegungen à la Putnam (oder Überlegungen zur radikalen Interpretation à la Davidson) für folgende Standardinterpretation der Tanksprache sprechen:

- (12) In der Tanksprache ist der Satz (*) "Hier sind Tiger" genau dann wahr, wenn sich an geeigneter Stelle im Speicher des Simulationscomputers Bit-Tiger befinden.

Die unerwünschte Alternative zum Bikonditional (12) lautet:

- (13) In der Tanksprache ist der Satz (*) "Hier sind Tiger" genau dann wahr, wenn sich in der Nähe des Sprechers Tiger befinden.

Dies Bikonditional zerstört einerseits unseren geplanten Beweis (daher ist es für die Zwecke unserer Untersuchung unerwünscht); andererseits legt es eine irreführende Interpretation der Tanksprache nahe, da ihm zufolge eingetankte Gehirne — trotz Mangel an kausalem Kontakt — Sätze äussern können, in deren Wahrheitsbedingungen echte Tiger verwickelt sind (daher ist das Bikonditional (13) für Davidsons Theorie der wahrheitskonditionalen Semantik gefährlich; denn diese Theorie sagt, dass sich in Bikonditionalen wie (12) und (13) die *Bedeutungen* objektsprachlicher Sätze wie (*) widerspiegeln).

In Isolation von anderen Bikonditionalen können wir das unerwünschte Bikonditional nicht ausschalten, wie wir gesehen haben: Falls der Zufall es so will, könnten sich immer dann Tiger in der Nähe der Tankkonstellation tummeln, wenn dem Gehirn im Tank irgendwelche (von Bit-Tigern herkommende) Tiger-Vorstellungen eingeflösst werden und wenn also laut (12) der Satz (**) wahr ist. (Ganz entsprechend ist durch einen zufälligen Zug unserer Welt Gras grün, wenn der britische Satz "Snow is white" wahr ist).

Lässt sich das unerwünschte Bikonditional (so wie im Gras-Fall) durch Betrachtung weiterer Sätze der betreffenden Sprache ausschalten? Es scheint nicht so. Denn betrachten wir irgendeinen anderen Tank-Satz über Bit-Tiger, etwa:

- (**) In Indien leben mehr als tausend Tiger.

Die Standardinterpretation lässt sich von folgendem Bikonditional ablesen:

- (14) In der Tanksprache ist Satz (**) genau dann wahr, wenn sich im abgespeicherten Bit-Indien mehr als tausend Bit-Tiger befinden.

Dies passt nicht zu der Interpretation, die folgendes konkurrierende Bikonditional nahelegt:

- (15) In der Tanksprache ist Satz (**) genau dann wahr, wenn sich in Indien mehr als tausend Tiger befinden.

Als Interpretationen im Sinne Davidsons passen die beiden Bikonditionale nicht zusammen; Davidson muss (15) aus dem Spiel werfen. Aber wird ihm das immer gelingen? Durch eine unerhörte Koinzidenz könnten dem Gehirn im Tank zoologische Beobachtungen vorgegaukelt werden, die auf mehr als tausend Bit-Tiger in Bit-Indien zurückgehen und den tanksprachlichen Satz (**) wahr machen, während die gesamte Tankkonfiguration zugleich in einer Welt realisiert ist, in deren indischem Subkontinent sich zufälligerweise mehr als tausend echte Tiger tummeln. In diesem Fall wäre (15) wahr und legte eine Interpretation der Tanksprache nahe, die Davidson nicht zulassen kann.

Frage: Lässt sich Davidsons Sache durch Betrachtung weiterer Sätze retten? Man mag das Gegenteil fürchten, weil es ja durch einen noch unerhörteren Zufall so sein könnte, dass dem Gehirn im Tank genau die Welt vorgegaukelt wird, in der es sein Tank-Dasein tatsächlich fristet. In diesem Fall bestünde eine Isomorphie zwischen der wirklichen Welt und dem Speicher des Simulationscomputers, und die Standardinterpretation der Tanksprache ergäbe Wahrheitsbedingungen, in denen Speicherinhalte des Simulationscomputers verwickelt wären (Bit-Tiger, Bit-Subkontinente, Bit-Katzen usw.) — während die Wahrheitsbedingungen aus der Alternativinterpretation echte Tiger, Subkontinente und Katzen beträfe und *aufgrund der Isomorphie zwischen Welt und Computerspeicher* ebenso unangreifbar wäre wie die erwünschte Standardinterpretation.

Das genau ist die Situation, vor der sich Davidsons Anhänger immer gefürchtet haben: eine Situation, in der sich eine umfassende Interpretationsalternative bietet, die für die gesamte betrachtete Sprache funktioniert und sich genauso systematisch (durch Rekursionen) aufbauen lässt wie die erwünschte Interpretation. Kurz, die Geschichte vom Gehirn im Tank bedroht Davidsons philosophische Theorie der Bedeutung.

Sie bedroht diese Theorie; aber sie zerstört sie nicht, wie wir im nächsten Abschnitt sehen werden.

V. Moores Hand

Es stimmt nämlich nicht, dass die Welt eines eingetankten Gehirns *bis ins letzte Detail* so aussehen kann, wie die Welt uns (und dem eingetankten Gehirn) erscheint. Sowohl uns als auch dem eingetankten Gehirn scheint es so, als hätten wir einen menschlichen Körper und insbesondere zwei menschliche Hände. Ein eingetanktes Gehirn hat aber (nach Konstruktion) überhaupt keinen Körper, also auch keine Hände. Damit kann das folgende unerwünschte Bikonditional auf keinen Fall zutreffen:

- (16) In der Tanksprache ist der Satz (***) "Ich habe Hände" genau dann wahr, wenn der Sprecher Hände hat.

Denn das eingetankte Gehirn hat sehr wohl das, was es in seiner Sprache "Hände" nennt (es hat Bit-Hände) — während es ganz gewiss keine echten Hände hat.

Ob und wie sich Davidsons philosophische Theorie der Bedeutung mithilfe von Sätzen über den Körper des Sprechers vor der Bedrohung retten lässt, die uns im letzten Abschnitt beunruhigt hat, kann ich hier nicht erörtern; der Fall scheint zumindest nicht hoffnungslos. Für die Zwecke dieses Aufsatzes genügt es, dass wir einen Satz gefunden haben, auf den die unerwünschte Interpretation keineswegs zutreffen kann. Dies freudige Ergebnis liefert uns die gesuchte zweite Prämisse des geplanten antiskeptischen Beweises:

- (2*) Für alle eingetankten Gehirne a: In a's Sprache ist a's Äußerung (***) "Ich habe Hände" nicht genau dann wahr, wenn a Hände hat.

Das Gehirn im Tank ist disponiert, dem Satz (***) zuzustimmen, wenn ihm beruhigende Erlebnisse seiner eigenen Hände vorgegaukelt werden; also fast immer. (Ausnahme: Nach simulierten Verkehrsunfällen). Satz (***) ist ein Beobachtungssatz *par excellence*. In der Tanksprache bedeutet der Satz die Anwesenheit von Bit-Händen, die ja für die handartigen Erlebnisse des Gehirns im Tank verantwortlich sind. Der (negierte) Vordersatz unseres Bikonditionals ist also falsch — genauso wie dessen Nachsatz, da das eingetankte Gehirn in Wirklichkeit keine Hände hat. Und damit ist das Bikonditional als ganzes wahr; es eignet sich als Prämisse in unserem Beweis.

Unser endgültiger antiskeptischer Beweis läuft also wie folgt:

- (1*) In meiner Sprache ist meine Äußerung (***) "Ich habe Hände" genau dann wahr, wenn ich Hände habe.
- (2*) Für alle eingetankten Gehirne a: In a's Sprache ist a's Äußerung (***) "Ich habe Hände" nicht genau dann wahr, wenn a Hände hat.
- (3) Also unterscheidet sich meine Sprache von der Sprache eingetankter Gehirne. (Aus (1*) und (2*)).
- (4) Also bin ich kein Gehirn im Tank. (Aus (3)).

Anders als in Moores berühmtem Beweis der Existenz einer äusseren Welt¹² setzt mein Beweis an keiner Stelle voraus, was allererst zu zeigen ist. Mein Beweis hängt nicht davon ab, dass ich schon von vornherein weiss,

dass ich Hände habe. Der Beweis beruht nur darauf, dass es mir (so wie einst Moore) scheint, als hätte ich Hände. Und so scheint es mir in der Tat.

Anmerkungen

- ¹ Putnam (1981), 5-7, 12-13.
- ² Putnam (1975), 223-238.
- ³ Putnam (1981), 7-8, 15.
- ⁴ Ich habe im zweiten Teil meiner Habilitationsschrift versucht zu zeigen, wie man bei dieser Suche fündig werden kann; das Ergebnis führt nicht in eine erkenntnistheoretische Skepsis, sondern in die Metaphysik, siehe Müller (2001).
- ⁵ Wright (1994), 224, 236/7 *et passim*.
- ⁶ Quine (1968), Davidson (1979), Putnam (1981), 22-48, 217-218.
- ⁷ Davidson (1990), 314.
- ⁸ Davidson (1969), 23.
- ⁹ Davidson (1969), 27.
- ¹⁰ Davidson (1969), 25-26.
- ¹¹ Diese Formulierung ist nicht ganz genau und betrifft zunächst nur aussagenlogisch komplexe Sätze; bei prädikatenlogisch komplexen Sätzen muss Davidsons Theorie zu trickreicheren Rekursionen Zuflucht nehmen, die ich hier um der Kürze willen nicht akribisch genau durchspielen mag.
- ¹² Moore (1939), 146-150.

Literatur

- Davidson, D. (1967): Truth and meaning, in: D. Davidson: *Inquiries into truth and interpretation*, Oxford: Clarendon Press, 1984, 17-36.
- Davidson, D. (1979): The inscrutability of reference, in: D. Davidson: *Inquiries into truth and interpretation*, Oxford: Clarendon Press, 1984, 227-241.
- Davidson, D. (1990): The structure and content of truth, in: *The Journal of Philosophy* LXXXVII No.6, 279-328.
- Moore, G.E. (1939): Proof of an external world, in: G.E. Moore: *Philosophical papers*, London: George Allen & Unwin, 1959, 127-150.
- Müller, O. (2001): *Skeptizismus und semantische Stabilität. Über die Grenzen der kausalen Theorie des Bezeichnens*, Habilitationsschrift, eingereicht an der philosophischen Fakultät der Georg-August-Universität Göttingen. (Erscheint vermutlich unter dem Titel *Wirklichkeit ohne Illusionen*).
- Putnam, H. (1975): The meaning of 'meaning', in H. Putnam: *Mind, language and reality. Philosophical papers, volume 2*, Cambridge: Cambridge University Press, 215-271.

- Putnam, H. (1981): *Reason, truth and history*, Cambridge: Cambridge UP.
- Quine, W.V.O. (1968): Ontological relativity, in: W.V.O. Quine: *Ontological relativity and other essays*, New York: Columbia UP, 1969, 26-68.
- Quine, W.V.O. (1995): *From stimulus to science*, Cambridge / Mass.: Harvard UP.
- Wright, Crispin (1994): On Putnam's proof that we are not brains in a vat, in: P. Clark / B. Hale (ed): *Reading Putnam*, Cambridge / Mass.: Blackwell, 216-241.