

[Gekürzt erschienen in: Jan-Christoph Heilinger (ed): *Die Naturgeschichte der Freiheit*. (Berlin: de Gruyter, 2007), pp. 335-364].

## **Die Diebe der Freiheit**

### **Libet und die Neurophysiologen vor dem Tribunal der Metaphysik**

*Olaf L. Müller, Humboldt-Universität zu Berlin (www.gehirnimtank.de)*

#### *GLIEDERUNG:*

- I. Einleitung
  - II. Libet und kein Ende
  - III. Ein extremes Szenario
  - IV. Woran Libets eingetankte Kollegen scheitern
  - V. Metaphysische Freiheit
- Anmerkungen  
Literatur

*ZUSAMMENFASSUNG:* Ich möchte den Neurowissenschaftlern, die glauben, mit *empirischen* Mitteln etwas über menschliche Freiheit herausfinden zu können, eine *philosophische* Herausforderung entgegensetzen. Meine These lautet: Die Frage nach der menschlichen Freiheit ist ein metaphysisches Problem, das sich empirischer Naturforschung entzieht. Um das zu begründen, werde ich ein extremes Gedankenexperiment durchführen. Ich werde zuerst hypothetisch die Situation eines Subjektes beschreiben, dessen Naturwissenschaft berechtigterweise einen durchgängigen kausalen Determinismus im Gehirn postuliert und dessen Libet-Experimente für *all* seine Handlungen fatal ausgehen (nicht nur für unbedeutende Handbewegungen). Dann werde ich zeigen, dass die deterministische Neurowissenschaft in dieser gedachten Situation gar nichts für oder gegen die Entscheidungsfreiheit des Subjekts austrägt, weil sich die Entscheidungen des hypothetischen Subjekts nicht dort abspielen, wo seine Naturwissenschaft hinzielt. – Wenn es mir gelingt, so eine Situation zu konstruieren, stellt sich die Frage, ob wir nicht in einer ähnlichen Situation stecken könnten. Meiner Ansicht nach lässt sich diese Frage zwar aufwerfen, aber nicht beantworten. Sie gehört ins Reich der metaphysischen Spekulation, genau wie die Frage, ob unser geistiges Leben nach dem biologischen Tod im Jenseits weitergehen könnte. Ziel der Überlegungen ist es, verwirrende Redeweisen wie die von meta-physischer oder transzendentaler Freiheit durch konkrete Modelle verständlicher zu machen. Die Position, die hierbei herauskommen wird, hat einige Gemeinsamkeiten mit Kants Position zur Freiheit.

Was Denken und Dichten verbindet,  
ist die Metapher. In der Philosophie  
nennt man Begriff, was in der Dichtung  
Metapher heisst. Das Denken schöpft  
aus dem Sichtbaren seine "Begriffe",  
um das Unsichtbare zu bezeichnen.

Hannah Arendt,  
*Denktagebuch*<sup>1</sup>

## **Die Diebe der Freiheit**

### **Libet und die Neurophysiologen vor dem Tribunal der Metaphysik**

#### *I. Einleitung*

Wissenschaftler widerlegen Willensfreiheit: Unter dieser Schlagzeile versammeln sich seit den achtziger Jahren des vergangenen Jahrhunderts immer mehr experimentelle Evidenzen, die offenbar gegen die Annahme sprechen, dass wir unser Tun mithilfe bewusster Entscheidungen frei zu steuern vermögen. In der Tat entspricht es gutem naturwissenschaftlichem Brauch, täglich vorm Frühstück eine Lieblingshypothese einzustampfen.<sup>2</sup> Nur: Können wir es uns leisten, mit der Annahme von Willensfreiheit so rabiatisch umzuspringen wie mit jeder x-beliebigen Lieblingshypothese? *Wollen* wir uns das leisten? Oder bleibt uns auch hier keine Wahl? Meine Antwort auf alle drei Fragen lautet: nein.

Um mein Plädoyer fürs Vertrauen in die Freiheit abzustützen, möchte ich auf den folgenden Seiten eine neue und extreme Strategie ausprobieren. Sie beginnt mit einem radikalen Zugeständnis an die Gegner der Freiheit und deren neurophysiologische Extrapolationen, so radikal, wie es meines Wissens bis heute kein Freund der Freiheit gewagt hat. Und zwar möchte ich zum Zweck des Arguments annehmen, dass uns die Empirie mit Fakten über das Gehirn überraschen wird, die alles in den Schatten stellen, was uns im letzten Vierteljahrhundert zugemutet worden ist (Abschnitte II bis III). Dass Freiheit selbst unter diesen unangenehmen Annahmen noch lange nicht verloren ist, wird den zweiten Teil meiner Antwort ausmachen (Abschnitte IV bis V). Wie ich zeigen will, ist Willensfreiheit nicht nur ein Thema für die Neurophysiologie – sie gehört auch ins gute alte Arbeitsgebiet der Metaphysik. Etwas ausführlicher: Neurowissenschaftliche Evidenzen können allenfalls *vor dem Hintergrund einer*

*ganz bestimmten – naturalistischen – Metaphysik* gegen Willensfreiheit sprechen. (Laut naturalistischer Metaphysik spielen sich alle Ereignisse, die es überhaupt gibt, gab oder geben wird, in der einheitlichen kausalen Ordnung unseres natürlichen Universums ab, und zwar innerhalb seiner räumlichen Grenzen). Wie wir sehen werden, lässt sich die naturalistische Metaphysik selber nicht mehr naturwissenschaftlich begründen. Gäbe es keine Alternativen zu dieser Metaphysik, so stünde es *vielleicht* wirklich schlecht um die Willensfreiheit (genau wie die eingangs zitierte Schlagzeile sagt). Doch darüber brauchen wir nicht zu streiten, denn: Es gibt Alternativen zur naturalistischen Metaphysik; die Wirklichkeit könnte weit über die räumlichen Grenzen unseres natürlichen Universums ausgreifen. Wer solche Alternativen nicht im Blick hat, leidet an Phantasielosigkeit, genauer gesagt: sein Realitätssinn ist über- und sein Möglichkeitssinn unterentwickelt. Dem möchte ich hier mit einer kleinen Lockerungsübung entgegenreten. Ich werde eine metaphysische Position *vorführen* (nicht: begründen), die den Ort unserer Entscheidungen ins Jenseits verlegt und so dem Zugriff unserer Naturwissenschaften entzieht. Da es gar nicht so leicht ist zu sagen, was das heissen soll, werde ich viel Sorgfalt aufbieten, um die Rede von Freiheit aus dem Jenseits verständlich zu machen, und zwar mithilfe eines Modells, das sich mit rein naturalistischen Ressourcen beschreiben lässt. So hoffe ich, den freiheitsfeindlichen Naturalismus mit seinen eigenen Mitteln aus den Angeln zu heben.

Bevor es soweit ist, möchte ich meinen Gegnern in den nächsten beiden Abschnitten ein weites Stück entgegenkommen, wie versprochen. Ich werde neurophysiologische Experimente beschreiben, die entweder schon durchgeführt worden sind oder *vielleicht* eines Tages durchgeführt werden können und die eines gemeinsam haben: Sie sind allesamt geeignet, unseren Glauben an die Willensfreiheit zu erschüttern; es handelt sich um Experimente mit schockierendem Versuchsausgang. Mir kommen Philosophen, die das nicht zugeben wollen, erstaunlich abgebrüht vor. Und ich kann mir diese Abgebrühtheit der Philosophen nicht anders erklären als mit Blick auf die Philosophiegeschichte. Jahrhunderte lang haben sich die Philosophen ganz allgemein mit der Frage herumgeschlagen, wie sich ein Determinismus, der die gesamte Natur durchzieht, mit Willensfreiheit versöhnen lassen könnte. Das ist eine höchst abstrakte Frage, für deren Antwort viel Raffinement aufgeboten wurde. Aber der Determinismus,

## Die Diebe der Freiheit

der in dieser Debatte durchdekliniert werden musste, blieb die ganze Zeit hindurch ein abstraktes Postulat, das uns Menschen sozusagen nicht nahe genug kam; der Determinismus regierte die ganze Welt, nicht zielgenau ausgerechnet uns Menschen. Und so konnten wir uns an den Determinismus in aller Ruhe philosophisch gewöhnen; daher die Abgebrühtheit.<sup>3</sup> Nur: Uns entgeht etwas ungeheuer Wichtiges, wenn wir die neueren Experimente der Neurophysiologie bloss durch die Brille der alten Debatte über Freiheit und Determinismus ansehen; uns entgeht dabei, dass die neuen Experimente direkt mit uns zu tun haben. Die ganze Debatte über Freiheit bekommt im Lichte dieser Experimente eine ungewohnte Farbe. Um diese Farbe schnell zum Leuchten zu bringen, werde ich es mir erlauben, die wohldokumentierte Debatte über Freiheit und Determinismus links liegen zu lassen und stattdessen als erstes den belebenden Schock vorzuführen, den wir der Neurophysiologie verdanken.

### *II. Libet und kein Ende?*

Betrachten wir zum Auftakt das berüchtigte Experiment, das die Lawine der Unfreiheit ins Rollen gebracht hat: Benjamin Libets zeitliche Untersuchungen zu elektrischen Bereitschaftspotentialen, die sich im Vorfeld willkürlicher Körperbewegungen aufbauen.<sup>4</sup> Für diese Messungen wurden Versuchspersonen aufgefordert, ihre Hand zu einem selber, spontan und frei gewählten Zeitpunkt zu bewegen und dabei (mithilfe einer sehr schnellen Uhr) präzise zu protokollieren, in welchem Augenblick ihnen jeweils der Entschluss zu der fraglichen Handbewegung bewusst vor Augen stand. Zusätzlich hat Libets Arbeitsgruppe am Kopf der spontan entscheidenden Versuchspersonen den Aufbau des sogenannten Bereitschaftspotentials gemessen, einer Grösse, deren jäher Anstieg (wie man seit längerem wusste) einer jeden willkürlichen Körperbewegung zuverlässig vorausgeht.

Als Libet die Zeitpunkte des Aufbaus jener Bereitschaftspotentiale mit den Zeitpunkten verglich, an denen seinen Versuchspersonen die jeweilige Bewegungsentscheidung bewusst wurde, kam eine Überraschung heraus: Die Bereitschaftspotentiale pflegten sich aufzubauen, lange bevor den Versuchspersonen die zugehörigen Entscheidungen für die Handbewegung bewusst werden – das Bereitschaftspotential ist durchschnittlich eine Drittelsekunde früher da.<sup>5</sup>

Anders als Libet selbst gehen einige der radikaleren Interpreten dieser Befunde so weit zu behaupten, dass sich solche und ähnliche Befunde *überhaupt* nicht mit der Annahme verträgen, wir seien in unseren Entscheidungen frei; sie behaupten, unser gegenteiliger Eindruck von Freiheit beruhe am Ende auf nichts anderem als Illusion.<sup>6</sup>

Gegen diesen schnellen Schluss haben sich viele Freunde der Willensfreiheit zur Wehr gesetzt. Hier ist eines ihrer einflussreichsten Argumente, das in vielen verschiedenen Fassungen vorgetragen worden ist. Unsere Entscheidungsfreiheit üben wir nicht so sehr dann aus, wenn wir plötzlich aus heiterem Himmel die Hand bewegen, ohne dass dies irgendeine grössere Bedeutung für unser Leben hätte; vielmehr üben wir sie dann aus, wenn etwas Wichtiges auf dem Spiel steht – und in diesen Fällen geht der endgültigen Handlung oft ein langer Prozess voraus, in dem wir uns unsere Optionen, Chancen und Ängste, Ziele und Hoffnungen, unser Wissen und vieles mehr vor Augen führen und miteinander ins Verhältnis setzen. Alle diese Elemente unserer Entscheidung (so läuft das freiheitsliebende Gegenargument weiter) gehen uns mitunter Tage, Wochen, Monate oder Jahre vor der endgültigen Handlung durch den Kopf, also auch Tage, Wochen, Monate oder Jahre vor dem Aufbau des Bereitschaftspotentials, das dann jene Körperbewegungen einleitet, mit denen wir die fragliche Handlung schliesslich vollziehen.<sup>7</sup>

Trotz aller Vernünftigkeit hat dieses freiheitsliebende Gegenargument einen Haken. Denn Libets ursprüngliches Experiment bildet vielleicht nur die Spitze eines riesigen Eisberges voller Unfreiheit.<sup>8</sup> Zugegeben, im ursprünglichen Experiment ging es nur um einen winzigen Ausschnitt dessen, was Entscheidungsfreiheit sein kann oder doch sein sollte. Aber die Experimentalwissenschaftler feiern ihre *ersten* Erfolge fast immer unter den exzessiven Einschränkungen innerhalb ihres Labors; dann lösen sie sich Schritt für Schritt von diesen Einschränkungen, bis es ihnen gelingt, auch die echten und komplizierten Phänomene ausserhalb des Labors in den Griff zu kriegen. Und in der Tat hat Libet mit seinem ursprünglichen Versuchsaufbau ein gigantisches Experimentierfeld aufgetan, dessen Erträge noch lange nicht ausgeschöpft zu sein brauchen.<sup>9</sup> Vielleicht ist es nur noch eine Frage der Zeit, bis unsere vornehmsten Willensakte vom selben unfreiwilligen Schicksal ereilt werden wie vor einem

## Die Diebe der Freiheit

Vierteljahrhundert die läppischen Handbewegungen der ersten Probanden von Libet.

Fassen wir diese schreckliche Möglichkeit etwas genauer ins Auge. Vielleicht können Libets Nachfolger schon bald nachweisen, dass die neuronale Aktivität<sup>10</sup> im Gehirn des durchschnittlichen Roulette-Spielers auf ganz spezifische Weise in Schwung kommt, lange bevor dem Roulette-Spieler bewusst wird, dass er all seine Jetons auf Rot setzen will statt auf Schwarz.<sup>11</sup> Das wäre bereits bedrohlicher für unsere Entscheidungsfreiheit als Libets ursprüngliches Experiment, denn anders als damals wären jetzt Entscheidungen betroffen, bei denen im wahrsten Wortsinn etwas auf dem Spiel steht.

Zugegeben, solche Roulette-Entscheidungen machen wieder nur einen Bruchteil dessen aus, worum es beim Streit über Entscheidungsfreiheit geht. Aber genau so funktioniert die Salamtaktik der freiheitsfeindlichen Naturwissenschaftler: Stück für Stück wollen sie uns die Entscheidungsfreiheit beschneiden; und wenn sich Libets Methode wirklich auf Roulette-Entscheidungen ausdehnen liesse, dann wären die Gegner der Freiheit einen Schritt weiter als jetzt – nicht mehr und nicht weniger.<sup>12</sup>

Wo könnten sie ihre übernächsten Erfolge feiern? Selbst nach dem eben anvisierten Triumph bei Roulette-Entscheidungen hätten meine Gegner noch keinerlei Evidenzen gegen Freiheit bei längerfristigen Entscheidungen aufgespürt. Aber ich finde es nicht schwer, mir vorzustellen, was sie als nächstes versuchen werden. Sie werden darauf verweisen, dass sich mittelfristige Entscheidungen in viele kleine, kurzfristige Einzelentscheidungen zerlegen lassen. Betrachten wir z.B. die Entscheidung für eine bestimmte Speise im Schnellimbiss. Der Gast wird durch ein menügesteuertes Computerprogramm geführt, das ihn in folgenden Dialog verwickelt:

Möchten Sie vegetarisch essen, so drücken Sie bitte (mit dem linken Zeigefinger) die linke Maustaste, andernfalls (mit dem rechten Zeigefinger) die rechte Maustaste.<sup>13</sup>

[Gast drückt die rechte Maustaste, doch lange bevor ihm die Entscheidung zugunsten dieser Fingerbewegung bewusst wird, hat sich schon das entsprechende laterale Bereitschaftspotential aufgebaut].

Möchten Sie ein Fleischbrötchen essen, so drücken Sie die linke Maustaste, möchten Sie ein Fischbrötchen essen, so drücken Sie bitte die

rechte Maustaste. Wenn Sie Ihre Entscheidung auf die vorige Frage rückgängig machen wollen, so drücken Sie bitte beide Maustasten.

[Gast entscheidet sich fürs Fischbrötchen, und wieder ist das für die rechte Maustaste einschlägige Bereitschaftspotential da, bevor ihm die Entscheidung zugunsten von Fisch bewusst geworden ist].

Möchten Sie Salzwasserfisch auf Ihrem Brötchen, so drücken Sie bitte die linke Maustaste, andernfalls die rechte Maustaste. Wenn Sie Ihre Entscheidung auf die vorige Frage rückgängig machen wollen, so drücken Sie bitte beide Maustasten.

[Bereitschaftspotential zeigt die Entscheidung für Süßwasserfisch an, bevor dies dem Gast bewusst geworden ist].

Ich weiss, ich weiss, so ein Versuchsablauf ist noch nicht in Sichtweite. Aber ich möchte mein Vertrauen in unsere Entscheidungsfreiheit nicht gerne darauf bauen müssen, dass das so bleibt. Und wäre es nicht bedrohlich, wenn die Zeitschrift *Brain* eines Tages solche Versuchsprotokolle zu publizieren begönne?<sup>14</sup> Gewiss, die eben skizzierte unfreie Farce im vollautomatisierten Schnellimbiss muss niemanden anfechten, der im Edelrestaurant sitzt und genüsslich die Speisekarte studiert. Aber was wäre, wenn sich die kulinarische Unfreiheit empirisch weiter ausbreitete? Was wäre mit Bereitschaftspotentialen, die sich aufbauen, lange bevor dem Probanden bewusst wird, welches Tellerchen Nigiri er gleich vom Fließband seines Sushi-Restaurants herunternehmen möchte? Und wer weiss, vielleicht geht es in diesem Stil immer weiter bergab mit unserer Freiheit. Die menügesteuerte Partnerwahl im weltweiten Netz könnte das nächste Opfer der Libet-Methode werden genauso wie das Stimmverhalten unserer Parlamentarier bei Abstimmungen ohne Fraktionszwang.

*Gründe! Gründe! Gründe!* Endlich verschafft sich ein Protestschrei meiner philosophischen Kollegen Gehör, denen unsere ausgedachten Experimente gegen den Strich gehen. Sie sagen: Wir sind in dem Masse frei, in dem wir unsere Entscheidungen an *echten Gründen* ausrichten, und die ausgedachten Experimente sind für diesen Faktor allesamt blind.<sup>15</sup>

Gründe zählen – einverstanden. Aber sie zählen natürlich nur, wenn ihnen handfeste Taten folgen. Solange sich die Gründe und Gegengründe im luftleeren Raum nutzlose Gefechte liefern, solange sind sie für unser Problem irrelevant.<sup>16</sup> Das legt folgendes Experiment nahe: Wir setzen der Versuchsperson eine echte Entscheidung vor, z.B. zwischen zwei schönen Wohnungen, deren Miete wir ein

## Die Diebe der Freiheit

Jahr lang übernehmen werden. (Das Experiment könnte z.B. mit Auslandsstipendiaten durchgeführt werden). Die beiden Wohnungen sind sehr verschieden: die eine ist hell, geräumig und liegt in Fussgängerentfernung vom Institut – die andere liegt in der Nähe des Freibads und des besten Kinos, sie ist ruhig, etwas kleiner und dunkler, hat dafür aber einen schönen Balkon und den schnelleren Anschluss zum Netz. Folgt dreistündiger Auftritt der Berater – Experten und Gegenexpertinnen tragen der Versuchsperson die Gründe und Gegengründe vor und gehen immer stärker ins Detail. Plötzlich heisst es: "Wir müssen uns beeilen, die Zeit drängt! Alle wesentlichen Aspekte sind erwähnt worden. Heben Sie bitte jetzt Ihre linke Hand, wenn Sie die Wohnung mit Balkon wollen; anderenfalls heben Sie die rechte Hand". Und wieder kommt es, fürchte ich, wie es kommen muss – bevor dem Stipendiaten bewusst wird, welche Wohnung er will, verrät ihn schon der Aufbau des Bereitschaftspotentials.<sup>17</sup>

Kurzum, wir Anhänger der Willensfreiheit sollten mit dem Schlimmsten rechnen. Wir sollten unsere Verteidigungslinien auf einer grundlegenderen Ebene setzen, statt uns darauf zu verlassen, dass Libets Methode bei komplexeren und wichtigeren Handlungen (als ausgerechnet Fingerbewegungen) steckenbleiben muss. Einen neuartigen Versuch dieser Art werde ich in den nächsten Abschnitten ausprobieren.<sup>18</sup>

### *III. Ein extremes Szenario*

Ich möchte nun eine Betrachtung vorführen, in deren Licht die Entscheidungsfreiheit selbst dann unangetastet bleibt, wenn es empirisch so schlimm kommt wie im letzten Abschnitt ausgemalt. Genauer gesagt, möchte ich von aussen eine Gemeinschaft fiktiver Wissenschaftler betrachten, deren Libet-Experimente hinsichtlich aller denkbaren Handlungen niederschmetternde Resultate zeitigen und die dennoch falsch lägen, wenn sie behaupteten: "Unsere Entscheidungen werden von Gehirnprozessen determiniert, bevor wir uns der fraglichen Entscheidungen bewusst werden". Was wäre mit der Beschreibung einer solchen fiktiven Wissenschaftlergemeinschaft erreicht? Eine ganze Menge: Sollte sich so eine Situation konsistent beschreiben lassen, dann wäre bewiesen, dass wissenschaftlich korrekte neurophysiologische Analysen nichts über Willensfreiheit besagen müssen, dass also der freiheitszerstörende Schluss

unzulässig ist. Und wenn der Schluss unzulässig ist, brauchen wir uns von unseren Libet-Experimenten auch nicht ins Bockshorn jagen zu lassen.

Die fiktiven Wissenschaftler, die ich nun betrachten möchte, haben Gehirne genau wie wir, aber diese Gehirne stecken nicht in normalen menschlichen Körpern. Vielmehr schwimmen sie in einem Tank mit Nährflüssigkeit herum, in dem sie künstlich am Leben gehalten werden. Die Input-Nervenbahnen dieser Gehirne (die im Fall normaler, vollständiger Menschen mit Sinnesorganen verbunden wären) kommen aus einem gigantischen Simulationscomputer, der die eingetankten Wissenschaftler zuverlässig mit sensorischer Information versorgt – und zwar mit genau denselben Sinnesreizen, denen normale Menschen typischerweise ausgesetzt sind. Mehr noch, sobald sich in den eingetankten Gehirnen Bereitschaftspotentiale aufbauen (die im Falle normaler vollständiger Menschen zu willkürlichen Körperbewegungen führen würden, etwa zur Drehung des Kopfes), werden diese Impulse von den Output-Bahnen des Gehirns in den Simulationscomputer geleitet und dort verrechnet – mit dem Ergebnis, dass die ins Gehirn strömende visuelle Information genau solche Verschiebungen des Blickfelds anzeigt, wie sie beim kompletten Menschen mit Kopfdrehungen einhergehen; und um die Simulation perfekt zu machen, errechnet und produziert der Computer sogar die propriozeptiven Signale, die – im Fall von Kopfdrehungen eines kompletten Menschen – aus dessen Nackenmuskeln ins Gehirn strömen würden.<sup>19</sup>

Wie gelingt dem Simulationscomputer diese umfassende Umwandlung von Output-Signalen in Input-Signale täuschend echter Qualität? Da die ganze Szenerie fiktiv ist, können wir aus dem vollen schöpfen und den Computer mit dem allerschnellsten Prozessor und mit gigantischen Speicherkapazitäten ausstatten. Etwas genauer: In einem Universalspeicher des Computers ist die gesamte physikalische Repräsentation *unseres* Universums geladen. Jeder Speicherplatz des Universalspeichers hat eine dreistellige Adresse  $\langle x, y, z \rangle$  aus rationalen Zahlen und enthält einen der folgenden vier Binärcodes: 01; 10; 11; 00.

Der erste dieser Codes steht im fraglichen Speicherplatz, wenn sich an der Raumstelle  $\langle x, y, z \rangle$  unseres Universums ein Elektron befindet; der zweite dieser Codes, wenn dort ein Neutron steckt; der dritte dieser Codes repräsentiert ein

dortiges Proton; und der vierte Code ("00") repräsentiert eine materiefreie Raumstelle  $\langle x, y, z \rangle$ .

Um Codes nicht dauernd zu verwechseln, möchte ich eine neue Redeweise einführen. Codes vom Typ '01' sollen *Bit-Elektronen* heißen, Codes vom Typ '10' *Bit-Neutronen* und Codes vom Typ '11' *Bit-Protonen*. Diese Redeweise lässt sich bequem erweitern. In unserer Welt bilden zwei Neutronen und zwei Protonen in enger Nachbarschaft einen Heliumkern – im Universalspeicher bilden die vier Codes "10", "10", "11", "11" dann einen *Bit-Heliumkern*, wenn sie in Speicherplätzen abgelegt sind, deren Adressen benachbarte Raumpositionen repräsentieren. (Es geht wirklich nur um die dreistelligen Namen der vier Speicherplätze; die Speicherplätze selbst könnten in der physischen Architektur des Simulationscomputers weit auseinanderliegen). Im selben Stil werden Bit-Atome, Bit-Moleküle usw. gebildet. Und während in unserer Welt Sterne und Wasserstoffbomben aus vielen Wasserstoffkernen bestehen, setzen sich Bit-Sterne und Bit-Wasserstoffbomben der Bit-Welt aus Bit-Wasserstoffkernen zusammen: aus vielen Codes "10" und "11", die in passenden Speicherplätzen realisiert sind.

Wenn wir uns zur Vereinfachung vorstellen, dass unser Universum nur aus Elektronen, Protonen, Neutronen und deren Zusammensetzungen besteht, so wird man mit Recht sagen können, dass der Universalspeicher jenes Computers gleichsam ein Standbild des Universums bietet. Um die Sache zu dynamisieren, nehmen wir an, dass in einem weiteren gigantischen Speicher die Geschwindigkeiten und Richtungen aller Elementarteilchen codiert sind und dass im Computer ein Algorithmus läuft, dem alle Naturgesetze bekannt sind und der den Universalspeicher in Echtzeit weiterrechnet. Ein Code 01 im Speicher  $\langle x, y, z \rangle$  mit hohem Geschwindigkeitscode zur Zeit  $t_0$  wird also zur Zeit  $t_1 > t_0$  aus dem Speicher  $\langle x, y, z \rangle$  gelöscht und dafür im Speicher  $\langle x+d, y+d, z+d \rangle$  abgespeichert. (Wir nehmen also an, dass sich das zugehörige Elektron während des Zeitintervalls  $t_1 - t_0$  in allen drei Raumrichtungen um  $d$  Längeneinheiten weiterbewegt). Alle simulierten Experimente der eingetankten Wissenschaftler sollen newtonisch funktionieren; dann sind im Tank natürlich besonders dramatische Befunde à la Libet zu erwarten.<sup>20</sup> Zudem wollen wir annehmen, dass die eingetankten Wissenschaftler alle an ein und denselben Computer angeschlossen sind und sich mithilfe von Output-Signalen verständigen können, die im Falle normaler Menschen zu Zungenbewegungen führen würden und die

stattdessen vom Simulationscomputer zuverlässig in akustische Input-Signale für alle Beteiligten umzumünzen sind.<sup>21</sup>

Soweit meine fiktive Beschreibung einer Gemeinschaft von Wissenschaftlern, die genau dieselben Erlebnisse und Versuchsergebnisse registrieren werden wie wir. Ich gebe es zu, eine solche Konstellation ist weit davon entfernt, tatsächlich realisiert zu werden. Aber ich bestehe darauf, dass ihre Realisierung prinzipiell möglich ist; jedenfalls, wenn diejenigen naturalistischen Neurophysiologen recht haben, die all unser Erleben gerne auf Gehirnzustände und neuronale Feuermuster zurückführen. (Und da sich die bevorstehenden Überlegungen gegen diese Neurophysiologen wenden, darf ich mir zum Zwecke des Arguments alle Annahmen zueigen machen, die von meinen Gegnern unterschrieben werden. Ich bekomme diese naturalistischen Annahmen sozusagen kostenlos).

Meine erste Frage lautet: Was bedeuten und bezeichnen die wissenschaftlichen Wörter der eingetankten Gehirne? Wir wollen annehmen, dass diese Gehirne seit Beginn ihrer Existenz im Tank stecken und seitdem genau denselben deutschsprachigen akustischen Spracherwerbs-Reizen ausgesetzt worden sind wie wir. Wovon reden diese Gehirne, wenn sie z.B. "Elektron" oder "Hand" oder "Bereitschaftspotential" oder "Wasser" sagen?<sup>22</sup>

Die Antwort auf diese Frage hängt vom sprachphilosophischen Standpunkt ab. Man könnte behaupten, dass die eingetankten Gehirne mit jenen Wörtern genau dasselbe bezeichnen und meinen wie wir; immerhin verbinden sie mit ihnen genau dieselben Assoziationen und Vorstellungsbilder. Aber mit dieser Interpretationsidee geraten wir ins Stocken, sobald wir uns fragen, wie z.B. handartige Assoziationen und Vorstellungsbilder (die den Tankgebrauch des Worts "Hand" begleiten) dafür sorgen sollen, dass das tanksprachliche Wort ausgerechnet echte Hände bezeichnen kann. Denn die eingetankten Gehirne haben keine echten Hände, und sie haben echte Hände nie gesehen, geschüttelt, geküsst, gerochen oder gebrochen. Alle ihre handartigen Erlebnisse stammen aus dem Simulationscomputer und gehen auf ganz bestimmte Konfigurationen von Codes zurück. Wer Sprechen, Bezeichnen und Meinen gern im guten naturalistischen Geiste verstehen möchte, dem sollte der mangelnde kausale Kontakt zwischen eingetankten Gehirnen und echten Händen zu denken geben. Es wäre die reinste

Zauberei, wenn eingetankte Gehirne echte Hände bezeichnen könnten, obwohl sie von diesen Händen kausal vollständig isoliert sind.<sup>23</sup>

Wie sähe eine weniger magisch anmutende Sicht der semantischen Dinge aus? Einfach: Ihr zufolge gebrauchen eingetankte Gehirne das Wort "Hand" im Beisein enger kausaler Kontakte zu ganz bestimmten Kombinationen von Codes im Universalspeicher; genau diese Codes bezeichnet das tanksprachliche Wort "Hand". Wir werden die fraglichen Codes als "Bit-Hände" bezeichnen.<sup>24</sup>

Darüber kann man natürlich lange streiten; aber in unserer dialektischen Lage ist der Streit nicht nötig. Denn wie gesagt richten sich meine Überlegungen gegen naturalistische Denker. Und im Rahmen einer naturalistischen Auffassung von Sprache steht fest, dass sich Wörter ihre Bezugsobjekte nicht durch Zauberei herausuchen, sondern durch gediegene kausale Verbindungen. (Abermals mache ich mir hier die Voraussetzungen meiner Gegner zueigen, ganz so, wie's immer erlaubt ist).

Was ich eben anhand des tanksprachlichen Worts "Hand" vorgeführt habe, dürfte sich aufs gesamte Vokabular der tanksprachlichen Naturwissenschaft ausdehnen lassen. Das tanksprachliche Wort "Tiger" bezeichnet keine echten Tiger, sondern Bit-Tiger, das tanksprachliche Wort "Schädeloberfläche" keine echten Schädeloberflächen, sondern Bit-Schädeloberflächen, das tanksprachliche Wort "Bereitschaftspotential" keine Ansammlungen von Elektronen im Gehirn, sondern Ansammlungen von Bit-Elektronen im Bit-Gehirn, usw.

Diese naturalistische Interpretation der Tanksprache hat drei wichtige Konsequenzen. Erstens handelt die tanksprachliche Naturwissenschaft nicht von derjenigen Welt, in die Tank, Simulationscomputer, Energiequelle usw. eingebettet sind, sondern nur von einem *Ausschnitt* dieser Welt: vom Universalspeicher des Simulationscomputers und dessen Speicherzuständen.

Die zweite Konsequenz der naturalistischen Interpretation der Tanksprache sichert die Zuverlässigkeit der Wissenschaft im Tank und deren Beobachtungsfundament. Laut naturalistischer Interpretation der Tanksprache sind die Beobachtungssätze der eingetankten Wissenschaftler typischerweise *wahr*. Denn die eingetankten Wissenschaftler werden z.B. den Satz: "Da ist ein Tiger" sagen (und damit die Anwesenheit von Bit-Tigern behaupten), wenn sie in den Genuss tigerartiger

Simulationen aus dem Computer kommen. Und da diese Simulationen typischerweise von Bit-Tigern herrühren, passt der semantische Gegenstand des betreffenden Satzes gut zu dem, was wirklich im fraglichen Moment da ist: ein Bit-Tiger.

In der magischen Interpretation der Tanksprache, die wir zurückgewiesen haben, lägen die Dinge schlechter. Dort würde der Satz von echten Tigern handeln, die natürlich nicht da zu sein brauchen, wenn die Tankwissenschaftler tigerartige Sinneseindrücke geniessen. Tanksprachliche Beobachtungssätze wie "Da ist ein Tiger" wären bestenfalls zufällig wahr – wenn nämlich im passenden Moment ein Tiger um den Tank herumstriche –, also wären sie oftmals falsch und mithin ganz unzuverlässig. Das bedeutet, dass die magische Interpretation der Tanksprache nicht wohlwollend wäre; sie verstiesse gegen die Interpretationsmaxime, einem Sprecher keinesfalls unnötig viele Irrtümer zuzuschreiben.<sup>25</sup>

Die dritte Konsequenz der naturalistischen Interpretation der Tanksprache wird für unsere weiteren Überlegungen besonders wichtig sein: Die eingetankten Gehirne können nicht über das Organ reden, in dem sich ihr mentales Leben *de facto* abspielt – sie können nicht über echte Gehirne reden. Denn wenn sie das tanksprachliche Wort "Gehirn" in den Mund nehmen, so meinen sie damit bestimmte Codes im Universalspeicher, die Bit-Gehirne. Und die bestehen (wie überhaupt alles, wovon die naturwissenschaftliche Tanksprache handelt) aus Stromspannungen an Transistoren (also letztlich aus Elektronenhaufen) – sie bestehen nicht aus Biomasse wie die echten Gehirne, in denen sich das mentale Leben der eingetankten Wissenschaftler abspielt. Und mehr noch, nicht nur können die eingetankten Wissenschaftler keine echten Gehirne bezeichnen oder beschreiben, sie können sie ebensowenig beobachten, manipulieren, zerlegen oder durchleuchten. Alle neurophysiologischen Bemühungen der eingetankten Wissenschaftler gelten den Bit-Gehirnen, nicht den echten Gehirnen: nicht den Trägern ihrer Gedanken, Sinneswahrnehmungen, Wünsche oder Entschlüsse.

Wenn das stimmt, dann sind die neurophysiologischen Bemühungen der eingetankten Wissenschaftler völlig irrelevant für die Frage, wie ihr eigenes Urteilen, Fühlen, Wahrnehmen und Entscheiden funktioniert. Denn der tanksprachliche Satz:

## Die Diebe der Freiheit

- (1) Gehirne urteilen, fühlen und entscheiden, indem ihre Neuronen ein Gewitter von Elektronensignalen abfeuern und dabei elektrische Bereitschaftspotentiale aufbauen,

bedeutet bei angemessener Übersetzung in unsere Sprache:

- (1<sub>ü</sub>) Bit-Gehirne urteilen, fühlen und entscheiden, indem ihre Bit-Neuronen ein kybernetisches Gewitter von Bit-Elektronensignalen abfeuern und dabei bit-elektrische Bit-Bereitschaftspotentiale aufbauen.<sup>26</sup>

Zwar ist der Schlussteil dieses Satzes völlig richtig, denn in Bit-Gehirnen spielt sich genau das bit-neuronale Geschehen ab, das der letzte Teil des Satzes benennt. Aber der Auftakt des Satzes ist falsch – es sind (wenn überhaupt) Gehirne, nicht Bit-Gehirne, denen man Urteile, Gefühle und Entscheidungen zuschreiben muss.<sup>27</sup>

Hiergegen könnte man einwenden, dass wir den Bit-Gehirnen besser doch ein echtes mentales Leben zugestehen sollten. Denn sind sie nicht strukturell isomorph zu echten Gehirnen, also komplex genug, um alle mentalen Leistungen zu vollbringen, die echte Gehirne vollbringen? Vor dieser verführerischen Idee möchte ich warnen. Stellen Sie sich vor, Sie wären *vor kurzem* wissentlich Ihres Körpers beraubt, in den Tank gesteckt und an den Simulationscomputer angeschlossen worden, in dem unter anderem eine isomorphe Repräsentation Ihres ehemaligen Körpers gespeichert ist, einschliesslich einer isomorphen Repräsentation Ihres noch vorhandenen Gehirns. Nehmen Sie an, plötzlich müssten Kosten gespart werden und der Experimentator stellt Sie vor folgende unangenehme Alternative: Entweder wird Ihr Gehirn annulliert oder das Bit-Gehirn Ihres Bit-Körpers. Wofür werden Sie sich entscheiden? Wenn Sie Ihr mentales Leben fortzuführen wünschen, dürfte die Antwort klar sein – jedenfalls für Naturalisten: Um mental zu überleben, brauchen Sie Ihr echtes Gehirn aus Biomasse, keine Codes aus Abermillionen Nullen und Einsen in irgendeinem Computerspeicher. Und wenn Sie sich also für die Annullierung des Bit-Gehirns entscheiden, so werden Sie sich nicht als Mörder fühlen, nicht als Zerstörer von geistigem Leben. Und das ist ein Anzeichen dafür, dass ich im letzten Absatz recht hatte, als ich behauptet habe, dass der Auftakt des Satzes (1<sub>ü</sub>) falsch sei.

Was bedeutet das für die Neurophysiologie? Und was bedeutet es für unser Thema – für Libet-Experimente und deren Interpretation? Die Antwort auf die erste Frage ist nur zum Teil niederschmetternd. Solange sich der Neurophysiologe darauf beschränkt, Gehirne in seiner technischen Sprache der elektrischen

Potentiale, neuronalen Feuermuster usw. zu beschreiben (wie im Schlussteil des Satzes (1)), solange droht ihm von unserem Gedankenexperiment keine Irrtumsgefahr. Erst dann, wenn er die Grenzen seines Faches überschreitet und mentale Vokabeln in den Mund nimmt (wie im Auftakt des Satzes (1)), sind seine Schlussfolgerungen vom Irrtum bedroht. Die eingetankten Kollegen des Neurophysiologen sind diesem Irrtum auf den Leim gegangen; ob es um unsere Neurophysiologie genauso schlimm steht, können wir nicht wissen, denn wir *könnten* allesamt in einer *ähnlich* misslichen Lage stecken wie die eingetankten Kollegen. Genau wie sie würden wir das nicht bemerken.<sup>28</sup>

Kurzum, die Neurophysiologie bewegt sich nur solange auf gesichertem Terrain, wie sie ihren gesicherten Aussagen über Feuermuster, Potentiale des Gehirns usw. keine Aussagen über dessen Urteile, Gefühle, Wahrnehmungen und Entscheidungen hinzuzugesellen versucht. Was das für den letzten dieser vier Begriffe bedeutet, um den es uns hier in erster Linie zu tun ist, will ich im nächsten Abschnitt genauer beleuchten.

#### *IV. Woran Libets eingetankte Kollegen scheitern*

Ich behaupte: Kein noch so drastisches Libet-Experiment der Tankwissenschaft könnte etwas darüber besagen, ob die wirklichen Entscheidungen der eingetankten Versuchspersonen frei sind oder nicht – denn das hinge von den tatsächlichen Bereitschaftspotentialen der eingetankten Gehirne selbst ab (die deren empirischen und sprachlichen Zugriff entzogen sind) und nicht von den Bit-Bereitschaftspotentialen irgendwelcher Bit-Gehirne im Simulationscomputer.

An diesem Punkt drängt sich ein Einwand auf. Er ist einigermaßen lang und lautet: Wenn wir in unserer Sprache von den "Entschlüssen", "Entscheidungen" und "Willensakten" des eingetankten Gehirns reden, dann meinen wir damit die echten neuronalen Prozesse im fraglichen Gehirn, nicht die bit-neuronalen Prozesse im Bit-Gehirn. Und diese echten neuronalen Prozesse, für die *wir* uns also interessieren, kann in der Tat kein eingetankter Wissenschaftler empirisch untersuchen, denn dessen Libet-Experimente setzen bloss am Bit-Gehirn an. Nur: Die seit Beginn ihrer Existenz eingetankten Neurowissenschaftler bedienen sich *ihrer eigenen* Sprache, und in der Tanksprache bezeichnen dieselben Wörter ("Entschlüsse", "Entscheidungen", "Willensakte") etwas anderes – nicht die

echten neuronalen Prozesse, die sich in echten Gehirnen abspielen, sondern (eine Ebene tiefer) gewisse bit-neuronale Prozesse in Bit-Gehirnen.<sup>29</sup> Demzufolge bietet das tanksprachliche Resultat:

- (2) Jede willkürliche freie Entscheidung ist von spezifischen Bereitschaftspotentialen determiniert, die sich aufbauen, lange bevor uns die Entscheidung bewusst wird,

eine *korrekte* Beschreibung der tatsächlichen Lage. Denn bei angemessenem Verständnis dieses tanksprachlichen Satzes müssen wir dessen Übersetzung in unsere Sprache betrachten und auswerten, und hierfür sind *alle* einschlägigen Wörter umzuinterpretieren:

- (2<sub>ü</sub>) Jede willkürliche freie Bit-Entscheidung ist von spezifischen Bit-Bereitschaftspotentialen determiniert, die sich aufbauen, lange bevor uns die Bit-Entscheidung bit-bewusst wird.

Fazit des Einwandes: Auch in der Tankwissenschaft führen wahrheitsgemäße Beobachtungen à la Libet zu korrekten Verdikten gegen das, was die *Tankwissenschaftler* "Entscheidungsfreiheit" nennen. Wenn das richtig wäre, könnte ich das Tankszenario nicht wie geplant heranziehen, um den parallelen Schluss *unserer* Neurowissenschaftler zu kritisieren. Denn diese Kritik kann nur überzeugen, wenn sie einen Schluss auszuschalten versucht, der erwiesenermassen manchmal in die Irre führt.

Um dem Einwand zu begegnen, möchte ich genauer auf die Übersetzung der tanksprachlichen Wörter blicken, von der er lebt. Laut Einwand bezeichnet unser Wort "Entscheidung" gewisse neuronale Prozesse in echten Gehirnen, während das tanksprachliche Wort "Entscheidung" – eine Ebene tiefer – bit-neuronale Prozesse in Bit-Gehirnen bezeichnet. Was ist davon zu halten? Die erste Teilbehauptung meiner Gegner betrifft unsere Sprache und ist strittig; meine naturalistischen Gegner verstehen mentale Ausdrücke als verkappte Bezeichnungen für natürliche Vorgänge im Gehirn – das lässt sich meiner Ansicht nach zwar widerlegen, aber das ist ein anderes Thema, und so steht es in dieser Sache bis auf weiteres unentschieden.<sup>30</sup> Anders bei der zweiten Teilbehauptung meiner Gegner, die von der Sprache eingetankter Gehirne handelt, genauer, von deren Ausdruck "Entscheidung". Ich möchte zeigen, dass dieser tanksprachliche Ausdruck bei Übersetzung in unsere Sprache nicht umzuinterpretieren ist. Wenn das stimmt, sollte der Einwand erledigt sein. – Ein eingetankter Sprecher sagt:

(3) Ich entscheide mich jetzt zu einer Handbewegung.

Was bedeutet das? Da der Sprecher keine Hände aus Fleisch und Blut steuern kann, wohl aber Bit-Hände aus Nullen und Einsen, steht eines fest. Die angemessene Übersetzung von (3) in unsere Sprache muss von Bit-Händen handeln. Aber wie steht es mit dem *Anfang* des Satzes? Wenn der eingetankte Sprecher damit nur gewisse kybernetische Ereignisse im Bit-Gehirn meinen würde, dann würde er sich selber mit einer komplizierten Speicher-Konfiguration verwechseln (mit dem Bit-Gehirn). Dieser (untergeordneten) Speicherkonfiguration würde er alle diejenigen Aktivitäten zuschreiben, die er bei Lichte besehen auf einer höheren Ebene selbst vollzieht – in seinem Gehirn!

Da sich diese Angelegenheit nicht leicht durchschauen lässt, will ich sie von einer anderen Seite beleuchten. Wenn Gehirne im Tank aktiv werden und Veränderungen ihrer Aussenwelt bewirken (nämlich Veränderungen in den Inhalten des Universalspeichers), dann fühlt sich das für sie genauso an wie für uns, wenn *wir* Veränderungen unserer Aussenwelt bewirken; auch spielt sich dann in ihrem Gehirn dasselbe ab wie in unserem Gehirn; zudem bedienen sich die eingetankten Gehirne genau derselben voluntativen Vokabeln wie wir; und sie beachten dabei dieselben Regeln wie wir. Der *einzig*e Unterschied zwischen eingetankten Willensentscheidungen und unseren Willensentscheidungen besteht im *Aktionsradius* dieser Entscheidungen: Wir bewirken durch unsere Entscheidungen Änderungen in der Aussenwelt – die eingetankten Gehirne Änderungen im Universalspeicher des Simulationscomputers (eines kleinen Ausschnitts unserer Aussenwelt). Um diese Unterschiede beim Übersetzen angemessen wiederzugeben, genügt es, den tanksprachlichen Satz:

(3) Ich entscheide mich zu einer Handbewegung,

so zu übersetzen:

(3<sub>ü</sub>) Ich entscheide mich zu einer Bit-Handbewegung,

ohne Änderung des schwierigen Ausdrucks "ich entscheide mich". Die Übersetzung meiner Gegner:

(3<sub>ü'</sub>) Ich bit-entscheide mich zu einer Bit-Handbewegung,

entspringt dem übereifrigen Bedürfnis, alle Vokabeln der Tanksprache so zu behandeln wie deren naturwissenschaftliche Vokabeln. Was sind denn Bit-Entscheidungen? Hirngespinnste meiner Gegner! *Wenn* es Bit-Entscheidungen gibt, dann sind das vermutlich ganz bestimmte bit-neuronale Ereignisse im Bit-Gehirn, die einen gewissen Output bewirken.<sup>31</sup> Die angebliche Übersetzung (3<sub>ü</sub>') wäre also in Wirklichkeit eine Abkürzung für folgendes:

(3<sub>ü</sub>\*) In meinem Bit-Gehirn spielen sich gewisse bit-neuronale Ereignisse ab, die zu einer Bit-Handbewegung führen.

Aber das bietet keine gute Übersetzung des Satzes (3); es bietet allenfalls die Übersetzung eines anderen Satzes:

(3\*) In meinem Gehirn spielen sich gewisse neuronale Ereignisse ab, die zu einer Handbewegung führen.

Und an der Übersetzung dieses Satzes waren wir nicht interessiert; uns interessierte der ursprüngliche Satz:

(3) *Ich entscheide mich* zu einer Handbewegung.

*Diesen* Satz haben unsere Neurowissenschaftler im Auge, wenn sie uns die Entscheidungsfreiheit abspenstig machen wollen. Wenn sie sich stattdessen mit dem Satz (3\*) begnügten, so läge gar kein Angriff auf die Entscheidungsfreiheit vor. Nur wenn sie das gefährliche Wort "Entscheidung" in den Mund nehmen, können sie ihren Angriff lancieren.

Ein weiterer Einwand gegen meine Überlegung versucht auf einem anderen Weg, den *tanksprachlichen* Schluss von der neurowissenschaftlichen Behauptung:

(4) Das Bereitschaftspotential hat sich 350 Millisekunden vor dem Zeitpunkt aufgebaut, den das Protokoll des Probanden als Zeitpunkt der bewussten Entscheidung angibt,

zur folgenden weitergehenden Behauptung zu verteidigen:

(5) Also: Das Bereitschaftspotential baute sich 350 Millisekunden vor der bewussten Entscheidung auf.

Der Einwand sagt nichts gegen meine Übersetzungsvorschläge der beiden Sätze in unsere Sprache:

- (4<sub>ii</sub>) Das Bit-Bereitschaftspotential hat sich 350 Millisekunden vor dem Zeitpunkt aufgebaut, den das Bit-Protokoll als Zeitpunkt der bewussten Entscheidung angibt.
- (5<sub>ii</sub>) Also: Das Bit-Bereitschaftspotential baute sich 350 Millisekunden vor der bewussten Entscheidung auf.<sup>32</sup>

Vielmehr behauptet der Einwand im Einklang mit meinen Übersetzungen, dass die beiden Sätze wahr sind. Verhielte es sich so, dann wäre es mir nicht gelungen, einen Fall zu beschreiben, in dem der Schluss von Wahrem zu Falschem führt, und meine Kritik am gleichlautenden Schluss unserer Neurowissenschaftler verlöre ihre Überzeugungskraft.

Dass der Ausgangspunkt (4<sub>ii</sub>) des tanksprachlichen Arguments wahr ist, liegt (sagt der Einwand) auf der Hand: Wir haben die Bit-Gehirne (genau wie den ganzen Rest des Bit-Universums: des Universalspeichers) als isomorphes Abbild dessen konstruiert, was eine Ebene höher der Fall ist. (Jedem Elektron, Proton, Neutron unserer Welt entspricht ein Bit-Elektron, Bit-Proton, Bit-Neutron im Universalspeicher; den räumlichen und topologischen Beziehungen zwischen den Elementarteilchen unserer Welt entsprechen gleichstrukturierte Beziehungen zwischen den Namen der Speicherplätze, in denen ihre Codes abgespeichert sind; also hat alles, was bei uns aus Elektronen, Protonen und Neutronen zusammengesetzt ist, ein isomorphes Gegenstück im Universalspeicher und gehorcht parallelen Gesetzmässigkeiten).

Zwar können (sagt der Einwand weiter) eingetankte Neurowissenschaftler ihre eigenen Bereitschaftspotentiale nicht direkt empirisch untersuchen – aber ihre empirischen Untersuchungen der Bit-Bereitschaftspotentiale betreffen exakt isomorphe Strukturen. Alle empirischen Ergebnisse auf dieser untergeordneten Ebene sind mithin untrügliche Anzeichen für die parallelen Verhältnisse auf der darüberliegenden Ebene; dafür sorgt die Isomorphie, die wir von Anbeginn in das Szenario der eingetankten Gehirne hineingesteckt haben. Also gibt es (triumphiert der Einwand) auf der Ebene echter Gehirne keinen zeitlichen Spielraum für Entscheidungsfreiheit, denn dort bauen sich die Bereitschaftspotentiale ebenfalls zu früh auf.

Auf diesen Einwand antworte ich in drei Schritten. Im ersten Schritt gebe ich zu, dass der strittige tanksprachliche Schluss von (4<sub>ii</sub>) auf (5<sub>ii</sub>) *bei vollkommener Isomorphie* zwischen übergeordneter Welt und Universalspeicher des

Simulationscomputers tatsächlich von wahren Voraussetzungen zu einer wahren Konklusion führen würde. Aber selbst in dieser Situation sollten wir uns (zweiter Schritt) nicht auf die Zulässigkeit des fraglichen Schlusses verlassen; es gibt viele unzulässige "Schlüsse" von wahren Prämissen zu wahren Konklusionen. Schlüsse sind nur zulässig, wenn die Konklusion wahr sein *muss*, falls die Prämissen wahr sind. Und es ist leicht einzusehen, warum dieser enge Zusammenhang zwischen Prämissen und Konklusion in unserem Fall fehlt: Die eingetankten Neurowissenschaftler untersuchen Bit-Gehirne (Prämisse) und haben mit ihrer Konklusion nur recht, wenn es bei echten Gehirnen strukturell genauso zugeht wie bei Bit-Gehirnen. *Muss* das so sein? Zugegeben, wir haben unsere Geschichte so konstruiert, dass das eine isomorph zum anderen passt. Aber nichts hält uns davon ab, die Geschichte leicht zu verändern: Damit bin ich beim dritten Schritt meiner Antwort auf den Einwand. Anders als bislang bieten die Codes im Universalspeicher diesmal kein *vollständig* isomorphes Abbild der Elektronen, Protonen und Neutronen unserer Welt. Vielmehr soll für unsere neue Fassung der Geschichte nur ausserhalb der Gehirne bzw. Bit-Gehirne Isomorphie walten. Wir stellen uns also auf unserer Ebene eine Welt vor, in der alle Libet-Experimente mit echten Gehirnen scheitern. Bei allen Gehirnen aus Biomasse (so wollen wir annehmen) baut sich das Bereitschaftspotential erst auf, nachdem den Gehirnen die jeweilige Entscheidung bewusst vor Augen steht. (Niemand wird bestreiten, dass wir uns eine solche Situation kohärent vorstellen können – die allermeisten von uns hätten sogar mit diesem beruhigenden Ergebnis der Libet-Experimente gerechnet. Die tatsächlichen Versuchsergebnisse waren überraschend, darum sind gegenteilige Versuchsergebnisse jedenfalls *denkbar*. Das genügt für unsere Zwecke).

Um nun die Isomorphie zwischen übergeordneter Wirklichkeit und Bit-Wirklichkeit genau bei den Gehirnen zu durchbrechen, wollen wir annehmen, dass die eingetankten Libet-Experimente (anders als ihre ausgetankten Gegenstücke) nicht scheitern: Bit-Bereitschaftspotentiale bauen sich auf, lange bevor den Versuchspersonen die Entscheidung zur Bit-Handbewegung bewusst wird.

An dieser Stelle drängt sich ein kniffliger Einwand für Freunde des Denksports auf. Er besagt, dass die soeben ins Auge gefasste Reihenfolge der Ereignisse unmöglich realisiert werden kann: Wenn *zuerst* die freie Entscheidung des

eingetankten Gehirns da sein und sich erst *danach* das Bereitschaftspotential an diesem Gehirn aufbauen soll, dann kann erst *nach* Aufbau dieses Bereitschaftspotentials ein entsprechendes Signal zum Universalspeicher gelangen und ins zugehörige Bit-Bereitschaftspotential umgerechnet werden. (Denn natürlich wird der Simulationsalgorithmus nicht vorhersehen können, zu welchen freien Entschlüssen sich das Gehirn durchringt, das er mit neuronalem Input versorgt). Kurzum, die bewusste Entscheidung muss auf jeden Fall früher auftreten als das Bit-Bereitschaftspotential, und die eingetankten Bit-Experimente à la Libet können nur dasselbe Ergebnis bringen wie deren ausgetankte Gegenstücke; Freiheit auf der oberen Ebene zieht automatisch Freiheit in der simulierten Welt nach sich.

Um den Einwand zu entkräften, müssen wir uns genauer vor Augen führen, wie Libets ursprüngliches Experiment angelegt war. Die Zeitmessung subjektiver Ereignisse funktioniert nicht so einfach, wie es vom Einwand vorausgesetzt wurde. Im Gegenteil, Libet schlug vor, die Zeitangaben seiner Probanden systematisch zu korrigieren, indem er sie mit den Ergebnissen eines anderen Experiments abglich. Hierfür wurden dieselben Probanden gebeten, wieder durch Blick auf die schnelle Uhr festzuhalten, wann sie den subjektiven Eindruck hatten, dass ihre Hand einen bestimmten Reiz empfangen hat, der ihnen in einem zufälligen Moment von aussen versetzt wurde.<sup>33</sup>

Diese Zeitangaben waren zwar genauso subjektiv wie die Zeitangaben fürs Bewusstsein der Willensentscheidung, aber anders als diese liessen sie sich objektiv überprüfen. Es kam heraus, dass der subjektiv festgehaltene Eindruck einen *zu frühen* Zeitpunkt liefert (der also vor dem tatsächlichen Zeitpunkt der Stimulierung liegt).<sup>34</sup> Daher hat Libet in Betracht gezogen, die subjektiven Zeitangaben seiner Versuche systematisch durch Addition eines Zeitbetrages von etwa 50 Millisekunden so zu korrigieren, dass immer ein späterer Zeitpunkt herauskam, als seine Versuchspersonen angaben.<sup>35</sup>

Seine eingetankten Kollegen werden dasselbe tun, und mit Recht. An dieser Stelle können wir ansetzen, um dem Einwand zu begegnen: Wir sorgen dafür, dass der Korrekturbetrag weiter erhöht werden muss. Wie das? Einfach: Wir *verlangsamen* die Signalübertragung vom Universalspeicher zum Gehirn im Tank. Denn wenn diese Informationen deutlich mehr Zeit von der Bit-Retina zur visuellen

Schnittstelle des Gehirns im Tank verbrauchen als die entsprechenden Informationsübertragungen beim kompletten Menschen, dann werden alle Zeitprotokolle des Versuchsaufbaus verzerrt, weil sie von bereits lange überholten Zeigerpositionen der Bit-Uhr stammen. Das wird den eingetankten Wissenschaftlern nicht entgehen; sie verbessern die zeitlichen Berichte ihrer Versuchspersonen durch Addition eines höheren Korrekturbetrags. Und wenn wir zusätzlich das Bereitschaftspotential am eingetankten Gehirn besonders schnell in ein passendes Bit-Bereitschaftspotential am Bit-Gehirn ummünzen, dann werden Libets eingetankte Kollegen genau den zeitlichen Ablauf herausbekommen, den ich oben in Anspruch nehme. Kurzum, Freiheit auf der übergeordneten Ebene lässt sich sehr wohl mit scheinbarer Unfreiheit auf der unteren Ebene vereinbaren.<sup>36</sup>

### *V. Metaphysische Freiheit*

Wenn die Überlegungen aus dem letzten Abschnitt richtig sind, scheitert die freiheitsfeindliche Interpretation der eingetankten Libet-Experimente. Sie beruht zwar auf korrekten, unverdächtigen Beschreibungen dessen, was sich tatsächlich beobachten lässt, wie z.B.

- (4') Das Bereitschaftspotential hat sich 600 Millisekunden vor der Fingerbewegung aufgebaut.
- (4'') Der Proband gab zu Protokoll "Die Uhr zeigte einen Zeigerstand T, als ich mir meiner Entscheidung bewusst wurde".
- (4''') Den Zeigerstand T hatte die Uhr 250 Millisekunden vor der Fingerbewegung des Probanden erreicht.
- (4''''') Also hat sich das Bereitschaftspotential 350 Millisekunden vor dem Zeitpunkt aufgebaut, den das Protokoll des Probanden als "Zeitpunkt seiner Entscheidung" ausweist. (Denn  $350 = 600 - 250$ ).

Alle diese empirischen Aussagen aus der Aussenperspektive werden auch im Tankszenario zuverlässig das beschreiben, wovon sie handeln (Bit-Bereitschaftspotentiale, Bit-Finger, Bit-Uhren, Bit-Protokolle). Aber die hieraus abgeleitete Interpretation:

- (5') Das Bereitschaftspotential hat sich aufgebaut, bevor dem Probanden seine *Entscheidung bewusst* wurde, und zwar 350 Millisekunden vorher,

enthält (kursivgesetzte) Ausdrücke, deren Bezugsobjekte ausserhalb der experimentellen Reichweite der eingetankten Neurowissenschaftler liegen, ohne dass ihnen dies auffiele. Und so kann es (wie wir gesehen haben) den eingetankten Neurowissenschaftlern nicht auffallen, dass Satz (5') falsch ist, obwohl die Sätze (4') bis (4''') wahr sind.

Derselbe unbemerkte Fehler könnte sich bei freiheitsfeindlichen Interpretationen *unserer* korrekten Versuchsbeschreibungen eingeschlichen haben. Vielleicht fällt auch uns nicht auf, dass unsere Entscheidungen an einem Ort stattfinden, der unserem experimentellen Zugriff entzogen ist. Und wenn sie ausserhalb unserer experimentellen Reichweite stattfinden, dann finden sie dort vielleicht auch anders statt, als irreführende Interpretationen hiesiger neuronaler Verhältnisse nahelegen mögen. Vielleicht finden sie dort *freier* statt, als es in unseren Experimenten scheint – nicht etwa blind verursacht vom gedankenlosen Synapsengewitter in unseren Gehirnen.<sup>37</sup>

Sie fragen, wo denn, wenn überhaupt, dieser Ort der Freiheit liegen soll? Die Antwort lautet: nirgends, d.h. nicht in unserem räumlichen Bezugssystem. Etwas gewagter: ausserhalb unseres räumlichen Bezugssystems. Noch gewagter: im Jenseits.

Was heisst Jenseits? Um diese Frage in Andeutungen zu beantworten, möchte ich zum letzten Mal die Geschichte vom Gehirn im Tank heranziehen. Denn die Rede vom "Jenseits" lässt sich – aus der Aussenperspektive – besser überblicken als die parallele Redeweise in unserer Sprache. Wenn die eingetankten Gehirne vermuten würden, dass ihre Entscheidungen ausserhalb der Grenzen dessen stattfinden, was sie den "physikalischen Raum" nennen (und was bei korrektem Verständnis ihrer Sprache nur der Universalspeicher des Simulationscomputers ist), dann hätten sie recht; sie könnten diese korrekte Vermutung zwar artikulieren, aber sie könnten sie nicht begründen. Und genauso hätten sie recht (im Lichte ihrer Evidenzen: unbegründeterweise), wenn sie vermuteten, dass sich ihre Entscheidungen auf "immaterieller" Grundlage ergeben. (Denn was sie "Materie" nennen, besteht aus Nullen und Einsen im Computer, und ihr Gehirn ist aus anderem Stoff gewebt).

Nicht anders steht es vielleicht bei uns. *Wenn* wir in einer analogen Lage wie die Gehirne im Tank stecken, dann gibt es einen übergeordneten, ja: übernatürlichen Bereich der Wirklichkeit, in dem unsere Naturwissenschaften nichts zu sagen

haben, von dem sie nie etwas erfahren können und der dennoch von eminenter Bedeutung wäre. Über diesen Bereich können wir mit der Sprache der Naturwissenschaften nicht reden und noch nicht einmal Vermutungen anstellen. Die Metaphysik hat hingegen mit Ausdrücken wie "Jenseits" einen grösseren sprachlichen Aktionsradius, aber auch sie wird uns hier kein Wissen verschaffen. Ob es diesen übergeordneten Bereich der Wirklichkeit gibt, lässt sich weder naturwissenschaftlich noch apriori entscheiden. Naturalisten verneinen die Frage, aber ohne Beweis – ohne empirische Evidenz.<sup>38</sup> (Oder schlimmer noch, sie ignorieren die Frage aus Mangel an Phantasie).

Es lohnt sich vielleicht, meinen Vorschlag im Spiegel dessen anzusehen, was Singer über seinen dualistischen Gegner sagt; der nämlich (so Singer) "postuliert für die wollende Ich-Instanz einen immateriellen Dirigenten, der das neuronale Substrat nur nutzt, um sich über die Welt zu informieren und seine Entscheidung in Handlungen zu verwandeln. Diese Position ist mit dem Verursachungsproblem konfrontiert und mit bekannten Naturgesetzen unvereinbar. Sie hat den Status unwiderlegbarer Überzeugungen".<sup>39</sup> Dazu in umgekehrter Reihenfolge dreierlei. Erstens hat die naturalistische Behauptung, dass sich alle realen Ereignisse innerhalb der kausalen und räumlichen Grenzen unseres Weltalls abspielen und dass es darüber hinaus nichts gibt, gleichfalls den Status unwiderlegbarer Überzeugungen; hier steht es eins zu eins. Zweitens widersprechen die eingetankten Hypothesen über "Immaterielles" und "Übernatürliches" nicht den im Tank bekannten Naturgesetzen – warum *muss* es bei uns anders sein, wenn wir Wort für Wort dieselben Hypothesen wiederholen? Immerhin haben wir *genau dieselben empirischen Evidenzen* wie die Gehirne im Tank! Drittens zeigt mein Gedankenspiel, wie das Verursachungsproblem gelöst werden könnte: Was die Gehirne im Tank "immaterielle Verhältnisse" nennen, kann fraglos mit dem, was sie "materielle Verhältnisse" nennen, kausal wechselwirken, wenn es eine "höhere Kausalität" gibt, die beide Bereiche umfasst. Die *höhere* Kausalität aus Sicht der Gehirne spielt sich aus unserer Sicht *innerhalb* des bekannten kausalen Terrains ab, in dem wir das Gedankenspiel aufgebaut haben. Aber es könnte, analog, für uns noch eine höhere Kausalität geben als die, um die sich unsere Naturwissenschaftler kümmern. Um diese Möglichkeit zu *illustrieren*, die von Naturalisten immer rundheraus geleugnet wird, habe ich das Gedankenspiel vom Gehirn im Tank hier so ausführlich beschrieben.<sup>40</sup> Aus der Aussenperspektive

folgt das Gedankenspiel allen kausalen Vorgaben der Naturalisten. Doch sobald sie sich auf die Innenperspektive einlassen und sich in das Gehirn im Tank hineinversetzen, sprengt das Gedankenspiel den engen naturalistischen Rahmen.

Gegen diese Überlegung hat Oliver Wachsmuth (in einer elegraphischen Mitteilung) folgendes eingewandt: Trotz allem bleibt es dabei, dass die freien Entscheidungen der Gehirne im Tank das durchbrechen, was sie "durchgängigen kausalen Determinismus innerhalb unseres Weltalls" nennen, und zwar genau an der Stelle, wo die echten Entscheidungen dieser Gehirne in Bit-Elektronen umgewandelt und von aussen in die Bit-Welt eingespeist werden. Zeigt dies nicht, dass Singer recht hat, wenn er behauptet, dass sich immaterielle Freiheit nicht mit "bekannten Naturgesetzen" vereinbaren lasse?

Ich antworte: Sehen wir Singer zuliebe weiter von allen Komplikationen ab, die sich aus dem Indeterminismus der Quantenphysik ergeben (siehe oben Anmerkung 20), betrachten wir also weiter von aussen die newtonische Welt der Gehirne im Tank. Ich behaupte, der Simulationsalgorithmus lässt sich so einrichten, dass die eingetankten Naturforscher ihre jeweiligen Bit-Gehirne berechtigterweise im Griff durchgängig determinierter Kausalketten sehen. Um das zu begründen, möchte ich zuerst daran erinnern, wie schwer es ist, Kausalketten durchs chaotische Milliarden-Gewirr von Neuronen und Synapsen *tatsächlich* zu verfolgen; man kann der Kausalität nicht bei jedem Zahnrad über die Schulter schauen. Mithin wird es im Tank keinen empirischen *Beweis* "durchgängiger Gehirnkausalität" geben, und es fragt sich nur, ob deren *Postulat* berechtigt ist. Dass dies Postulat den eingetankten Neurobeobachtungen zumindest nie widersprechen muss, lässt sich leicht einsehen. Denn die kausale Analyse der bitneuronalen Vorgeschichte von Entscheidungen wird immer nur eine kleine Zeitspanne in die Vergangenheit eindringen, bevor sie sich im Dunkel des bitneuronalen Chaos verliert. Beim gegenwärtigen Stand der Wissenschaft reicht diese Analyse bis zum Aufbau der Bit-Bereitschaftspotentiale zurück, was sehr grob ist angesichts der vielen feinen Schwingungsmuster, die diesem Aufbau vorangehen und ihn *vermutlich* verursachen.<sup>41</sup> Aber selbst bei Verfeinerung und Ausdehnung dieser Rückwärtsanalysen wird sich die Spur der Kausalität ganz sicher in irgendwelchen Details verlieren. Das letzte dann bekannte Glied der Kausalkette heisse Y. Der Simulationsalgorithmus darf natürlich weder Y noch die Wirkungen von Y antasten; er muss die echten Entscheidungen echter Gehirne

(als Bit-Ereignis namens X) *kurz vor Y* ins Bit-Geschehen einspeisen – und zwar ohne beobachtbare Verletzung eingetankter Gesetzmässigkeiten wie "Energie-, Impuls- oder Materie-Erhaltungssatz". Diese Gesetzmässigkeiten lassen genug Spielraum frei; problemlos lassen sich neuronale Netze oder Turingmaschinen konstruieren, die den erwünschten Output X als notwendige kausale Wirkung früherer Zustände erscheinen lassen; und es kann sehr wohl vernünftig sein, sich auf diese Möglichkeit zu berufen, um am Postulat der durchgehenden kausalen Determination aller Gehirnprozesse festzuhalten.<sup>42</sup>

Dass diese Antwort in unserer dialektischen Lage kein Trick ist, sollte auf der Hand liegen. Wie eingangs dargelegt, geht es mir hier nicht um den alten Gegensatz zwischen Freiheit und abstraktem Determinismus. Mir geht es um die Frage, ob sich Freiheit mit den konkreten Ergebnissen vereinbaren lässt, die uns von Neurophysiologen vorgelegt worden sind oder noch vorgelegt werden könnten. Und natürlich werden die Neurophysiologen niemals unser gesamtes Gehirn *en détail* kausal durchleuchten können, soviel Zeit haben sie nicht. Nichts von dem, was sie jemals konkret empirisch herausfinden werden, wird der Idee von Freiheit aus dem Jenseits das Wasser abgraben.

Um Missverständnissen vorzubeugen: Ich sage nicht, dass Freiheit aus dem Jenseits *genau* so oder *perfekt* analog funktionieren müsse wie in dem Gedankenspiel, das ich vorgeführt habe. Nein, das Gedankenspiel soll nur ein überschaubares Modell für Freiheit ausserhalb der natürlichen Ordnung bieten. Laut Modell wären unsere Entscheidungen an eine Art Übergehirn aus völlig anderem Stoff gebunden. Wir wären dann zwar Bürger zweier Welten (so wie die eingetankten Gehirne erstens Bürger im Tank wären und zweitens Bürger im Simulationscomputer), aber diese beiden Welten wären nicht so radikal verschieden wie in Kants berühmter oder berüchtigter Sicht der Dinge.<sup>43</sup> Das macht nichts; mit meinem Modell habe ich eine Möglichkeit vorgeführt, die vielleicht ein bisschen besser zu verstehen ist als die Möglichkeit, der sich Kant verschrieben hat. Wichtig ist nur eines: Sobald die Tür zu *einer einzigen* solchen Möglichkeit aufgestossen ist, entspannt sich die Lage; wo es eine Möglichkeit gibt, da gibt es auch zwei, drei, viele Möglichkeiten. Libets Experimente brauchen uns dann nicht mehr zu beunruhigen; selbst wenn sie sich *ad infinitum* und *ad libetum* vermehren sollten.<sup>44</sup>

*Anmerkungen*

- 1 Siehe Arendt [D]/2:728.
- 2 Das empfiehlt uns jedenfalls in anderem Zusammenhang Konrad Lorenz [SB]:20.
- 3 Die Abgebrühtheit zeigt sich in vielen Texten kompatibilistischer Coleur; überzeugende Versionen des Kompatibilismus haben in letzter Zeit Peter Bieri [URGF] und Ansgar Beckermann [SBDF] angeboten.
- 4 Der *locus classicus* ist Libet et al [ToCI] und Libet [UCIR]. Für neuere Überlegungen siehe Libet [DWHF], übersetzt als Libet [HWFw] und Libet [MT], viertes Kapitel.
- 5 Dieser Durchschnitt lag, um genau zu sein, bei 343 Millisekunden und kam bei Handbewegungen heraus, die ohne vorherige Planung vollzogen wurden (und denen Bereitschaftspotentiale vom "Typ II" vorausgehen). Siehe Libet et al [ToCI]:623, 635, 631 (Tabelle 2 C); siehe Libet [UCIR]:529, 532.
- 6 Wolf Singer z.B. nennt unsere Erfahrung, frei zu sein, eine Illusion und plädiert dafür, unsere Rechtspraxis zu überprüfen (Singer [VLUF]:50, 63/4). Dass dies die zentrale Botschaft seiner Überlegungen ist, zeigt schon deren Untertitel: "Wir sollten aufhören, von Freiheit zu sprechen" (Singer [VLUF]:30). – Libet war und blieb in dieser Sache weitaus vorsichtiger, indem er darauf verwies, dass sich der freie Wille möglicherweise noch mit einem Veto ins Geschehen einmischen könnte; denn nachdem der Versuchsperson bewusst wird, dass sie sich entschieden hat, bleiben ihr vor Umsetzung der Handlung immer noch ca. 100 Millisekunden, und genau in dieser Zeitspanne könnte laut Libet die Versuchsperson ihre Handlung vielleicht noch verhindern, sozusagen in letzter Millisekunde. Siehe Libet et al [ToCI]:641 und Libet [UCIR]:529, 536-539.
- 7 Beispielsweise bezweifelt Kristian Köchy (unter Berufung auf Henrik Walter), ob "ein Willensakt tatsächlich, wie bei Libet vorausgesetzt, ein zeitlich genau umschriebenes und singuläres Ereignis ist. Fasst man den Willensbildungsprozess anders auf, nämlich als *komplexes und langwieriges Geschehen*, [...] dann ist die im Experiment von den Probanden gemachte Aussage kein Indiz mehr für die gerade gefallene Willensentscheidung. Diese ist vielmehr bereits in dem Moment gefallen, in dem die Versuchsperson einwilligt, der Instruktion des Experimentators zu folgen" (Köchy [WKNN]:157, mein Kursivdruck). – Nun scheint die letzte Behauptung dieses Zitats nicht recht zu dem zu passen, was ich in dessen Mitte kursiv hervorgehoben habe; denn die Einwilligung der Versuchsperson könnte sehr wohl als "zeitlich genau umschriebenes und singuläres Ereignis" erfolgen, und dann liesse sich Libets Test vielleicht auch darauf anwenden. (Wie das? Mit einer vorgeschalteten Messung: Nach Installation der Elektroden, nach einer Beschreibung der durchzuführenden Uhrbeobachtung und Handbewegung hört die Versuchsperson plötzlich folgendes: "Wenn Sie also am soeben lang und breit beschriebenen Experiment teilnehmen wollen, so drücken sie bitte jetzt *diese* Taste". Vielleicht lässt sich schon dann ein Bereitschaftspotential nachweisen, das sich aufbaut, bevor der Versuchsperson bewusst wird, dass sie die Taste drücken und also danach am eigentlichen Experiment teilnehmen will?) – Diese Kritik trifft natürlich nicht den Beginn des Zitats, wo Kristian Köchy mit Recht darauf verweist, dass viele unserer Willensentscheidungen weniger punktuell ablaufen als diejenigen, die Libet untersucht hat. Libet selbst hat von Anbeginn diesen Ausweg für die Freunde der Freiheit im Blick gehabt, siehe Libet et al [ToCI]:641. – Norbert Meuter unternimmt in [NKF] einen grundstürzenden Versuch in diese Richtung (der von den Gedankenspielen des vorliegenden Abschnitts nicht getroffen werden kann, so radikal ist er). Siehe Anmerkung 18.
- 8 So schreibt Libet in jener Mischung aus Vorsicht und Kühnheit, wie sie für Naturwissenschaftler so bezeichnend ist: "The present evidence for the unconscious initiation of a voluntary act of course applies to one very limited form of such acts. However, the simple voluntary motor act studied here has in fact often been regarded as an incontrovertible and ideal example of a fully endogenous and 'freely voluntary' act. [...] It [...] *invites the extrapolation* that other relatively 'spontaneous' voluntary acts, performed without conscious deliberation or planning, may also be initiated by cerebral activities proceeding unconsciously" (Libet et al [ToCI]:640/1, mein Kursivdruck; ähnlich Libet [HWFw]:281/2).
- 9 Spätere Experimentatoren sind in verschiedenen Richtungen über die Beschränkungen in Libets ursprünglichem Experiment hinausgegangen. Hier nur ein Beispiel: Statt (wie bei Libet) die Versuchspersonen vor die Entscheidung zu stellen, spontan *eine* Art von Handbewegung durchzuführen, haben Haggard und Eimer ihren Versuchspersonen die Wahl zwischen *zwei verschiedenen* willkürlichen Bewegungen vorgesetzt; mit demselben (für die Freiheit unangenehmen) Ergebnis. Siehe Haggard / Eimer [oRbB]:129, 131/2.
- 10 Ich spreche hier allgemein von "neuronaler Aktivität" (statt punktgenauer von "Bereitschaftspotentialen"), weil ich mich nicht auf neurophysiologische Details festlegen möchte. Im folgenden wird es auf solche Details nicht ankommen. Daher werde ich von nun an *pars pro toto*

## Die Diebe der Freiheit

wieder mit Bereitschaftspotentialen weiterarbeiten, obwohl deren Ermittlung streng genommen viel zu wünschen übrig lässt: Denn anders als man nach flüchtigem Blick auf die Debatte um Libet meinen möchte, können die Neurophysiologen dem Gehirn bislang bei *einzelnen* Entscheidungen nicht mithilfe des Bereitschaftspotentials auf die Schliche kommen; die zugehörigen Messkurve versinkt nämlich im aleatorischen Rauschen der elektrischen Gesamtaktivität des Gehirns. Um dieses Rauschen zu neutralisieren, messen die Neurophysiologen pro Versuchsdurchgang die Bereitschaftspotentiale bei insgesamt vierzig Entscheidungen des Probanden und bilden dann (für jeden Zeitpunkt) den Mittelwert der vierzig Potentialwerte (siehe Libet et al [ToCI]:624, 632-634, Libet et al [UCIR]:533). Erst nach dieser Durchschnittsbildung entsteht eine anständige und aussagekräftige Kurve. Dass der jähe Anstieg solcher Durchschnittskurven nicht so viel hergibt, wie der jähe Anstieg einer *einzelnen* Potentialkurve hergeben würde, kann man sich leicht klarmachen: Wenn eine von vierzig Kurven eine Ausreisserin ist und extrem früh von Null auf höhere Werte ansteigt, während die anderen 39 Kurven dies viel später tun, so zeigt die Durchschnittskurve einen künstlich frühen Anstieg – genau dort, wo die Ausreisserkurve ansteigt. (Schon aus Libets eigenen Daten lässt sich mit etwas Geduld ablesen, dass die gemittelten Bereitschaftspotentialkurven viel früher ansteigen, wenn die Versuchsperson nur eine einzige der vierzig Entscheidungen langfristiger geplant statt spontan durchgeführt hat; dann entstehen gemittelte Potentialkurven vom Typ I statt vom Typ II oder III, siehe Libet et al [ToCI]:623, 632 und Libet et al [UCIR]:532). Mehr zu dieser Detailkritik bieten Trevena et al [CMPb], Miller [CMPC]; siehe auch Rösler [NKH]. Doch möchte ich diese berechnete innerwissenschaftliche Kritik hier nicht weiterverfolgen. Denn mich beunruhigt die Frage, wie es um die Freiheit steht, falls es den Neurophysiologen irgendwie gelingt, unseren Gehirnen bei deren Einzelentscheidungen auf die Schliche zu kommen – gleichgültig, ob ihnen das mithilfe von Potentialen gelingt, die an der Kopfoberfläche abgeleitet werden, oder mithilfe neuerer bildgebender Verfahren oder womit auch immer. Um es zu wiederholen: In meinem Text sollen von nun an Bereitschaftspotentiale als Stellvertreter für das verstanden werden, was uns die Neurophysiologie noch alles bieten kann.

- 11 Wer das Experiment unter kontrollierten Bedingungen durchführen will, wird viele Faktoren konstant halten müssen, die im Casino durcheinandergehen. Der Roulette-Spieler muss immer an derselben Stelle des Spieltischs stehen, so dass das Feld für Rot schräg links vor ihm liegt und das Feld für Schwarz gleich weit und im selben Winkel schräg rechts vor ihm; seine Jetons müssen sich immer an einem fixen Platz genau in der Mitte vor ihm befinden; keine Hindernisse dürfen ihm beim Setzen in die Quere kommen; usw. Um alle diese Schwierigkeiten zu umgehen, werde ich die Entscheidungen im nächsten Beispiel an Mausclicks koppeln.
- 12 Und dann könnten sie sich gegen einen Einwand wehren, den Lutz Wingert so formuliert hat: "Zum Handeln gehört ein Urteil: 'Es ist (jetzt) besser, das und das zu tun, als es zu unterlassen.' Die Probanden von Libet & Co. hatten allerdings keine Hinsicht, unter der es ihnen besser erschien, diese oder jene Taste zu drücken. Deshalb mußten sie im Labor Handlungen simulieren" (Wingert [GZ]:197).
- 13 Da Libets Methode bislang mit Handbewegungen besonders erfolgreich gewesen ist, scheint sich mir die Maustastenbedienung besonders gut zu eignen, um so etwas wie Kommunikation zu untersuchen. Bei längeren (geschriebenen oder gesprochenen) Sprachspielzügen der Probanden dürfte Libets Methode dagegen eher an ihre Grenzen stossen. Denn die Bereitschaftspotentiale, die wir messen können, sind weit davon entfernt, spezifisch anzuzeigen, welche *Inhalte* der Proband zu kommunizieren sich anschickt. (Daher vermutlich setzt Libet in seinen extrapolierenden Ausführungen nur auf Bereitschaftspotentiale, die "*dem Beginn* des Sprechens oder Schreibens vorausgehen", siehe Libet [HFWF]:281, mein Kursivdruck. Das ist fast nichts, wenn man bedenkt, dass man schon vor dem Traualtar zwei dramatisch verschiedene Sprechakte *beginnen* kann!)
- 14 Es lohnt sich, genau zu durchdenken, was solche Versuchsprotokolle beweisen würden und was nicht. Nicht beweisen können sie die Behauptung, dass die Essensentscheidung des Opfers bereits feststünde oder gar prognostiziert werden könne, bevor der Gast den Schnellimbiss betritt und sich an den Computer setzt. Vielmehr steht jede einzelne Teilentscheidung immer erst unmittelbar vor ihrer Bewusstwerdung fest. Schon die jeweils übernächste Teilentscheidung verschwindet gleichsam hinter dem neuronalen Horizont, und da die Gesamtentscheidung erst mit der allerletzten Teilentscheidung komplett wird, wären wir im Lichte jener Versuchsprotokolle weit davon entfernt, im voraus prognostizieren zu können, an welches Ende des Entscheidungsbaumes der Gast klettern wird. – Ohnehin kann man darüber streiten, ob die anvisierte Zerlegung einer Entscheidung in eine Serie von Einzelentscheidungen nicht doch die Phänomenologie unserer lebensechten Entscheidungen allzu sehr zerrüttet.
- 15 So argumentiert z.B. Lutz Wingert (mit Blick auf Libets ursprüngliches Experiment). Siehe Wingert [GZ]; den Titel seines Aufsatzes habe ich mir für den Auftakt meines nächsten Satzes ausgeliehen.
- 16 So ähnlich argumentiert Libet [UCIR]:530, 532, 536, 539.

- 17 *Einwand*. Der Auslandsstipendiat hat die eigentliche Entscheidung schon in dem Moment gefällt, wo er sich auf unsere Spielregeln einlässt und damit einverstanden ist, seine Wahl auf zwei Wohnungen zu beschränken, deren Miete wir ihm schenken. (Vergl. das Köchy-Zitat aus Anmerkung 7). *Erwiderung*. Das ausgerechnet soll die Freiheit retten? Wer wagt es, gegen meine Prognose zu wetten, dass so gut wie jeder Auslandsstipendiat die geschenkte Miete wird haben wollen? (Ich werde natürlich nicht gegen den Auslandsstipendiaten oder seine Freunde, Verwandten und Bekannten wetten, denn wenn der Einsatz hoch ist, könnte das den Stipendiaten in die falsche Richtung locken; solche Störeinflüsse müssen wir ausblenden).
- 18 Falls Libets Methode nicht steckenbleibt, dürften sich immer noch viele verschiedene Möglichkeiten für die Verteidigung der Freiheit empfehlen. Ich werde im folgenden nur eine dieser Möglichkeiten durchspielen; sie ist meines Wissens neu. Attraktive Ressourcen für eine andere solche Möglichkeit hat Norbert Meuter bei Henri Bergson aufgespürt: Wenn man echte Entscheidungen ganz anders auffasst, nämlich fließender, als wir's im Zeitalter der Entscheidungsbäume gewohnt sind, dann könnte Libets Methode keine echten Entscheidungen erfassen, sondern schlimmstenfalls solche Surrogate wie aus meinem fiktiven Dialog zwischen Gast und Maus. Siehe Meuter [NKF], dritter Abschnitt.
- 19 Das Gedankenspiel der Gehirne im Tank ist von Hilary Putnam bekannt gemacht worden. Siehe [RTH]:5 ff.
- 20 Zumindest gibt es dann kein indeterministisches Schlupfloch für die Willensfreiheit, wie es z.B. Pascual Jordan aus der Quantenphysik hat entnehmen wollen. Siehe Jordan [PZJ]:126. – Sicherheitshalber möchte ich darauf aufmerksam machen, dass der newtonisch-deterministische Ausgang aller Experimente nicht auf die Feststellung hinausläuft, der durchgängige Determinismus der Welt sei *beobachtet* worden. Nein, so ein Determinismus geht weit über das hinaus, was sich beobachten lässt; er ist und bleibt ein theoretisches *Postulat* der Experimentatoren, das allerdings gut gerechtfertigt sein kann.
- 21 Genauer gesagt, geht beim Gespräch zwischen eingetankten Gehirnen folgendes vor sich. Das sprechende Gehirn produziert elektronische Outputsignale, die für seine (fehlende) Zungenmuskulatur bestimmt wären. Sie werden in ein Kommunikationsmodul des Rechners geleitet und dort in Bit-Elektronen umgewandelt, die auf die zugehörigen Bit-Zungenmuskelfasern einwirken und deren Code-Bestandteile in andere Speicherplätze verschieben – eine simulierte Muskelkontraktion. In der Folge werden viele andere Codes ihren Speicherort wechseln (simulierte Luftwellen), und von gewissen Bit-Trommelfellen werden Bit-Elektronen auf Bit-Nervenbahnen in Richtung von Bit-Gehirnen wandern, wobei sie vom eingangs erwähnten Kommunikationsmodul aufgeschnappt werden. Das wandelt sie in echte Elektronen um und leitet diese auf den akustischen Input-Bahnen in die eingetankten Gehirne der Zuhörer und des Sprechers. (Anderswo mehr zu derartigen Details, siehe Olaf Müller [MSS]:§4.2 bis §4.7).
- 22 Viele Philosophen bestreiten, dass *Gehirne* etwas sagen oder tun können oder dass sie sich entscheiden können (siehe z.B. Kemmerling [IMGM] und Wingert [GZ]:198/9). Doch da sich meine Gegner dieser Redeweise befehligen, wird sie auch mir erlaubt sein. Und tatsächlich, in einer Neurobiologie nach Geschmack von Gerhard Roth treffen Aussagen zu wie "Nicht mein bewusster Willensakt, sondern mein Gehirn hat entschieden!", während laut Roth folgendes nicht richtig ist: "Mein Arm und meine Hand haben nach der Kaffeetasse gegriffen, weil *ich* dies so gewollt habe" (Roth [WDHR]:73, Kursivdruck im Original). – Natürlich würden Philosophen wie Kemmerling nicht sagen, dass *der Arm* nach der Kaffeetasse greift (sondern die Person); und erst recht nicht würden sie sagen: "Der Willensakt hat entschieden", denn auch das tut in ihrer Sicht wieder nur die Person.
- 23 Hilary Putnams Einsicht, siehe [RTH]:12-14, 3-5.
- 24 Den Vorschlag hat Hilary Putnam ins Spiel gebracht (Hilary Putnam [RTH]:14), ohne ihn vor konkurrierenden nicht-magischen Vorschlägen auszuzeichnen (Hilary Putnam [RTH]:14/15). Warum diese konkurrierenden Vorschläge nicht plausibel sind, habe ich anderswo begründet, siehe Olaf Müller [MSS]:§4.14 bis §4.17.
- 25 Plädoyers fürs Prinzip des Wohlwollens (in ganz unterschiedlichen Fassungen) bieten Quine [WO]:58/9, Davidson [CToT]:316, Williamson [PISa]:137-147. Anderswo habe ich das Prinzip in aller Ausführlichkeit auf die Tanksprache angewandt, siehe Olaf Müller [MSS]:§3.3 bis §3.9.
- 26 Hier setze ich voraus, dass mentale Vokabeln im Gegensatz zu physischen Vokabeln nicht umzuinterpretieren sind, wenn sie von der Tanksprache in unsere Sprache übertragen werden. Ich habe das anderswo genauer begründet, siehe Olaf Müller [MSS]:§25.12 bis §25.15, siehe auch Olaf Müller [J]:142-144. Mehr dazu im nächsten Abschnitt.
- 27 Der Zusatz in Klammern war nötig im Lichte der Anmerkung 22.
- 28 An dieser Stelle muss man sehr vorsichtig sein. Dass wir nicht in *derselben* misslichen Lage stecken können, dass wir also keine Gehirne im Tank sind, lässt sich zwar nicht empirisch feststellen, aber es

## Die Diebe der Freiheit

- lässt sich apriori beweisen. (Den bahnbrechenden Beweis hat als erster Hilary Putnam gesehen, siehe Hilary Putnam [RTH]:12-17. Ich habe anderswo Verbesserungen des Beweises vorgeschlagen, siehe Olaf Müller [HPAS], Abschnitte 18-21). Doch der Beweis ist nicht stark genug, um zu entscheiden, ob wir nicht in einer *ähnlichen, analogen* Lage wie die Gehirne im Tank stecken könnten. Siehe Olaf Müller [MSS], Abschnitt 15 (insbesondere §15.6); Abschnitte 19-23 (insbesondere §23.5 ff). Weitere Texte zum Thema der Gehirne im Tank habe ich ins Netz gestellt unter [Www.GehirnImTank.De](http://www.GehirnImTank.De).
- 29 Wer dagegen *seit kurzem* im Tank steckt, spricht immer noch unsere Sprache, die er ja unter normalen Umständen gelernt hat, mithilfe von Kausalketten zu den echten Dingen ausserhalb des Computers. (Auf dieser Tatsache beruht meine Antwort auf den letzten Einwand aus Abschnitt III).
  - 30 Genauer gesagt, kann meiner Ansicht nach folgende Behauptung widerlegt werden: "Mentale Ausdrücke bedeuten begrifflich (d.h. analytisch) dasselbe wie bestimmte Ausdrücke für natürliche Vorgänge im Gehirn". Die Widerlegung einer solchen Behauptung habe ich exemplarisch vorgeführt in Olaf Müller [J]:146-149. – Da sich Libet (anders als die Radikalen aus der Neurophysiologie) diese unvorsichtige Behauptung nicht auf die Fahnen schreibt, fällt er meiner Widerlegung auch nicht zum Opfer. Er hält vorsichtigerweise fest: "Conscious subjective experience, in this case an awareness of the endogenous urge or intention to move, is a primary phenomenon; *it cannot be defined in an a priori way, by recourse to any externally observable physical event*" (Libet et al [UCIR]:532; mein Kursivdruck).
  - 31 Dass unter diesen Vorzeichen sowieso keine plausible Übersetzung entstehen kann, lehrt auch ein kleiner Blick auf die Pragmatik: Wir und die Gehirne im Tank benutzen den Ausdruck "entscheiden" nicht immer zum Zweck der Beschreibung (sozusagen im Regierungsbereich von Freges Urteilsstrich). Vielmehr drücken wir mit seiner Hilfe oft ganz andersartige Sprechakte aus, etwa wenn es nach dem Verfahren heisst: "Das Hohe Gericht entscheidet ..." Dies zusätzliche Problem meiner Gegner will ich hier nicht weiterverfolgen.
  - 32 Warum zeitliche Vokabeln der Tanksprache bei Übersetzung in unsere Sprache nicht verändert zu werden brauchen, habe ich anderswo angedeutet. Siehe Olaf Müller [MSS]:§22.16.
  - 33 Siehe Libet et al [ToCI]:625.
  - 34 Siehe Libet et al [ToCI]:630/1. Wie kann man sich dieses überraschende Ergebnis erklären? Hier ist eine Möglichkeit: Die Verarbeitung der visuellen Uhreindrücke kostet Zeit, genau wie alle anderen Verarbeitungsprozesse im Gehirn; in dem Moment, wo diese Eindrücke ins Bewusstsein gelangen, sind sie also schon veraltet – und veraltete Zifferblatteindrücke zeigen einen zu frühen Zeitpunkt an.
  - 35 Siehe Libet et al [ToCI]:631, 634, 637/8 sowie Libet [UCIR]:529, 534. Für diese Korrektur spricht die – bisher von mir nicht erwähnte – empirische Tatsache, dass andere objektiv überprüfbare Zeitprotokolle à la Libet ebenfalls veraltete Zeitpunkte liefern, und zwar ungefähr gleich stark veraltete; hierfür hat Libet seine Versuchspersonen nach dem frühesten Augenblick gefragt, in dem ihnen die tatsächliche Ausführung der Handbewegung bewusst wurde. Siehe Libet et al [ToCI]:627/8, 631, 639.
  - 36 Dennett hat eine scharfsinnige Kritik an Libets Hauptergebnis vorgetragen, die ebenfalls mit der Dauer von Signalübertragungen (und der Verarbeitungsgeschwindigkeit im Gehirn) zu tun hat, siehe Dennett [FE]:231-236. Doch Dennett verzichtet dort darauf, die zeitlichen Korrekturen zu berücksichtigen, die Libet ins Auge fasst.
  - 37 Dass sich aus Libets Experimenten nichts darüber ableiten lasse, ob unsere bewussten Willensentscheidungen wirklich von unbewussten Bereitschaftspotentialen *verursacht* sind, sollte man meiner Ansicht nach besser nicht mit Anleihen bei Humes Ursachenskeptizismus begründen, wie es z.B. Kristian Köchy versucht ([WKNN]:157/8). Denn das wäre ein Dummdumgeschoss, das mit einem einzigen Schuss die gesamte kausale Arbeit der Naturwissenschaft zerfetzen müsste. Mein Argument soll präziser treffen: Ich habe anhand der eingetankten Wissenschaftler einen konkreten Fall vorgeführt, wo Libets Evidenzen zu fehlerhaften Kausalbehauptungen führen – das braucht nicht unser Zutrauen in Kausalität überhaupt zu erschüttern; es erschüttert nur unser Zutrauen in die Kausalbehauptungen der Freiheitsgegner. Immerhin haben wir genau dieselben Evidenzen wie die eingetankten Wissenschaftler (oder sogar weniger).
  - 38 Soll man hierauf etwa erwidern, dass jene Möglichkeit zwar denkbar sei, aber sehr, sehr unwahrscheinlich? Das hilft nicht, denn die Wahrscheinlichkeiten, auf die sich diese Erwidern stützen müsste, entstammen wieder nur dem Bereich unserer empirisch gegebenen Natur; ausserhalb der Natur greifen sie nicht.
  - 39 Singer [VLUF]:57.
  - 40 So fragt der Naturalist Ansgar Beckermann in ähnlichem Zusammenhang: "Wie *kann* etwas Nichtkörperliches physische Wirkungen haben?" (Beckermann [SBDF]:20, mein Kursivdruck). Das klingt so, als hielte Beckermann die fragliche Möglichkeit für unverständlich.

- 41 Libet erwähnt im Vorübergehen passende Indizien dafür, siehe Libet et al [ToCI]:636; siehe auch Libet [UCIR]:535,
- 42 Siehe oben Anmerkung 20. – Neuauflage des Einwandes, diesmal mit Bezug auf den Materie-Erhaltungssatz (auf den sich in ähnlichem Zusammenhang z.B. Beckermann beruft, siehe Beckermann [SBDF]:20): Wird dieser Satz nicht genau in dem Augenblick verletzt, wo sich die eingetankte Entscheidung in Bit-Elektronen verwandelt, die von aussen in die Bit-Welt eingespeist werden? – *Antwort:* Das hängt abermals vom Simulationsalgorithmus ab; die Sache lässt sich auch ohne Verletzung des Materie-Erhaltungssatzes programmieren. Denn die fraglichen Bit-Elektronen können schon an passender Stelle – und (dem Energie-Erhaltungssatz zuliebe) mit geeigneter kinetischer Bit-Energie – abgespeichert sein, bevor sie vom Algorithmus als bitneuronale Signale instrumentalisiert werden, die zu den Bit-Muskeln reisen.
- 43 Die intelligible Welt, in der wir nach Kant ebenso Bürger sind wie in der Welt der Erscheinungen, ist weder räumlich noch zeitlich organisiert (Kant [KRV]: A 539/40, B 567/8). Die höhere Wirklichkeit aus meinem Modell ist zwar aus Sicht des jeweiligen Sprechers nicht räumlich organisiert (sondern sozusagen nur analog zum Raum), aber sie bleibt zeitlich organisiert. Trotz dieser und anderer Unterschiede bestehen viele Gemeinsamkeiten zwischen Kants Vorschlag und meinem Modell. Der epistemische Status von Aussagen über transzendente Freiheit liegt unter der Schwelle echten Wissens, anders als bei Aussagen zur durchgehenden Kausalität im Reich der Erscheinungen (Kant) bzw. im Reich dessen, wovon unsere empirische Naturwissenschaft handelt. Wir wissen nach allen Anstrengungen nur, dass Freiheit in keinen Widerspruch zur durchgehenden Naturkausalität geraten muss (Kant [KRV]: A 536, 558, B 564, 586). Wer um der Moral willen auf Freiheit nicht verzichten mag oder kann, der ist berechtigt oder gar genötigt, sie zu postulieren (Kant [KRV]: A 533, B 561). Übertragen wir das in meine Redeweise. Wenn sich die Libet-Experimente wirklich auf all unser Tun ausbreiten sollten, dann können wir trotzdem unsere starke Intuition retten, wir seien in unseren Entscheidungen nicht durchs nackte Naturgeschehen determiniert. Wie das? Einfach: Wir bringen dann unser erschüttertes Meinungssystem wieder ins Überlegungsgleichgewicht, indem wir postulieren, dass wir in einer analogen Lage stecken wie die Gehirne im Tank, deren Entscheidungen nicht von dem determiniert werden, was sie "Naturablauf" nennen. Hier ist eine weitere Parallele meiner Überlegungen zu denen von Kant: Gott, Freiheit und Unsterblichkeit sind für Kant die zentralen Fragen der Metaphysik (Kant [KRV]: B 7). Überraschenderweise lassen sich *alle drei Fragen* mithilfe der Methode beleuchten, die ich hier nur anhand von Freiheit vorgeführt habe. Auf die Frage der Unsterblichkeit habe ich meine Methode bereits angewandt (siehe Olaf Müller [JJ]). Dass sie sich auch auf die Frage nach Gott übertragen lässt, kann man sich ideengeschichtlich klarmachen. Der böse Wissenschaftler, der den Simulationscomputer programmiert hat und den ich hier nur hinter den Kulissen habe mitspielen lassen, geht auf Descartes' bösen Täuscherdämon zurück, der seinerseits als Echo scholastischer Debatten über die Frage angesehen werden kann, ob der allgütige Gott in stände wäre, uns umfassend zu täuschen, wenn Er wollte.
- 44 Dieser Aufsatz (insbes. in Abschnitt III bis V) geht auf einen Vortrag zurück, den ich unter der Überschrift *Freiheit aus dem Jenseits* am 9.11.2002 auf der Tagung "Determinismus und Freiheit in der Physik und der Philosophie" in Iffeldorf gehalten habe und am 20.1.2006 im Humanprojekt an der Berlin-Brandenburgischen Akademie der Wissenschaften. Ich danke allen Teilnehmern der beiden Diskussionen für saftige Kritik und schwierige Fragen, die überhaupt nur in meine Seele dringen konnten dank der Einladung zwischen die Stühle der Disziplinen von Harald Lesch, Wilhelm Vossenkuhl, Bettina Walde (Iffeldorf) bzw. Volker Gerhardt und Julian Nida-Rümelin (BBAW): Mein herzlicher Dank an alle fünf. – Annika Keyzers danke ich dafür, dass sie mich gegen meinen freien Willen gezwungen hat, Libet & Co. zu lesen; Dank an Jürgen Müller für Mühen mit Exegese und Bibliographie sowie an Sabine Hassel fürs zuverlässige Abschreiben der Bänder. Schliesslich danke ich den Mitgliedern meines wissenschaftsphilosophischen Kolloquiums, die mir in allerletzter Sekunde geholfen haben einzusehen, welche Debatte ich besser nicht führen sollte.

### Literatur

Arendt, Hannah [D]/2: *Denktagebuch. 1950-1973. Zweiter Band.* (Ursula Ludz / Ingeborg Nordmann (eds); München: Piper, 2002).

Bieri, Peter [URGF]: "Untergräbt die Regie des Gehirns die Freiheit des Willens?" In Gestrich et al (eds) [FOUW]:20-36.

## Die Diebe der Freiheit

- Beckermann, Ansgar [SBDF]: "Schliesst biologische Determiniertheit Freiheit aus?" In Hermanni et al (eds) [FUW]:19-32.
- Davidson, Donald [CToT]: "A coherence theory of truth and knowledge". In LePore (ed.) [TI]:307-319.
- Dennett, Daniel C. [FE]: *Freedom evolves*. (New York: Viking, 2003).
- Elsner, Norbert / Lüer, Gerd (eds) [GSG]: *Das Gehirn und sein Geist*. (Göttingen: Wallstein, 2000).
- Ganten, Detlev / Gerhardt, Volker / Nida-Rümelin, Julian [NF]: *Die Naturgeschichte der Freiheit*. (Berlin: de Gruyter, 2006).
- Geyer, Christian (ed) [HW]: *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente*. (Frankfurt: Suhrkamp, 2004).
- Haggard, Patrick / Eimer, Martin [oRBB]: "On the relation between brain potentials and the awareness of voluntary movements". *Experimental Brain Research* 126 (1999), pp. 128-133.
- Hermanni, Friedrich / Koslowski, Peter (eds) [FUW]: *Der freie und der unfreie Wille – Philosophische und theologische Perspektiven*. (Paderborn: Fink, 2004).
- Jordan, Pascual [PZJ]: *Die Physik des 20. Jahrhunderts. Einführung in den Gedankeninhalt der modernen Physik*. (Braunschweig: Vieweg, fünfte Auflage 1943).
- Kant, Immanuel [KRV]: *Kritik der reinen Vernunft*. Nach der ersten und zweiten Original-Ausgabe neu herausgegeben von Raymund Schmidt. (Hamburg: Meiner, 1926; durchgesehener Nachdruck 1976). [Erschien zuerst 1781 bzw. 1787 (Riga: Hartknoch); ich zitiere die Randnummern der Meiner-Ausgabe, die sich auf die erste (A) bzw. zweite Auflage (B) beziehen].
- Kemmerling, Andreas [IMGM]: "Ich, mein Gehirn und mein Geist: Echte Unterschiede oder falsche Begriffe?" In Elsner et al (eds) [GSG]:221-243.
- Köchy, Kristian / Stederoth, Dirk (eds) [WaIP]: *Willensfreiheit als interdisziplinäres Problem*. (Freiburg: Karl Alber, 2006).
- Köchy, Kristian [WKNN]: "Was kann die Neurobiologie nicht wissen? Bemerkungen zum Rahmen eines Forschungsprogramms". In Köchy et al (eds) [WaIP]:145-164.
- LePore, Ernest (ed) [TI]: *Truth and interpretation. Perspectives on the philosophy of Donald Davidson*. (Oxford: Blackwell, 1986).
- Libet, Benjamin [DWHF]: "Do we have a free will?" *Journal of Consciousness Studies* 6 No. 8/9 (1999), pp. 47-57.
- Libet, Benjamin [HWFw]: "Haben wir einen freien Willen?" In Geyer (ed) [HW]:268-289. [Erschien zuerst englisch als Libet [DWHF]; deutsch von Jürgen Schröder].
- Libet, Benjamin [MT]: *Mind Time. Wie das Gehirn Bewusstsein produziert*. Aus dem Amerikanischen von Jürgen Schröder. (Frankfurt am Main: Suhrkamp, 2005). [Erschien zuerst englisch 2004].

- Libet, Benjamin [UCIR]: "Unconscious cerebral initiative and the role of conscious will in voluntary action". *Behavioral and Brain Sciences* 8 (1985), pp. 529-539.
- Libet, Benjamin / Gleason, Curtis A. / Wright, Elwood W. / Pearl, Dennis K. [ToCI]: "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act". *Brain* 106 (1983), pp. 623-642.
- Lorenz, Konrad [SB]: *Das sogenannte Böse. Zur Naturgeschichte der Aggression.* (München: Deutscher Taschenbuch Verlag, 1983). [Erschien zuerst 1974].
- Meuter, Norbert [NKF]: "Natur und Kultur der Freiheit". In Ganten et al (eds): [NF]. [In diesem Band].
- Miller, Jeff / Trevena, Judy Arnel [CMPC]: "Cortical Movement Preparation and Conscious Decisions: Averaging Artifacts and Timing Biases." *Consciousness and Cognition* 11 (2002), pp. 308-313.
- Müller, Olaf [HPAS]: *Hilary Putnam und der Abschied vom Skeptizismus oder Warum die Welt keine Computersimulation ist. Wirklichkeit ohne Illusionen, Band 1.* (Paderborn: Mentis, 2003).
- Müller, Olaf [JJ]: "Jenseits. Eine metaphysische Provokation für Naturalisten". In Sukopp et al (eds) [N]:137-154. [Im Erscheinen].
- Müller, Olaf [MSS]: *Metaphysik und semantische Stabilität oder Was es heisst, nach höheren Wirklichkeiten zu fragen. Wirklichkeit ohne Illusionen, Band 2.* (Paderborn: Mentis, 2003).
- Putnam, Hilary [RTH]: *Reason, truth and history.* (Cambridge: Cambridge UP, 1981).
- Quine, Willard Van Orman [WO]: *Word and object.* (Cambridge / Mass.: MIT Press, 1960).
- Rösler, Frank [NKH]: "Neuronale Korrelate der Handlungsausführung im Elektroenzephalogramm: Agieren wir vor oder nach dem bewussten Entschluß zu handeln? - Zur Validität der Experimente von Libet (1983)". In Köchy et al (eds) [WaIP]:165-190.
- Roth, Gerhard [WDHR]: "Worüber dürfen Hirnforscher reden – und in welcher Weise?" In Geyer (ed) [HW]:66-85.
- Singer, Wolf [VLUF]: "Verschaltungen legen uns fest: Wir sollten aufhören, von Freiheit zu sprechen". In Geyer (ed) [HW]:30-65.
- Sukopp, Thomas / Vollmer, Gerhard (eds) [N]: *Naturalismus: Positionen, Perspektiven, Probleme.* (Tübingen: Mohr Siebeck, 2007).
- Trevena, Judy Arnel / Miller, Jeff [CMPb]: "Cortical Movement Preparation before and after a Conscious Decision to Move." *Consciousness and Cognition* 11 (2002), pp. 162-190.
- Williamson, Timothy [PISa]: "Philosophical 'intuitions' and scepticism about judgement". *Dialectica* 58 No.1 (2004), pp. 109-153.
- Wingert, Lutz [GZ]: "Gründe zählen. Über einige Schwierigkeiten des Bionaturalismus". In Geyer (ed) [HW]:194-204.