

JOURNAL FOR THE PHILOSOPHY OF NATURE

*philosophia
naturalis*

50 / 2013 / I

Herausgeber / Editors Andreas Bartels
 Olaf L. Müller
 Manfred Stöckler
 Marcel Weber

Beirat / Editorial Board Werner Diederich (Hamburg)
 Michael Esfeld (Lausanne)
 Don Howard (Notre Dame)
 Andreas Hüttemann (Köln)
 Bernulf Kanitscheider (Gießen)
 James Lennox (Pittsburgh)
 Holger Lyre (Magdeburg)
 Felix Mühlhölzer (Göttingen)
 Friedrich Steinle (Berlin)
 Eckart Voland (Gießen)
 Gerhard Vollmer (Braunschweig)

persönliches Exemplar,
nur gemäß den Richtlinien
der Zeitschrift zu verwenden

KLOSTERMANN

Olaf L. Müller

Verschmierte Spuren der Unfreiheit

Wissenschaftsphilosophische Klarstellung zu angeblichen Artefakten bei Benjamin Libet

Zusammenfassung

Benjamin Libets bahnbrechende Experimente zur Willensfreiheit aus den Achtziger Jahren des vorigen Jahrhunderts gehören längst zu den Klassikern der Experimentalkunst. Aus *philosophischer* Perspektive sind sie oft kritisiert worden; unabhängig davon ist zu Beginn des neuen Jahrtausends ein *technischer* Einwand gegen Libet prominent geworden, der von Freiheitsfreunden gerne zitiert wird. Er geht auf Judy Trevena und Jeff Miller zurück und besteht in dem Vorwurf, dass Libet mit seiner Berechnung der *gemittelten* Bereitschaftspotential-Kurven die wahren Verhältnisse verschmiere und dabei aus mathematischen Gründen den Anstiegs-Zeitpunkt des Bereitschaftspotentials künstlich nach vorne verschiebe – das ist der Vorwurf des sog. Verschmierungsartefakts („smearing artifact“). Ich zeige, dass der Vorwurf nicht sticht. Er beruht auf einem Missverständnis der Schlussmethode Libets (über die sich Libet nicht explizit geäußert hat). Laut meiner wissenschaftstheoretischen Rekonstruktion zieht Libet weder deduktive noch induktive Schlüsse aus seinen Einzelbeobachtungen; er schließt abduktiv, d.h. er schließt auf die beste Erklärung seiner Einzelbeobachtungen. Wie eine genaue arithmetische Analyse zeigt, bieten die Alternativ-Erklärungen, unter denen sich das Verschmierungsartefakt bewahrheiten würde, schlechtere Erklärungen als Libets Erklärung – sie sind an den Haaren herbeigezogen und extrem unwahrscheinlich.

Abstract

Benjamin Libet's celebrated experiments concerning freedom (1983 etc.) elicited numerous attempts of *philosophical* repudiation. Ten years ago, however, Judy Trevena and Jeff Miller published a *technical* objection; they claim to have detected a „smearing artifact“ in Libet's calculations. This rests on a misunderstanding of Libet's methodology (about which he had not been quite explicit). In my reconstruction of Libet's argument, he draws an abductive inference to the best explanation. Now, Trevena's and Miller's objection does indeed lead to alternative explanations of Libet's measurements. These alter-

natives are *ad hoc* and extremely improbable (as can be seen by constructing them explicitly). They constitute worse explanations than the explanation offered by Libet.

I. Erinnerung an Libets klassisches Experiment

Jemand entscheidet sich spontan dazu, seinen rechten Zeigefinger zu bewegen. Bevor ihm diese Entscheidung bewusst wird, spielt sich in seinem Gehirn allerhand ab. Es ist nicht leicht, aus dem verschwommenen Grollen des neuronalen Trommelfeuers eindeutige Signale herauszulesen. Eine der vielen Methoden, die hierfür von den Neurophysiologen entwickelt worden sind, hat es in den Debatten über Willensfreiheit zu großer Berühmtheit gebracht: die Ermittlung des sogenannten Bereitschaftspotentials.¹ Auf der Schädeloberfläche der Versuchsperson werden an ganz bestimmten Stellen Elektroden angebracht, mit deren Hilfe sich blitzschnell elektrische Spannungsänderungen registrieren lassen (per Elektroenzephalographie, EEG). So entstehen Kurven, an denen man ablesen kann, wie sich die abgeleiteten elektrischen Potentiale im Laufe der Zeit ändern und wie sie sich z. B. kurz vor der spontanen Fingerbewegung der Versuchsperson aufbauen.

Die zeitliche Auflösung dieser Kurven ist hoch und liegt im Bereich von Millisekunden. Doch das Verfahren erlaubt keine scharfe räumliche Auflösung. Da die Elektroden an der Schädeloberfläche angebracht werden und nicht etwa am Ort des neuronalen Geschehens (nicht etwa im Innern des Gehirns), registrieren die Elektroden zu viele elektrische Geschehnisse, die sich im Gehirn gleichzeitig abspielen, sich also bei der Potentialmessung gegenseitig überlagern: das Ergebnis ist jedesmal ein undifferenziertes Rauschen und besagt für sich allein nicht viel. Um dieser Schwierigkeit zu entrinnen, haben die Neurophysiologen einen raffinierten Ausweg ersonnen. Zwar vermischen sich die elektrischen Geschehnisse, die mit der spontanen und bewussten Entscheidung der Versuchsperson korreliert sind, mit allerlei anderen elektrischen Geschehnissen, die sich in enger Nachbarschaft abspielen und die jede Messung stören. *Aber das neuronale Störfeuer aus der Nachbarschaft kann als zufällig betrachtet werden.* Und im statistischen Mittel müssten sich die störfürigen Zufallsausschläge (nach unten und oben) gegenseitig auslöschen. Das bedeutet: Wenn man vor vierzig spontanen Finger-

bewegungen ein und derselben Versuchsperson jedesmal den zeitlichen Verlauf des elektrischen Potentials ermittelt und aus diesen vierzig Kurven den Durchschnitt bildet, dann müsste in der Durchschnittskurve diejenige Größe sichtbar werden, die uns interessiert und die wirklich mit der spontanen Entscheidung zur Fingerbewegung zusammenhängt.

So verhält es sich in der Tat, wie Sie gleich sehen werden. Doch bevor ich an einige erstaunliche Messergebnisse erinnere, muss ich auf eine technische Komplikation aufmerksam machen: Der Durchschnitt der vierzig Potentialkurven ist erst dann eindeutig bestimmt, wenn wir festlegen, welche Zeitpunkte der vierzig Potentialkurven zusammengehören. Streng genommen bewegt unsere Versuchsperson den Finger zu vierzig verschiedenen Zeitpunkten; wir haben also vierzig getrennte Zeitabläufe, die sich nicht überlappen. Um den Durchschnitt zu bilden, müssen wir diese getrennten Abläufe künstlich synchronisieren, und hierbei kommt ein gewisses Element der Willkür ins Spiel (das uns allerdings nicht stark zu beunruhigen braucht).

Wir können z. B. vereinbaren, dass wir jeweils den Beginn der Fingerbewegung auf *0 Uhr* datieren. D. h., wir nennen jeweils denjenigen Zeitpunkt „0 Uhr“, an dem sich der rechte Zeigefinger der Versuchsperson zu bewegen beginnt. Erst im Lichte einer solchen Konvention können wir ausrechnen, wie groß das Potential zu „ein und demselben Zeitpunkt“ durchschnittlich gewesen ist. Hierbei identifizieren wir vierzig Zeitpunkte (die in Wirklichkeit nicht identisch sind) dadurch miteinander, dass sie allesamt z. B. exakt eine halbe Sekunde vor der jeweiligen Fingerbewegung liegen – jeweils eine halbe Sekunde vor Null.

Ohne derartige Konventionen lassen sich keine Durchschnitte berechnen. Wieviel Willkür steckt in diesem Verfahren? Nicht allzuviel. Zwar hängt es ausschließlich von den Experimentierenden ab, dass sie die vierzig getrennten Zeitverläufe immer mithilfe des Beginns der spontanen Fingerbewegung synchronisieren; das neuronale Geschehen auf seiten der Versuchspersonen diktiert den Experimentatoren kein bestimmtes Synchronisationsverfahren. Doch lässt sich das befolgte Synchronisationsverfahren *nachträglich* rechtfertigen. Wer sich auf völlig andere Synchronisationsverfahren stützt und z. B. als Nullpunkt den Zeitpunkt des siebzehnten Augenblinzeln der Versuchsleiterin festsetzt, wird keine deutlichen Durchschnittskurven erhalten.² Und es ist überraschend genug, dass ausgerechnet diejenigen Durchschnittskurven halbwegs deutliche Muster zeigen, die mithilfe des Beginns der

jeweiligen Fingerbewegung synchronisiert wurden – ein interessantes empirisches Ergebnis.³

Noch überraschender ist etwas anderes. Die Durchschnittskurven beginnen erstaunlich früh anzusteigen – etwas mehr als eine halbe Sekunde vor Null, vor Beginn der Fingerbewegung; genauer gesagt steigen sie 550 ms vor Null an.⁴

Wenn man diesen erstaunlich frühen Wert systematisch mit den Erlebnisberichten der Versuchspersonen vergleicht, kommt die größte Überraschung ans Licht. *Der Zeitpunkt der bewussten Entscheidung für die spontane Fingerbewegung liegt nämlich im Durchschnitt nur 200 ms vor Null.*⁵ Und das bedeutet offenbar: Im Durchschnitt beginnt der Anstieg des Bereitschaftspotentials ca. 350 ms vor demjenigen Zeitpunkt, den die Versuchspersonen als bewussten Augenblick ihrer spontanen Entscheidung angaben.⁶ Das ist das sensationelle Versuchsergebnis von Benjamin Libet, das vor 30 Jahren Wissenschaftsgeschichte geschrieben hat. Zwar gibt es inzwischen andere freiheitsgefährdende Versuchsergebnisse aus der Neurophysiologie, bei denen schlagkräftigere Methoden eingesetzt werden.⁷ Dennoch lohnt sich eine wissenschaftstheoretische Analyse der Libet-Resultate. Bis heute ist nicht eindeutig geklärt, was sie bedeuten. Welche Schlüsse können wir aus ihnen ziehen? Und mithilfe welcher Schluss-Logik?

Im kommenden Abschnitt will ich dazu einen Vorschlag machen; er drängt sich auf, ist aber in der Literatur meines Wissens nirgends mit der wünschenswerten Präzision explizit gemacht worden. Implizit lag er (so nehme ich an) den Überlegungen Libets zugrunde. Libet schloss auf die beste Erklärung seiner Messungen, vollzog also das, was oft als abduktiver Schluss bezeichnet wird (im Unterschied zum deduktiven und zum induktiven Schluss).

Aber selbst wenn meine wissenschaftsgeschichtliche Behauptung nicht stimmen sollte, gilt ihr *systematisches* Pendant: Libet wäre gut beraten gewesen, seinen Schluss so aufzubauen, wie ich es vorschlagen werde – abduktiv. Hätte er das deutlich genug herausgestrichen, so wäre eine Kritik an seiner Methode unterblieben, die in den letzten Jahren viel Eindruck gemacht hat und die ich im übernächsten Abschnitt III vorstellen werde. Ich werde sie in den Abschnitten IV bis VI wissenschaftsphilosophisch entkräften. Die Moral aus der Geschichte lautet: Ein wenig Wissenschaftsphilosophie hat noch niemandem geschadet.

II. Wissenschaftsphilosophische Rekonstruktion der Schlussweise Libets

Gegen Libets Überlegung haben viele Autoren mit den verschiedensten Gründen protestiert. Nur einen ihrer Gegenstände möchte ich in diesem Aufsatz vorführen und entkräften. Ich habe ihn deshalb für meine Betrachtung ausgewählt, weil ich den Denkstil reizvoll finde, dem er entspringt. Es ist der Denkstil von jemandem, der philosophische Abstraktion scheut und der stattdessen scharf auf konkrete Einzelheiten achtgibt, selbst wenn sie im Kleingedruckten stehen. So jemand will jeden Schritt seiner Gedanken nach klaren Regeln überprüfen, statt sich in tiefe philosophische Fragen verwickeln zu lassen wie etwa: Was ist Freiheit? Sind Freiheit und neuronaler Determinismus miteinander kompatibel? Wie weit müsste unser Freiheitsbegriff im Namen dieser Kompatibilität abgeschwächt werden? Und verträge sich das noch damit, wie wir moralische Verantwortlichkeit zuschreiben und ahnden? Wie lässt sich die normative Dimension der Handlungsgründe ohne Verluste einbetten in eine natürliche Welt, deren Teil wir sind, in der aber nur Ursachen und Zufälle vorkommen?

Solche Fragen sind wichtig, das bestreite ich nicht. Es sind die ewigen Fragen der Philosophie.⁸ Dennoch möchte ich sie diesmal an die Seite stellen, um den Blick für etwas anderes freizubekommen – für eine Debatte, die nicht weniger Scharfsinn erfordert als die ewigen Streitigkeiten der Philosophen, die sich aber meiner Ansicht nach eindeutig entscheiden lässt, anders als sonst in der Philosophie.

Die Debatte, die ich hier führen will, hat mit statistischen Methoden in den Naturwissenschaften zu tun, betrifft aber nur einen winzigen Ausschnitt aus diesem riesigen Minenfeld. Um sie ingangzubringen, möchte ich fragen, mit welchem Recht man aus bloßen Durchschnittsbetrachtungen eine Allaussage ableiten kann wie:

- (1) Bei allen Fingerbewegungen, deren Zeitpunkt eine Versuchsperson selber spontan und frei bestimmen kann, beginnt ca. 350ms vor der bewussten Entscheidung in ihrem Gehirn eine bestimmte elektrische Größe (das Bereitschaftspotential) langsam, aber deutlich anzusteigen.

Ob so eine Allaussage statistisch aus den Daten abgeleitet werden kann, hat natürlich u. a. mit Stichprobengröße, Signifikanz, Standardabweichung

chung usw. zu tun. Aber um diese Nettigkeiten aus dem Werkzeugkasten der Statistiker brauchen wir uns nicht zu scheren, denn Libets Experimente sind oft genug reproduziert worden und haben nahezu niemanden dazu gebracht, die Berechnungen der Neurophysiologen und deren statistische Interpretationen anzufechten.⁹

Selbst wenn man an dieser Front Ruhe gibt, bleibt die Frage nach der Berechtigung von Schlüssen wie (1) bestehen. Immerhin sollte es uns zu denken geben, *dass kein einziger Einzelfall der Allaussage (1) beobachtet worden ist*. Die elektrische Größe aus (1) lässt sich nicht einzeln beobachten, sie wird (wie Sie gesehen haben) nur im Durchschnitt aus vierzig Versuchsdurchläufen sichtbar; bei jeder Einzelmessung geht die fragliche elektrische Größe im statistischen Rauschen unter.

Das ist zwar auf den ersten Blick eine gute Nachricht für Freiheitsfreunde und gediegene Humanisten: Es ist nicht damit zu rechnen, dass sich mit Libets Mitteln jemals eine einzelne spontane Entscheidung vorhersagen lässt, bevor sie dem Entscheider bewusst wird. Denn die Kurve, die ca. 350 ms vor der bewussten Entscheidung langsam, aber deutlich ansteigt, existiert empirisch nur als Durchschnittskurve, also *post festum* nach vierzig Durchläufen; im Einzeldurchgang lässt sie sich nicht dingfest machen. Und das bedeutet in der Tat, dass uns kein Neurophysiologe à la Libet in die Karten gucken kann – nicht einmal, kurz bevor wir uns zur Fingerbewegung entscheiden.¹⁰

Mit dieser frohen Botschaft ist die Geschichte nicht zuende. Es mag beruhigen zu erfahren, dass unsere spontanen Fingerbewegungen nicht *ex ante* vorausgesagt werden können. Aber das lässt sich immer noch damit vereinbaren, dass sie *ex ante* (vor dem Zeitpunkt der bewussten Entscheidung) vorherbestimmt sind – ontologisch vorherbestimmt, nicht epistemisch oder prognostisch.¹¹

Ich will jetzt erklären, warum Libets Experimente den Schluss auf die ontologische Allaussage (1) nahelegen. Das ist kein statistischer oder induktiver Schluss, kein Schluss von beobachtbaren Einzelfällen auf ein allgemeines Gesetz. Es ist ein Schluss auf die beste Erklärung („inference to the best explanation“) – ein abduktiver Schluss. So ein Schluss führt genau wie induktive oder statistische Schlüsse über das hinaus, was beobachtet wurde; aber anders als sie postuliert er nicht einfach nur mehr Sachverhalte vom schon beobachteten Typ, sondern neuartige Sachverhalte, für die überhaupt noch keine Einzelfälle beobachtet wurden.

Die Taufe dieses Schlussmusters („abduction“) geht auf C. S. Peirce zurück.¹² Während der empiristischen Übertreibungen der ersten Hälfte des Zwanzigsten Jahrhunderts geriet es außer Mode. Erst von Gilbert Harman ist es wieder schmissig in die Debatte gebracht worden:

In making this inference one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis. In general, there will be several hypotheses which might explain the evidence, so one must be able to reject all such alternative hypotheses before one is warranted in making the inference. Thus one infers, from the premise that a given hypothesis would provide a “better” explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true (Harman, 1965, S. 89).

Meiner Ansicht nach beschreibt Harman hier exakt die Schlussmethode, an der sich Libet (implizit) orientiert hat. In seinem Fall läuft der Schluss auf die beste Erklärung so:

- (2) Empirische Einzelfakten: In der letzten Sekunde vor der Entscheidung einer beliebigen Versuchsperson, spontan zu einem frei gewählten Zeitpunkt den Finger zu bewegen, lässt sich an der Schädeloberfläche der Versuchsperson ein elektrisches Potential messen, dessen zeitlicher Verlauf aussieht wie undifferenziertes Rauschen.
- (3) Statistische Analyse: Der Durchschnitt dieser Kurven nach vierzig Versuchsdurchläufen ein und derselben Person liefert ein überraschend klares Bild. Das Durchschnittspotential steigt gut eine halbe Sekunde vor dem Nullzeitpunkt langsam, aber deutlich an.
- (4) Eine Erklärung des überraschenden Ergebnisses aus Schritt (3) lautet: Bei *jedem* Versuchsdurchlauf findet im Gehirn der Versuchsperson ein elektrischer Einzelprozess statt, der – ohne elektrische Störeinflüsse benachbarter Prozesse – an der Schädeloberfläche ungefähr denselben Potentialverlauf verursachen *würde*, den die berechnete Durchschnittskurve zeigt. Die benachbarten elektrischen Prozesse verlaufen unabhängig vom fraglichen Einzelprozess, sind mit ihm nicht korreliert und können daher als zufällige Prozesse betrachtet werden. Sie beeinflussen die einzelnen Versuchsdurchläufe so gravierend, dass jede einzelne Potentialmessung ein Ergebnis liefert, das wie zufällig aussieht. Doch im statistischen Mittel aus vierzig Versuchsdurchläufen heben sich

diese Störeinflüsse gegenseitig auf. Daher zeigt die Durchschnittskurve denjenigen Verlauf, den die gemessenen Einzelkurven hätten zeigen müssen, wären sie nicht von Nachbarprozessen gestört worden.

- (5) Eine andere Erklärung des überraschenden durchschnittlichen Faktums aus (3) lautet: Dass sich im Durchschnitt immer eine deutliche Potentialkurve zeigt, beruht auf einem gigantischen Zufall (nämlich auf den sich zufällig nicht exakt ausgleichenden Störprozessen aus der Nachbarschaft der Entscheidung).
- (6) Die Erklärung gemäß Schritt (4) ist besser als die gemäß Schritt (5). Andere Erklärungen für das überraschende statistische Faktum (3) ähneln in dieser Hinsicht eher der schlechteren Erklärung.
- (7) Also: Die Erklärung (4) ist die beste Erklärung für das überraschende statistische Faktum gemäß (3). (Aus (6)).
- (8) Also: Bei jedem Versuchsdurchlauf findet im Gehirn der Versuchsperson ein elektrischer Einzelprozess statt, der – ohne elektrische Störeinflüsse benachbarter Prozesse – an der Schädeloberfläche einen Potentialverlauf verursachen würde, wie sie die Berechnung zur Durchschnittskurve zeigt. Q.E.D. (Per Abduktion aus (7)).

Schlüsse auf die beste Erklärung sind in den Naturwissenschaften gang und gäbe. Der rekonstruierte Schluss von (2) und (3) zu (8) ist mithin methodologisch so respektabel wie weite Teile der Physik, Chemie, Biologie. Hier wie da gelten solche Schlüsse nur bis auf Abruf, sie sind nicht unfehlbar. So könnten später neue Daten eintreffen, mit denen die bislang beste Erklärung nicht mehr gut fertig wird; oder uns könnte eine noch bessere Erklärung für die bekannten Daten einfallen, auf die wir bislang mangels Phantasie nicht gekommen sind.¹³

Doch es wäre kein angemessener Schachzug, diese Möglichkeiten bloß hypothetisch ins Feld zu führen, um einen gegebenen Schluss auf die beste Erklärung anzugreifen. Solange wir keine neue, bessere Erklärung haben und solange keine neuen widerspenstigen Daten herkommen, solange gilt die bislang beste Erklärung für die Daten, die wir haben. Da es den Naturwissenschaftlern nur zu bewusst ist, dass ihre Erklärungen bloß bis auf Abruf gelten, sollte man ihnen aus der Fehlbarkeit ihrer Erklärungen keinen Strick drehen.

Es gibt freilich noch eine andere Form von Fehlbarkeit, die man im Sinn haben könnte, wenn man einen gegebenen Schluss auf die beste Erklärung angreifen möchte: Es könnte sich schon jetzt herausstellen, dass sich in der fraglichen Erklärung ein Denkfehler versteckt, der bislang niemandem aufgefallen ist. In diesem Falle verfehlte die fehlerhafte Erklärung nicht nur das Ziel, die beste unter den denkbaren Erklärungen zu sein; schlimmer: Sie verfehlte sogar das Ziel, überhaupt eine Erklärung zu sein. Denn eine fehlerhafte Erklärung ist nicht etwa eine schlechte Erklärung. Sie ist überhaupt keine Erklärung.

Auf die bloße Möglichkeit eines solchen Denkfehlers zu verweisen, wäre hier abermals kein angemessener Schachzug, um einen gegebenen Schluss auf die beste Erklärung anzugreifen. Denn bei halbwegs komplizierten Überlegungen besteht diese Möglichkeit im Prinzip immer; sogar bei den Beweisen aus der Mathematik. Erst wer einen Denkfehler namhaft machen kann, lanciert dadurch eine ernstzunehmende Kritik am fraglichen Schluss auf die beste Erklärung.

III. Libets Denkfehler laut Trevena und Miller

Dass Benjamin Libet und seine Kollegen einem Denkfehler zum Opfer gefallen sind, behaupten Judy Trevena und Jeff Miller in zwei Arbeiten, die im Jahr 2002 publiziert wurden und seitdem oft diskutiert worden sind.¹⁴ Der Vorwurf hat mit Libets statistischen Methoden zu tun und besagt: Selbst wenn der Beginn des Anstiegs des durchschnittlichen Bereitschaftspotentials bei jeder Versuchsperson vor dem Durchschnitt der Zeitpunkte liegen sollte, zu denen sich die Versuchsperson ihrer jeweiligen Bewegungsentscheidungen bewusst wird, könnte immer noch jeder dieser Bewusstseinszeitpunkte vor dem *zugehörigen* Beginn eines Potentialanstiegs liegen, den der Versuchsleiter nicht beobachten, sondern allenfalls abduktiv aus den Daten erschließen kann. Denn wenn man aus mehreren ansteigenden Potentialkurven eine Durchschnittskurve berechnet, so muss der Durchschnitt ihres jeweiligen Anstiegsbeginns nicht übereinstimmen mit dem Anstiegsbeginn der Durchschnittskurve. Abbildung 1 veranschaulicht diese Verhältnisse.

Auf der Zeitachse repräsentieren wir die letzte Sekunde vor dem Nullzeitpunkt (vor dem Beginn der Fingerbewegung); auf der y-Achse repräsentieren wir das elektrische Potential. Tun wir erst einmal so,

als wären drei ungestörte, saubere Potentialkurven A, B, C gegeben, die jeweils den im Gehirn ablaufenden Einzelprozess darstellen! (In Wirklichkeit können wir nur deren unbrauchbare Gegenstücke messen, die von zahllosen elektrischen Nachbarprozessen aus der neuronalen Umgebung des Einzelprozesses überlagert und gestört werden; solche Kurven werde ich Ihnen im Anhang zeigen, siehe z.B. Abbildung 6). Nehmen wir an, dass die ungestörten Potentialkurven mit dem tatsächlichen Entscheidungsprozess zu tun haben; sie stellen diejenige unbeobachtbare Größe dar, die im Gehirn für die Entscheidung sorgt.¹⁵ Wie ich annehmen will, verlaufen diese – rein hypothetischen – ungestörten Potentialkurven zunächst allesamt auf der Nulllinie, und zwar bis zum Zeitpunkt t_a , t_b bzw. t_c . Erst ab $t_a = -900$ ms, $t_b = -630$ ms bzw. $t_c = -360$ ms steigen die ungestörten Potentialkurven linear an, und sie erreichen zur Nullzeit ihren jeweiligen End- und Höhepunkt bei einem Potential von rund $-10\mu\text{V}$.¹⁶

In der Kurve A steigt das Potential (nach t_a) am flachsten, in der Kurve C (nach t_c) am steilsten. Und alle drei ungestörten Potentialkurven zeigen jeweils in ihrem Knick t_a , t_b bzw. t_c an, wo der Anstieg des jeweils ungestörten Potentials beginnt. Der Durchschnitt t_\emptyset des Anstiegsbeginns dieser drei Potentialkurven liegt bei:

$$t_\emptyset = \frac{(t_a + t_b + t_c)}{3} = t_b = -630 \text{ ms.}$$

Jetzt betrachten wir die Durchschnittskurve D. Sie entsteht durch Drittelung der Summe aus den Kurven A, B, C. Da jede dieser Kurven jeweils einen Knick hat und da jeder Knick zu einem verschiedenen Zeitpunkt stattfindet ($t_a \neq t_b \neq t_c$), muss die Summenkurve (und also auch die Durchschnittskurve D) insgesamt drei Knicke aufweisen – natürlich da, wo die drei ursprünglichen Kurven geknickt waren. Der erste dieser drei Knicke liegt bei t_a und zeigt den Anstiegsbeginn der Durchschnittskurve an, die anderen beiden Knicke bei t_b und t_c zeigen diejenigen Zeitpunkte an, zu denen sich der Anstieg der Durchschnittskurve D jeweils schlagartig beschleunigt. Da ich mich nur für den *Beginn* des Anstiegs der Durchschnittskurve interessiere, ignoriere ich diese letzten beiden Knicke, konzentriere mich auf den ersten und halte fest: Der Anstiegsbeginn der Durchschnittskurve liegt nicht beim Durchschnitt t_\emptyset des Anstiegsbeginns der Einzelkurven, sondern da, wo die *erste* Einzelkurve beginnt anzusteigen!

Dieses Ergebnis hängt nicht von den Besonderheiten meines Beispiels ab.¹⁷ Der Anstiegsbeginn einer Durchschnittskurve liegt mit mathematischer Notwendigkeit immer da, wo die erste Einzelkurve anzusteigen beginnt (nicht etwa beim Durchschnitt des Anstiegsbeginns der Einzelkurven).

Und das bedeutet offenbar, dass in Libets Überlegungen ein Denkfehler steckt.¹⁸ Die Durchschnittskurve repräsentiert zwar alle Einzelkurven, aus denen sie hervorgegangen ist – aber deren Anstiegsbeginn repräsentiert sie nicht, vielmehr verzerrt sie ihn; der Anstiegsbeginn der Durchschnittskurve erscheint unnatürlich früh. Diese Verzerrung der wahren Zeitverhältnisse kommt – offenbar unzulässigerweise – Libets Schluss zugute. Libet möchte uns davon überzeugen, dass das Potential überraschend früh anzusteigen beginnt – bevor sich die Versuchsperson bewusst dazu entscheidet, den Finger zu bewegen.

Kurzum, der frühe Anstieg des Bereitschaftspotentials, mit dem Libet Aufsehen erregt hat, ist ein mathematisches Artefakt; das sog. Verschmierungsartefakt.¹⁹

Daraus ergibt sich: Es könnte durchaus sein, dass *jede* bewusste Entscheidung vor dem jeweiligen Anstiegsbeginn der dazugehörigen Einzelkurve stattfand. Diese Möglichkeit wird in Abbildung 2 verdeutlicht, in der ich fiktive Beispiele für den frühesten bewussten Moment der jeweiligen Entscheidung mithilfe dreier Zeitpunkte c_a , c_b , c_c eingetragen habe. Sehen Sie selbst: Wenn die bewusste Entscheidung jeweils eine Zehntelsekunde (100 Millisekunden) vor dem Anstiegsbeginn des einzelnen (ungestörten) Bereitschaftspotentials stattfand, dann liegt der Durchschnitt c_0 dieser drei Zeitpunkte bei -730 ms, also 170 ms nach dem Anstiegsbeginn in der Durchschnittskurve! Und das bedeutet offenbar folgendes. Aus der unbestrittenen Behauptung

- (9) Der Anstiegsbeginn der Durchschnittspotentialkurve am Schädel einer Versuchsperson liegt vor dem Durchschnitt der Zeitpunkte, zu denen sich die Versuchsperson ihrer Bewegungsentscheidung bewusst wird,

folgt nicht, dass auch nur ein einziges Mal das ungestörte Einzelpotential im Schädel der Versuchsperson anzusteigen beginnt, bevor ihr die zugehörige Bewegungsentscheidung bewusst wurde. Und darum folgt erst recht nicht, dass das immer so ist. Und ebenso wenig folgt, dass es im durchschnittlichen Einzelfall – oder wahrscheinlich – so ist.²⁰

Warum ist dies Problem jahrelang niemandem aufgefallen? Vielleicht wegen der vertrackten Datenlage. Wie es um die neuronalen Zeitverhältnisse bei einzelnen freien Entscheidungen steht, wissen wir nicht so ohne weiteres. Zwar wissen wir, dass der Anstiegsbeginn einer Durchschnittskurve beim Anstiegsbeginn der zuerst ansteigenden Einzelkurve liegen muss, aber wie und wann die Einzelkurven ansteigen, wissen wir nicht – wir kennen nur die Durchschnittskurve. Und natürlich kann ein und dieselbe Kurve den Durchschnitt sehr verschiedener Einzelkurven zeigen. Die Kurve D aus Abbildung 1 bietet erstens den Durchschnitt aus den Einzelkurven A, B, C; zweitens den Durchschnitt aus drei identischen Einzelkurven D. Drittens, viertens und fünftens gibt es natürlich beliebig viele weitere Kurven A₃, B₃ und C₃ bzw. A₄, B₄ und C₄ usw., deren Durchschnitt exakt auf der Kurve D verläuft.²¹ Weitere Beispiele dafür kann ich Ihnen ersparen, denn wir haben schon einen für unsere Zwecke besonderes wichtigen Fall, das ist der zweite.

Wenn die empirisch vorgegebene Durchschnittskurve der gemessenen Potentiale auf zeitlich einheitliche (oder doch gleichartige) Einzelprozesse im Gehirn zurückgeführt werden kann, wenn sich also im Gehirn bei jeder einzelnen Bewegungsentscheidung zeitlich immer recht präzise dasselbe abspielt (abgesehen von den störfreudigen neuronalen Nachbarprozessen), dann wird der Anstiegsbeginn der Durchschnittskurve recht präzise den durchschnittlichen Beginn des jeweils zugrundeliegende Hirnprozesses anzeigen; dann steht Libets Schluss wieder besser da, als es eben noch schien. (Man kann darüber streiten, ob wir berechtigt sind, eine solche Einheitlichkeit der Gehirnprozesse anzunehmen; empirisch nachweisen lässt sie sich mit den Mitteln der Elektroenzephalographie nicht. Zwar ist eine Erklärung, in der einheitliche Prozesse postuliert werden, *ceteris paribus* besser, weil einfacher als eine uneinheitliche Erklärung. Aber Einfachheit ist sicher nicht der einzige Maßstab, an dem sich die Güte einer Erklärung ermesen lässt. Nichtsdestoweniger bietet sich eine solche Erklärung zunächst einmal an, als Ausgangspunkt für die weiteren Überlegungen).

IV. Wo greift die Kritik?

Steckt nun in Libets Schluss ein Denkfehler oder nicht? Im Lichte der zuletzt angestellten Überlegung sollte man Libet keinen Denkfehler vorwerfen. Ich zeige Ihnen noch einmal den entscheidenden Schluss seiner Argumentation (wie vorhin rekonstruiert, aber gekürzt):

- (3) Statistische Analyse: Das Durchschnittspotential steigt eine halbe Sekunde vor dem Nullzeitpunkt langsam, aber deutlich an.
- (4) Eine Erklärung für (3) lautet: Bei *jedem* Versuchsdurchlauf findet im Gehirn der Versuchsperson ein elektrischer Einzelprozess statt, der – ohne elektrische Störeinflüsse benachbarter Prozesse – an der Schädeloberfläche einen Potentialverlauf verursachen würde, wie sie die berechnete Durchschnittskurve zeigt. Viele zufällige benachbarte Prozesse beeinflussen die einzelnen Versuchsdurchläufe gravierend. Doch im statistischen Mittel heben sich diese Störeinflüsse gegenseitig auf. Daher zeigt die Durchschnittskurve denjenigen Verlauf, den die gemessenen Einzelkurven hätten zeigen müssen, wären sie nicht von Nachbarprozessen gestört worden.

Was Libet im Schritt (4) meiner Rekonstruktion als Erklärung für die Durchschnittskurve gemäß (3) anbietet, enthält keinen Denkfehler. In dieser Erklärung wird nirgends vom Anstiegsbeginn der Durchschnittskurve auf den durchschnittlichen Beginn der (ungestörten) Einzelkurven geschlossen; vielmehr wird *in entgegengesetzter Richtung* aus (postulierten) ungestörten Einzelkurven auf die empirisch gegebene Durchschnittskurve geschlossen. Dieser Schluss ist wasserdicht; von Einzelkurven lässt sich eindeutig auf die Durchschnittskurve schließen.²²

Nur in der umgekehrten Richtung drohen Fehlschlüsse, denn aus einer gegebenen Durchschnittskurve lässt sich nichts Eindeutiges über die Einzelkurven ableiten, aus denen der Durchschnitt entstanden ist.

Wenn das stimmt, dann bringt Trevenas und Millers Kritik an dieser Stelle keinen Denkfehler in Libets Überlegungen zutage; ihre Kritik sollte vielmehr anders verstanden werden. Sie greift weder Schritt (3) noch Schritt (4) meiner Rekonstruktion an, sondern den Schritt, den ich in der Fortsetzung der Argumentation hervorhebe:

- (5) Eine andere Erklärung des überraschenden Faktums aus (3) lautet: alles Zufall.
- (6) Die Erklärung gemäß Schritt (4) ist besser als die gemäß Schritt (5). *Andere Erklärungen für das überraschende statistische Faktum (3) ähneln in dieser Hinsicht eher der schlechteren Erklärung.*

Recht verstanden, führt Trevenas und Millers korrekter Hinweis auf Verschmierungsartefakte also nicht zu dem Ergebnis, dass Libets Erklärung (4) für die empirischen Daten (3) irgendwelche Fehlschlüsse enthält. Nein, wenn ihr Hinweis kritisches Potential enthält, dann entfaltet es sich an späteren Stellen der Argumentationskette, mit der ich Libets Überlegungen rekonstruiert habe. Indem Trevena und Miller etwas gegen Libet einwenden wollen, müssen sie Alternativerklärungen vorschlagen, die sie (gegen (6)) für besser halten oder für mindestens so gut wie Libets Erklärung (4). Und dafür genügt kein allgemeiner Hinweis auf Verschmierungsartefakte; die Kritiker müssen mit einer besseren Erklärung aufwarten. Dass das alles andere als einfach ist, will ich in den kommenden beiden Abschnitten vorführen. Dort möchte ich zeigen, auf welchen mathematischen Pfaden man wandeln muss, um eine alternative Erklärung für Libets Durchschnittskurven zu ersinnen – eine Alternative, die sich mit der Annahme unserer Willensfreiheit besser verträgt. Genauer gesagt: eine Alternative, in der die bewussten Augenblicke der Entscheidung so gut wie immer vor dem Anstiegsbeginn der einzelnen Potentialkurven liegen. Ich werde diese Alternative exemplarisch errechnen. Wer den Rechenweg nachvollzieht, lernt dabei ein Muster kennen, mit dessen Hilfe auch andere Alternativen bestimmt werden können. Wie Sie sehen werden, erlaubt uns die arithmetische Struktur der Situation keine Konstruktion sonderlich plausibler Alternativerklärungen.

V. Freiheitsfreunde auf mathematischen Pfaden

Nehmen wir an, Freiheitsfreunde wollten Trevenas und Millers Verschmierungsartefakte ausnutzen, um ein Schlupfloch für die Willensfreiheit aufzuzeigen. Was müssten sie tun? Sie müssten sagen: Libets berechnete Durchschnittskurve $D(t)$ bildet nicht nur den Durchschnitt aus vierzig gemessenen Potentialkurven P_0, P_1, \dots, P_{39} (in denen es

undifferenziert rauscht, wegen zufälliger Störungen aus benachbarten Gehirnprozessen, die mit der Entscheidung zur Fingerbewegung nichts zu tun haben):

$$(10) \quad D(t) = \frac{\sum_0^{39} P_i(t)}{40}.$$

Und sie bildet auch nicht nur den Durchschnitt aus vierzig gleichartigen Kurven D_0, D_1, \dots, D_{39} , die man als ungestörte Potentialkurven postulieren kann (also als Echo der neuronalen Einzelprozesse ohne Störfeuer von nebenan) und die nicht genug Zeit für die freie Entscheidung lassen, weil sie zu früh anzusteigen beginnen:

$$(11) \quad D(t) = \frac{\sum_0^{39} D_i(t)}{40}, \text{ mit } D_i(t) \approx D(t).$$

Vielmehr bildet die Durchschnittskurve auch den Durchschnitt von vierzig Kurven Z_0, Z_1, \dots, Z_{39} , die sich deshalb ganz gut mit der Annahme von Freiheit vereinbaren lassen, weil sie mit wenigen Ausnahmen immer erst nach der bewussten Entscheidung anzusteigen beginnen. Das bedeutet: Die gesuchten vierzig Kurven müssen zu unterschiedlichen Zeitpunkten anzusteigen beginnen und dann unterschiedlich steil ansteigen. Mindestens eine dieser Kurven Z_0 muss mit ihrem Anstieg schon mehr als eine halbe Sekunde vor Null beginnen (bei -550 ms); sonst könnte Libets gemessene Durchschnittskurve $D(t)$ nicht zu diesem Zeitpunkt anzusteigen beginnen (darin lag die Pointe des Hinweises auf Verschmierungsartefakte). Doch die meisten anderen dieser Kurven Z_1, \dots, Z_{39} können mit ihrem Anstieg weit später beginnen (nämlich stets nach dem durchschnittlichen Zeitpunkt der bewussten Bewegungsentscheidung, also irgendwann zwischen -300 ms und 0 ms).

Versuchen wir also, Libets Daten aus vierzig Versuchsdurchläufen so zu erklären, wie es die Freiheitsfreunde haben wollen! Ich erinnere noch einmal an Libets Daten:

(12) Durchschnittlicher Zeitpunkt der bewussten Entscheidung:
 -200 ms.

(13) Beginn des Anstiegs der durchschnittlichen Potentialkurve:
 -550 ms.

Um die Berechnungen transparenter zu machen, werde ich Libets errechnete Durchschnittskurve wie eine lineare Beziehung behandeln. Wie man sich klarmachen kann, ändert sich auch ohne diese Vereinfachung an den Größenordnungen der Situation nichts. Wir nehmen also an:

- (14) Form der durchschnittlichen Potentialkurve D: Zwischen -550 ms und 0 ms steigt die Kurve (dem Betrage nach) linear an, von $0 \mu\text{V}$ auf $-10 \mu\text{V}$.

$$D(t) = \begin{cases} -10/550 \frac{\mu\text{V}}{\text{ms}} (t + 550 \text{ ms}) & \text{für } t \text{ später als } -550 \text{ ms;} \\ 0 \mu\text{V} & \text{sonst.} \end{cases}$$

[Diese vereinfachte Annahme soll dazu dienen, die zugrundeliegende Arithmetik klarzumachen; bei weniger gut aufgeräumten Kurven ändern sich die Details, aber nicht die Größenordnungen].

Um diese Kurve zu erklären, postulieren wir eine Ausreißerkurve $A(t) = Z_0(t)$ und 39 kongruente (oder nahezu kongruente) Standardkurven $Z_1(t), \dots, Z_{39}(t)$. Der Durchschnitt aus allen diesen Kurven soll $D(t)$ ergeben. Das führt zu folgendem Postulat:

(15)
$$D(t) = \frac{A(t) + 39 Z(t)}{40}, \text{ mit } Z(t) \approx Z_1(t) \approx \dots \approx Z_{39}(t).$$

Wir postulieren (im Namen und zuliebe der Freiheit), dass *fast jede* bewusste Entscheidung 50 ms vor dem Anstiegsbeginn der zugehörigen ungestörten Potentialkurve $Z_i(t)$ getroffen worden ist; für $i = 1, 2 \dots 39$ – also für die Standardkurven, in denen sich die Versuchsperson bewusst frei hat entscheiden können.²³ Die Ausreißerkurve beschreibt den Ausnahmefall; hier kann das Bereitschaftspotential sehr wohl angestiegen sein, bevor dies der Versuchsperson bewusst wurde. Um nicht missverstanden zu werden: Ich nenne die Kurve A nicht ausreißerisch, weil ich an diesem Punkt meiner Rechnung schon wüsste, dass sie sich von den anderen Kurven stark unterscheidet. Nein, ich nenne sie deshalb Ausreißerkurve, weil sie in Sachen Freiheit eine Ausnahme bietet. Wann sie ansteigt, legen wir völlig unabhängig von den bewussten Entscheidungs-Zeitpunkten fest, und zwar so, dass sie und nur sie fürs Verschmierungsartefakt verantwortlich ist.²⁴

Wie muss die Ausreißerkurve A verlaufen? In Übereinstimmung mit

(I3) beginnt sie schon bei -550ms anzusteigen, und zwar linear und zunächst ungeknickt:

$$(I6) \quad A(t) = \begin{cases} 0\ \mu\text{V} & \text{für } t \text{ vor } -550\text{ms;} \\ -400/550 \frac{\mu\text{V}}{\text{ms}}(t + 550\text{ms}) & \text{für } t \text{ zw. } -550\text{ms} \text{ u. dem Knick.}^{25} \end{cases}$$

Wo liegt der Knick dieser Kurve? Genau bei dem Zeitpunkt, zu dem die kongruenten Standardkurven $Z_1(t), \dots, Z_{39}(t)$ anzusteigen beginnen.²⁶ Diesen Zeitpunkt will ich jetzt (im Namen und zuliebe der Freiheit) bestimmen. Wir müssen der Freiheit zuliebe behaupten, dass auch im Ausreißer-Durchlauf die bewusste Entscheidung 50ms vor dem Anstiegsbeginn der zugehörigen Potentialkurve A getroffen wurde, also zum Zeitpunkt -600ms .²⁷ Nun lag der Durchschnittszeitpunkt der Entscheidungen bei -200ms . Nehmen wir an, dass er in den 39 standardgemäß verlaufenden Durchgängen $Z_1(t), \dots, Z_{39}(t)$ immer zum gleichen Zeitpunkt t' stattfand. Aus diesen Annahmen ergibt sich:

$$(I7) \quad -200\text{ms} = (39t' + -600\text{ms})/40, \text{ also}$$

$$(I8) \quad t' = -7400\text{ms}/39 = -189,7\text{ms}.$$

Wie Sie sehen, ändert sich durch diese vielleicht überpenible Korrekturrechnung nichts wesentliches an dem Zeitpunkt, auf den ich es abgesehen habe; er verschiebt sich nur um 10ms nach hinten.²⁸ Da ich verlange, dass die 39 ungestörten Standardkurven immer genau 50ms nach der jeweiligen bewussten Entscheidung anzusteigen beginnen, steht nun ein wichtiger Parameter dieser 39 Kurven fest. Sie beginnen ihren Anstieg bei -140ms . Hier ihre Form:

$$(I9) \quad Z(t) = \begin{cases} \beta(t + 140\text{ms}) & \text{für } t \text{ später als } -140\text{ms;} \\ 0\ \mu\text{V} & \text{sonst.} \end{cases}$$

Wie steil sie ansteigen (gemäß dem Parameter β), müssen wir jetzt noch berechnen. Dazu ermitteln wir zunächst den Wert der Ausreißerkurve zum fraglichen Zeitpunkt, also:

$$(20) \quad A(-140\text{ms}) = -400/550 \frac{\mu\text{V}}{\text{ms}}(-140\text{ms} + 550\text{ms}) = -298\ \mu\text{V}.$$

Das ist wirklich eine extreme Ausreißerkurve, denn ihr Potential ist zu diesem Zeitpunkt fast *dreißigmal* so hoch, wie es am Ende der Geschichte durchschnittlich sein darf (zur Zeit 0 liegt D bei $-10\ \mu\text{V}$, siehe (I4)). Man mag fragen: Wieso ist dieses sehr hohe Potential (das wir

im Namen der Freiheit postuliert haben) niemandem bei der Messung des fraglichen Versuchsdurchgangs aufgefallen? Bedenken Sie, dass die tatsächlich gemessene Kurve P_0 genau so nichtssagend aussah wie die anderen gemessenen Potentialkurven $P_1 \dots, P_{39}$. Die Antwort lautet: Sehr starke Störprozesse S_0 aus Nachbarregionen im Gehirn haben die Ausreißerkurve so stark nivelliert, dass sie im Rauschen unterging – und zwar besonders kräftig just in dem Augenblick, in dem A auf ihren Höhepunkt anstieg. Welch ein Zufall! Denkbar, oder?

Aber gehen wir weiter. Nehmen wir an, dass diese Kurve $A(t)$ bis zur Zeit 0 genau auf den Durchschnittswert absinkt, und zwar linear. Ihre Steigung beträgt also nach dem schwindelerregenden Höhepunkt $288/140$. Und so bekommen wir für den letzten Zeitraum der Geschichte folgende Beziehung:

$$(21) \quad A(t) = 288/140 \frac{\mu\text{V}}{\text{ms}} (t + 140 \text{ ms}) - 298 \mu\text{V} \text{ für } t \text{ später als } -140 \text{ ms.}$$

Hieraus können wir via (15) und (19) die identische Steigung β der 39 Standardkurven berechnen. Denn wir wissen, dass im fraglichen Zeitraum die durchschnittliche Steigung gemäß (14) bei $-10 \mu\text{V}/550 \text{ ms} = -0,018 \frac{\mu\text{V}}{\text{ms}}$ liegt und dass die Ausreißerkurve eine Steigung von $288/140 \frac{\mu\text{V}}{\text{ms}}$ hat. Also gilt:

$$(22) \quad -10/550 \frac{\mu\text{V}}{\text{ms}} = \frac{(39\beta + 288/140 \frac{\mu\text{V}}{\text{ms}})}{40}.$$

$$(23) \quad \beta = -0,071 \frac{\mu\text{V}}{\text{ms}}.$$

Kurz und gut, wir müssen 39 Kurven folgender Form postulieren:

$$Z(t) = \begin{cases} -0,071 \frac{\mu\text{V}}{\text{ms}} (t + 140 \text{ ms}) & \text{für } t \text{ später als } -140 \text{ ms;} \\ 0 \mu\text{V} & \text{sonst.} \end{cases}$$

Und wir postulieren *eine* total anders aussehende, geknickte Kurve dieser Form:

$$A(t) = \begin{cases} -440/550 \frac{\mu\text{V}}{\text{ms}} (t + 550 \text{ ms}) & \text{für } t \text{ zw. } -550 \text{ ms u. dem Knick;} \\ 288/140 \frac{\mu\text{V}}{\text{ms}} (t + 140 \text{ ms}) - 298 \frac{\mu\text{V}}{\text{ms}} & \text{für } t \text{ später als } -140 \text{ ms.} \end{cases}$$

In Abbildung 3 sehen Sie den Unterschied zwischen den 39 Standardkurven (blau) und der einen Ausreißerkurve (rot). Er ist drastisch. Theoretisch denkbar sind solche Verhältnisse. Aber sind sie wahrscheinlich?

Besonderes Ärgernis erregt eine Tatsache, auf die ich vorhin schon im Vorübergehen hingewiesen habe. Natürlich können Freiheitsfreunde eine hypothetische Ausreißerkurve postulieren. (Auch Libets Kurven sind hypothetisch; ohne Hypothesen kommt keiner auf diesem Gebiet vom Fleck, denn die tatsächlich gemessenen Kurven sagen für sich allein nicht genug aus). Nur: Es grenzt an eine intellektuelle Zumutung zu postulieren, dass sich *manchmal* vor einer freien Fingerbewegung im Gehirn ein extremes Potential aufbaut und dass wir dies Potential deshalb nicht messen, weil es genau im richtigen Moment zufälligerweise von genauso starken Nachbarprozessen nivelliert wird (die mit der Entscheidung nichts zu tun haben). Ein solches Postulat wirkt wie reines Wunschdenken. Denn ähnlich krasse Verhältnisse müssten unsere Freiheitsfreunde in *jeder* Serie von 40 Durchläufen postulieren. (Immer eine exakt passend gestörte Ausreißerkurve und 39 total andere Kurven, deren Störung vergleichsweise gering ist). Mit Verlaub, da sind 6 Richtige im Lotto wahrscheinlicher.

VI. Andere Parameter?

Wie stark hängt mein Resultat aus dem vorigen Abschnitt bloß von den Besonderheiten meines Beispiels ab? Nicht sehr. Ich habe die Rechnung im Detail vorgeführt, um Ihnen ein Gespür für die Größenordnungen zu verschaffen. Es kommt für meine Ziele nicht darauf an, ob Libets Durchschnittskurve $D(t)$ oder die 39 postulierten Standardkurven wirklich linear ansteigen; sie tun es nicht.

Erst recht nicht kommt es darauf an, ob die Ausreißerkurve einen Knick hat und auf beiden Seiten des Knicks linear verläuft. Dass es im Gehirn so gerade und eckig zugeht, ist unwahrscheinlich. Aber nicht wegen der Ecken und Kanten habe ich die Ausreißerkurve vorhin extrem unwahrscheinlich genannt. Sondern wegen der *Größenordnungen* der Störprozesse, die man postulieren muss, um verständlich zu machen, warum sich die extreme Ausreißerkurve (die im Innern des Gehirns postuliert wird) nicht auf der Schädeloberfläche bemerkbar macht. Und diese Größenordnungen hängen nicht wesentlich von der exakten Form der Kurven ab – weder vom Unterschied zwischen geknickten und glatten Kurven noch vom Unterschied zwischen linearen und anderen Anstiegen dieser Kurven.

Wenn Sie z.B. einen geeigneten Ausschnitt einer Sinuskurve (zwischen $\frac{-\pi}{4}$ und $\frac{\pi}{2}$) passend parametrisieren und an die Stelle der Durchschnittskurve $D(t)$ oder der 39 postulierten Standardkurven setzen (vergl. Abb. 4), wird die Ausreißerkurve trotzdem durch die Decke gehen. Warum? Weil sie fürs Ansteigen der Durchschnittskurve in einem großen Zeitraum der betrachteten Prozesse ganz alleine verantwortlich sein soll, während sich die 39 Standardkurven noch nicht aufschwingen – und wer den Durchschnitt aus 39 Nullen und einer Ausnahme auch nur auf moderate Höhen treiben will (auf ca. $-10\mu\text{V}$), muss die Ausnahme um Größenordnungen höher veranschlagen als den Durchschnitt. Dieser triviale Sachverhalt gilt für geeignet ausgeschnittene sinusförmige Anstiege genauso wie für Geraden und für andere steigende Kurven. Durchschnitt ist Durchschnitt. (Im Anhang werde ich Ihnen ein Resultat vorführen, das mit Sinuskurven gewonnen wurde und gut zu meinen Überlegungen passt).

Bietet meine Rechnung aus dem vorigen Abschnitt vielleicht aufgrund anderer Vorentscheidungen keinen repräsentativen Fall? Hier sind einige Gesichtspunkte, die für diese Frage zu berücksichtigen sind. Erstens habe ich angenommen, dass die 39 Standardkurven allesamt (nahezu) kongruent verlaufen – was wäre, wenn wir ihnen eine gewisse Variabilität erlauben? In der Tat, das Maximum der Ausreißerkurve lässt sich so ein kleines Stückchen verringern – ein unwesentliches Stückchen. Wer eine Ausreißerkurve postuliert, kann die anderen Kurven vergleichsweise spät ansteigen lassen. Selbst wenn sie nicht alle exakt zum selben Zeitpunkt anzusteigen beginnen, müssen sie später ansteigen als die Ausreißerkurve; sie allein soll ja (im Namen des Verschmierungsartefakts) daran schuld sein, dass die Durchschnittskurve so früh anzusteigen beginnt. Fest steht: Solange der Anstieg einer Durchschnittskurve einzig und allein auf dem Anstieg einer der durchschnittsbildenden Kurven beruht, solange muss diese eine Kurve extrem steil ansteigen – um den anfänglichen Nichtanstieg der 39 anderen Kurven zu kompensieren.

Das bedeutet auch, dass die Karten für einen freiheitsfreundlichen Verweis aufs Verschmierungsartefakt schlechter würden, je mehr Durchläufe in die Durchschnittsbildung einfließen. Dass Libet immer nur mit 40 Durchläufen gearbeitet hat, ist von der Sache her nicht vorgegeben; er meinte, dass es für seine Zwecke genügt. Hätte er die Gefahr ernstgenommen, die seinen Zwecken aus dem Verschmierungsartefakt

drohte, so hätte er ohne Schwierigkeit auf 60 oder gar 100 Durchläufe hochgehen können.²⁹ Wenn wir annehmen – wie das Zufallsexperiment aus dem Anhang nahelegt –, dass Libets Durchschnittskurven dadurch noch ein wenig deutlicher, jedenfalls keineswegs schlechter geworden wären, so wird klar: Mit dieser unwesentlich verbesserten Methode wäre die verschmiert freiheitliche Alternativerklärung noch unplausibler geworden.

Hier ist ein zweiter Gesichtspunkt, den man ins Spiel bringen könnte, um meine Berechnung aus dem vorigen Abschnitt anzugreifen: Ich habe mit *einer* Ausreißerkurve gearbeitet; was ändert sich, wenn wir mit zwei (oder mehr) Ausreißerkurven arbeiten? – Zugegeben, rein arithmetisch verbessert sich die Situation, sobald nicht eine, sondern mehrere postulierte Kurven sehr früh anzusteigen beginnen; die Last der hohen Ausschläge verteilt sich dann auf mehrere Schultern und wird weniger extrem. Doch bedenken Sie: Je mehr Ausreißerkurven wir postulieren, desto stärker gefährden wir die bewusste Entscheidungsfreiheit der Versuchsperson; denn die Ausreißerkurven beginnen ihren Anstieg typischerweise vor den gemessenen Zeitpunkten der jeweiligen Entscheidung. Wir postulieren also (bei Vermehrung der Ausreißerkurven) entweder mehr Ablesefehler oder weniger Fälle von Freiheit.³⁰ – Weder das eine noch das andere wird der Freiheitsfreund übertreiben wollen. Nehmen wir daher z. B. nur zwei solcher Kurven (die gleichzeitig bei –550 ms anzusteigen beginnen). Dann ändert sich an der Dimension des Problems nichts. Das Unwahrscheinliche der resultierenden Alternativerklärung wird lediglich auf andere Weise unwahrscheinlich. Das will ich exemplarisch kurz begründen. Anstelle von (17) und (18) beginnt die Rechnung dann so:

$$(17^*) \quad -200 \text{ ms} = \frac{(38 t^* + -1200 \text{ ms})}{40}, \text{ also}$$

$$(18^*) \quad t^* = -6800 \text{ ms}/38 = -178,9 \text{ ms}.$$

Rein rechnerisch fällt diesmal die Steigung der beiden Ausreißerkurven A_1 und A_2 (bis zum Knick) geringer aus als vorhin, und beim Knick stehen die beiden Kurven auf:

$$(20^*) \quad A_{1;2}(-129 \text{ ms}) = -200/550 \frac{\mu\text{V}}{\text{ms}}(-129 \text{ ms} + 550 \text{ ms}) = -153,09 \mu\text{V}.$$

An ihrem höchsten Punkt nehmen die beiden Kurven dann also immer noch ein zig-faches Potential der Standardkurven an. Wieder müssen gigantische Störprozesse postuliert werden; sonst ließe sich nicht begründen, warum die Ausreißerkurven keine *sichtbaren* Potentialmessungen nach sich zogen. (Die gemessenen Einzelkurven bleiben ja so nichtssagend wie eh und je). Das Ausmaß der postulierten Störprozesse ist zwar jetzt geringer als im vorigen Abschnitt; aber dafür müssen diesmal *zwei* Störprozesse postuliert werden, die rein zufällig exakt im rechten Augenblick losgingen. Ein doppelter Zufall! Wie Sie sehen, wird die Sache nicht besser, wenn man zwei Ausreißerkurven postuliert.

Es gibt einige weitere Gesichtspunkte, die bei einer genauen Analyse der Experimente Libets zu berücksichtigen wären. Es würde meinen augenblicklichen Rahmen sprengen, sie mit der wünschenswerten Korngröße zu erörtern. Hier nur eine Andeutung: Je nach Anzahl der postulierten Ausreißerkurven können wir deren relative Häufigkeit (im Vergleich zu den Standardkurven) ermitteln. Im vorigen Abschnitt V lag diese relative Häufigkeit z.B. bei 2,5 % (und im vorigen Absatz lag sie bei 5 %). Hieraus lässt sich eine Voraussage dafür abschätzen, unter wievielen Libet-Serien die Ausreißerkurve aller Wahrscheinlichkeit nach *ausfallen* müsste. Entsprechend oft ginge Libets Experiment *ganz anders* aus. Die Durchschnittskurve der Bereitschaftspotentiale begönne viel später anzusteigen, und es läge in diesem Durchlauf keine scheinbare empirische Bedrohung für Willensfreiheit vor. Gibt oder gab es derartige harmlose Durchläufe? Und zwar hinreichend oft? Soweit ich sehe, hilft uns das veröffentlichte Zahlenmaterial bei dieser Frage nicht eindeutig weiter. Es wäre interessant, die Experimente hinreichend oft zu wiederholen, um hier weiterzukommen.³¹

VII. Libet ist unschuldig

Das Ergebnis aus den vorigen beiden Abschnitten ist niederschmetternd. Zwar lässt sich der erstaunlich frühe Anstiegsbeginn der Durchschnittskurven anders erklären, als Libet meinte. Aber alle Alternativen (die sich noch mit den Daten vertragen) wirken reichlich konstruiert.

Die Moral von der Geschichte lautet: Das sogenannte Verschmierungsartefakt kann sehr wohl die gemessene Durchschnittskurve so beeinflussen, wie Trevena und Miller vorgeführt haben. Aber das

Verschmierungsartefakt tritt genau dann auf, wenn man *schlechte* Erklärungen der Durchschnittskurve postuliert. Das berührt Libets Argument, recht verstanden, kein Stück. Denn Libet suchte nach der bestmöglichen Erklärung.

Natürlich kann man an Libets Methode und an seiner Schlussweise andere Dinge unplausibel finden als dessen Vernachlässigung des Verschmierungsartefakts. Diese Kritik an Libet dürfte uns tief in die Philosophie führen; dort tobt der ewige Streit. Auf die – mathematische – Kritik an Libet, die viele Autoren aus Trevenas und Millers Überlegungen herausgelesen haben, sollte man sich dagegen besser nicht stützen. Sie beruht auf einem wissenschaftstheoretischen Missverständnis.³²

Anhang: Ein mathematisches Zufallsexperiment zur Illustration

Um den mathematischen Mechanismus vorzuführen, auf dem Libets Erklärungen beruhen, haben wir ein Zufallsexperiment durchgeführt. Das zugehörige Programm ist von Matthias Herder geschrieben worden, und zwar in der Skriptsprache der Programmumgebung Matlab.³³

Wir betrachten zunächst eine ungestörte (hypothetische) Potentialkurve, die ungefähr auf die Weise ansteigt, wie es Libet postuliert hat. Anstelle des linearen Anstiegs (mit dem im Abschnitt V gearbeitet wurde) arbeiten wir jetzt mit einem geeigneten Ausschnitt einer Sinuskurve (Abbildung 4 oben); sie ist so parametrisiert, dass sie zu Libets Daten passt (Abbildung 4 unten). Wir nehmen an, dass eine solche Kurve mit dem neuronalen Korrelat der Entscheidung Hand in Hand geht. Sie steigt zunächst deutlich, ja sprunghaft an und nähert sich ihrem Maximum mit sinkender Steigung – glatt sozusagen.

Libets Einzelmessung an der Schädeloberfläche wird aber jedesmal von diversen Störprozessen laut rauschend überlagert, so dass die gemessene Einzelkurve keine Spur dessen zeigt, was in Abbildung 4 (unten) sichtbar ist.

Abbildung 5 zeigt ein Beispiel für dies Rauschen; wir stellen es mithilfe des sog. Rosa Rauschens ($1/f$ -Rauschens) dar. Das ist ein gut durchmisches Zufallssignal. Darin überlagern sich – per Zufall gleichverteilt – unzählige Schwingungen aller erdenklichen Frequenzen und Phasen; die Amplituden dieser Schwingungen sind ebenfalls zufällig, aber nicht

gleichverteilt (wie im sog. Weißen Rauschen), sondern so, dass (im Rosa Rauschen) Schwingungen mit Frequenz f und Amplitude A genauso wahrscheinlich auftauchen wie z.B. mit verdoppelter Amplitude $2A$ und halbiertes Frequenz $f/2$. (D.h. je höher die Frequenz einer der enthaltenen Schwingungen, desto wahrscheinlicher ist seine Amplitude gering). Das Rosa Rauschen eignet sich gut, um die mithilfe der Elektroenzephalographie gemessenen summierten elektrischen Aktivitäten des Gehirns zu modellieren; sie zeichnen sich ebenfalls durch relativ hohe Amplituden bei niedrigen Frequenzen und geringe Amplituden bei hochfrequenten Potentialschwankungen aus.

Für unser Zufallsexperiment haben wir sechzigmal jeweils irgendein zufälliges Rauschsignal auf jeweils eine ungestörte Kurve addiert, wie sie in Abbildung 4 (unten) gezeigt wurde. (Um die Sache nicht zu künstlich werden zu lassen, haben wir dafür nicht jedesmal dieselbe Kurve genommen, sondern gewisse zufällige Variationen der ungestörten Kurven zugelassen; auch im Gehirn wird ja sicher nicht jedesmal exakt nach demselben Zeitplan entschieden). Abbildung 6 bietet eines der Resultate dieser sechzig Additionen; wie gewünscht verschwindet im Rauschen das uns interessierende Signal (noch sichtbar in Abbildung 4 unten). Damit haben wir ein Modell für eine der Einzelmessungen Libets.

Was geschieht, wenn wir den Durchschnitt einiger der so modellierten Einzelmessungen bilden? Das kommt darauf an, wieviele der Einzelmessungen wir in den Durchschnitt einfließen lassen. Bei *zwei* gemittelten Einzelmessungen sieht das Resultat nicht besser aus als die Einzelmessung selbst (Abbildung 7; den Durchschnitt zeigen wir rot, die erste Einzelmessung hellgrau, die letzte Einzelmessung dunkelgrau). Schon bei zehn gemittelten Kurven wird das Resultat aussagekräftiger (Abbildung 8). Die rote Durchschnittskurve zeigt geringere Ausschläge als die (in verschiedenen Graustufen gezeigten) Einzelkurven; im rechten Ende der Graphik sieht der geübte Blick die Tendenz eines Anstiegs. Diese Tendenz tritt bei vierzig bzw. sechzig gemittelten Kurven immer deutlicher hervor (Abbildung 9 bzw. 10). Im Netz haben wir eine Computer-Animation hochgeladen, die das Zufallsexperiment in seiner Dynamik zeigt.³⁴

Die Erklärung für dies Phänomen ist nicht schwer: Je mehr Kurven wir in den Durchschnitt einfließen lassen, desto wahrscheinlicher kürzen sich Zufallsausschläge nach oben durch gleichstarke Zufallsaus-

schläge nach unten heraus. Wir haben das Experiment oft wiederholt (mit immer neuen, vom Zufallsgenerator erzeugten Störsignalen). Fast immer sah das Ergebnis nach 40 bis 60 Durchläufen so aus, wie unsere Abbildungen zeigen. *Fast* immer – aber nicht immer. Auch damit ist zu rechnen. Denn im Begriff des Wahrscheinlichen ist der Begriff des Unwahrscheinlichen bereits enthalten.

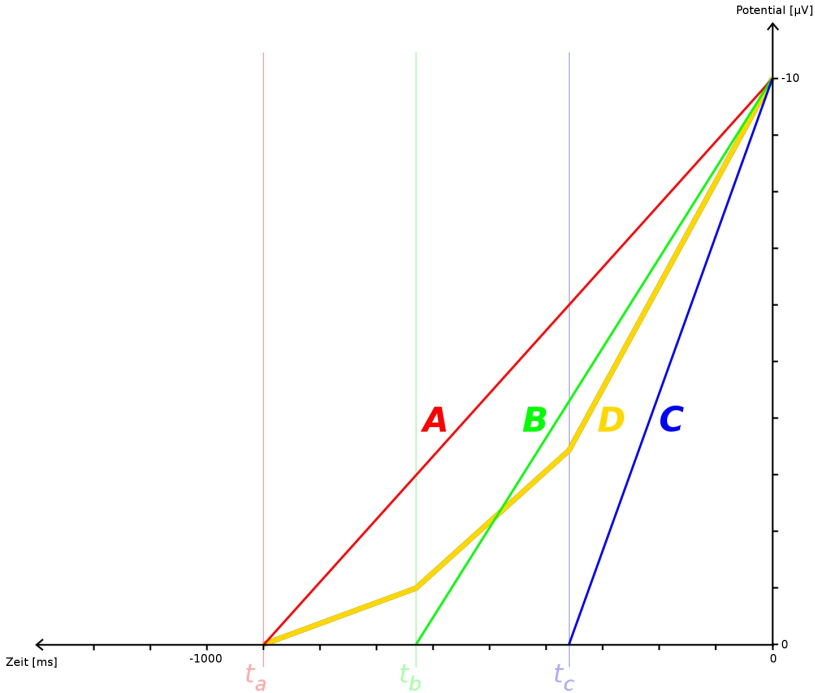


Abb. 1: Durchschnitt dreier Einzelkurven. Drei Potentialkurven A (rot), B (grün), C (blau) verlaufen zunächst auf der Nulllinie (nicht farbig eingezeichnet) und beginnen bei drei verschiedenen Zeitpunkten t_a , t_b bzw. t_c linear anzusteigen. Die Durchschnittskurve D (gelb) beginnt ihren Anstieg zu dem Zeitpunkt t_a , zu dem die *zuerst* steigenden Kurve A anzusteigen beginnt. D.h. der Durchschnitt der Anstiegsbeginne (hier: t_b) liegt im allgemeinen später als der Anstiegsbeginn der Durchschnittskurve D.

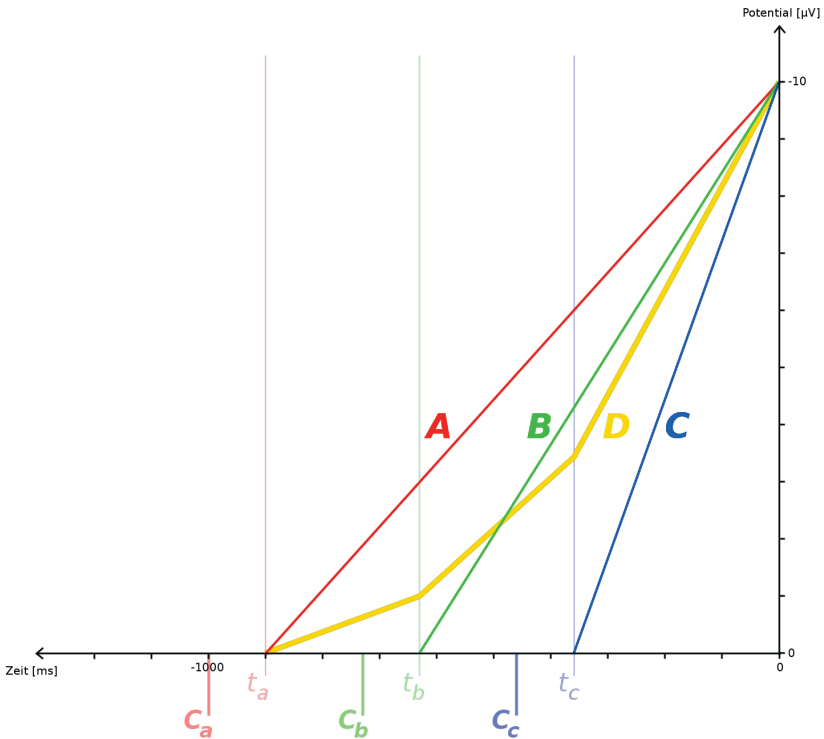


Abb. 2: Frühe (fiktive) Zeitpunkte bewusster Willensentscheidungen. Die neurophysiologischen Verhältnisse aus Abbildung 1 lassen sich mit erstaunlich frühen Zeitpunkten c_a , c_b bzw. c_c bewusster Willensentscheidungen kombinieren. Hier liegt z. B. jede Willensentscheidung 100 ms vor dem zugehörigen Anstiegsbeginn des Bereitschaftspotentials, und zwar – das ist der Clou – obwohl die Willensentscheidungen *im Durchschnitt* (hier bei c_b) erst nach dem Anstiegsbeginn t_a der *Durchschnittskurve* D stattfanden. (Um der Deutlichkeit willen haben wir in beiden Abbildungen die arithmetischen Verhältnisse zeitlich großzügiger dargestellt, als es realistisch ist; bei Libet steigen die Durchschnittskurven später an, und die durchschnittlichen Willensentscheidungen finden ebenfalls später statt; die tatsächlichen Abläufe drängen sich also enger zusammen. Das ändert nichts am Prinzip).

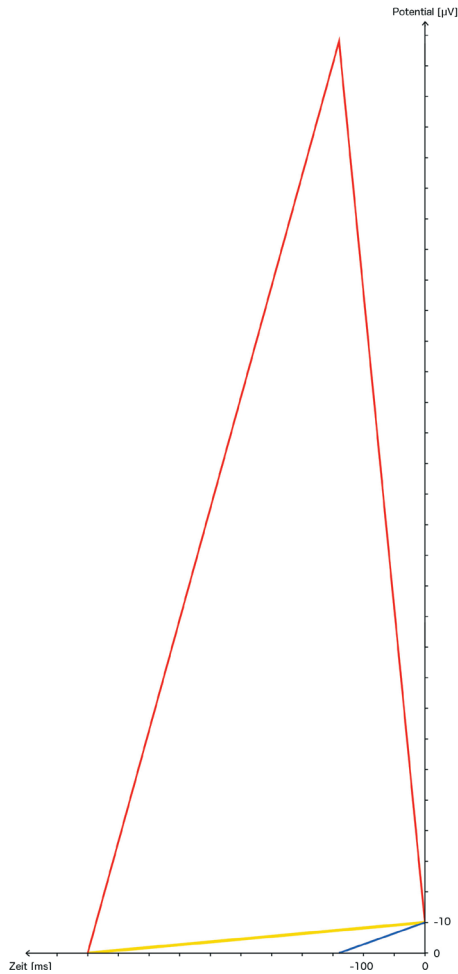


Abb. 3: Ausreißerkurve. Um die tatsächlichen Messungen Libets mit der Möglichkeit von Freiheit zu vereinbaren, kann man passende neurophysiologische Verhältnisse postulieren: Libets gemessene Durchschnittskurve (gelb) aus vierzig Versuchen kann sehr wohl durch (sich vierzigfach gegenseitig neutralisierendes) Zufallsrauschen (hier nicht dargestellt) und geeignet postulierte ungestörte (nicht messbare) Potentialkurven erklärt werden. Zum Beispiel kann man die Durchschnittskurve aus 39 (mehr oder minder) identischen Standardkurven (blau) und eine Ausreißerkurve (rot) erklären. Die Ausreißerkurve muss zunächst extrem steil ansteigen, wenn man allein mit ihrer Hilfe – via Verschmierungsartefakt – den frühen Anstieg der Durchschnittskurve zu erklären wünscht.

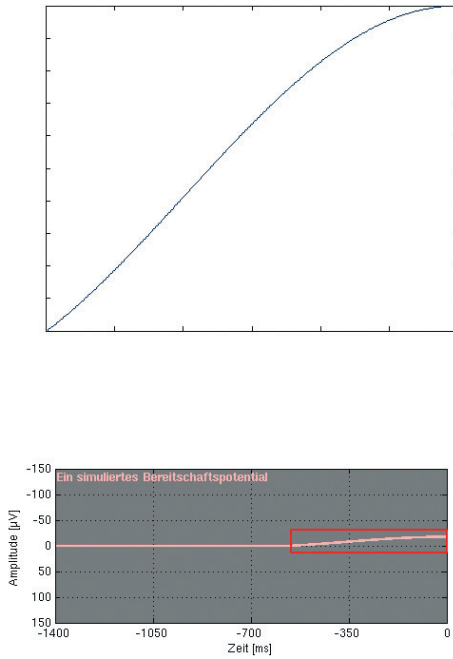


Abb. 4: Grundlagen fürs Zufallsexperiment. Im Zufallsexperiment haben wir keine linearen Anstiege benutzt (wie in Abbildungen 1 bis 3), sondern geeignete Ausschnitte aus Sinuskurven (oben), und zwar à la Libet parametrisiert (unten im eingerahmten rechten Teil der Abbildung). Bevor die Kurve unten sinusförmig ansteigt, verläuft sie linear auf der Null-Linie.

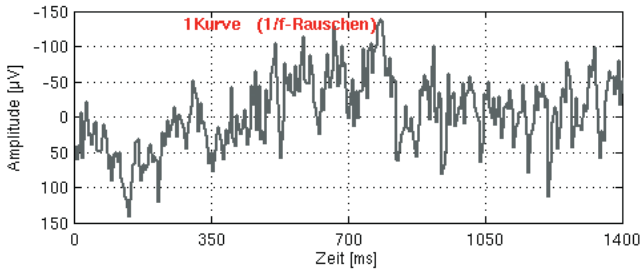


Abb. 5: Rosa Rauschen ($1/f$ -Rauschen). In diesem Zufallssignal (dessen Ausschläge weit über Libets Durchschnittskurve liegen) überlagern sich diverse Schwingungen der unterschiedlichsten Frequenzen, Amplituden und Phasen. Enthaltene Schwingungen mit hohen Amplituden haben öfter niedrige Frequenzen und umgekehrt.

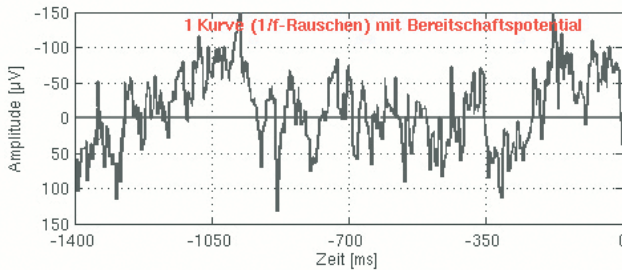


Abb. 6: Modell für eine Einzelmessung Libets. Hier ist das Rosa Rauschen (Abbildung 5) zu einem libet-artigen Anstieg des Bereitschaftspotentials (Abbildung 4 unten) addiert. Da das Rauschen weit höhere Ausschläge hat als das ungestörte Bereitschaftspotential, übertönt es das uns interessierende Signal. In den kommenden Abbildungen wird es durch Durchschnittsbildung allmählich wieder sichtbar.

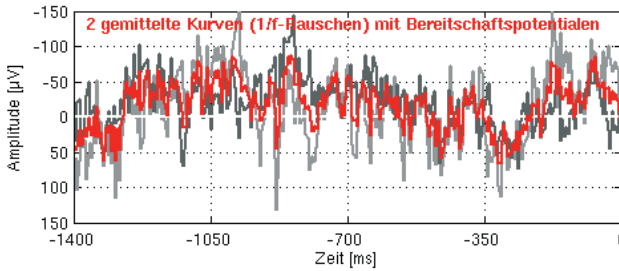


Abb. 7: Durchschnitt zweier Einzelmessungen. Das gewünschte Signal ist zwar noch nicht sichtbar. Aber wie man sieht, schlägt die Durchschnittskurve (rot) der Tendenz nach nicht mehr so stark nach oben und unten aus wie die Einzelkurven (hellgrau, dunkelgrau). Der Grund: Die zufälligen Signale neutralisieren sich gegenseitig, jedenfalls der Tendenz nach.

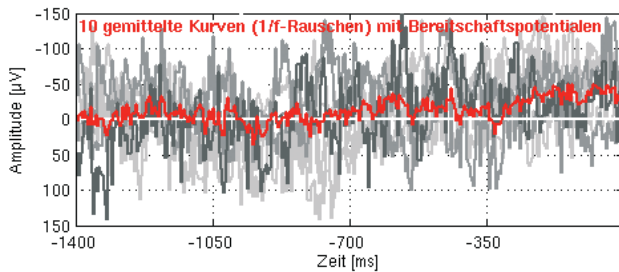


Abb. 8: Durchschnitt aus zehn Einzelmessungen. Das geübte Auge erkennt in der roten Durchschnittskurve bereits eine ansteigende Tendenz des Bereitschaftspotentials; richtig deutlich ist diese Tendenz noch nicht. Insbesondere lässt sich nicht gut erkennen, ob der Anstieg erst bei -350 ms beginnt oder schon bei -500 ms; nur wer letzteres *erwartet*, wird die Kurve so interpretieren, dass sie zu Libets Interpretationen passt – nach dem Motto: Man sieht nur, was man weiß. In der nächsten Abbildung sieht die Angelegenheit wesentlich eindeutiger aus.

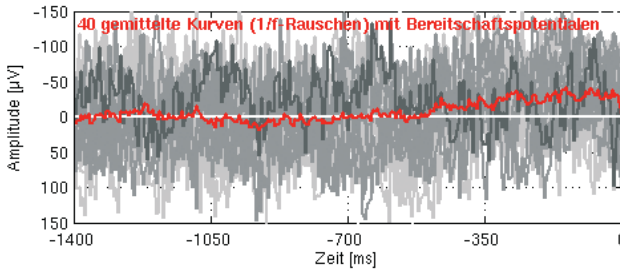


Abb. 9: Durchschnitt aus vierzig Einzelmessungen. Das gewünschte Signal lässt sich bereits gut aus der roten Durchschnittskurve erkennen. Die Einzelmessungen (deren letzte dunkelgrau und deren Vorgängerinnen immer blasser hellgrau gezeigt werden) neutralisieren sich in ihren hohen Ausschlägen, und sie lassen lediglich ein Zittern geringer Amplitude übrig.

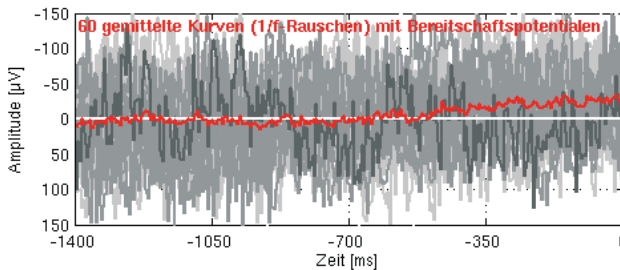


Abb. 10: Durchschnitt aus sechzig Einzelmessungen. Das bei Abbildung 9 erwähnte Zittern geht unmerklich zurück, die Durchschnittskurve nähert sich noch etwas deutlicher an das uns interessierende Signal an. Es ist vernünftig zu postulieren, dass allen sechzig Einzelmessungen jeweils eine ungestörte Kurve gemäß Abbildung 4 (unten) und jeweils eine Störung gemäß Abbildung 5 zugrundelag.

Anmerkungen

- 1 *Locus classicus* ist Kornhuber et al, 1965.
- 2 Aussagekräftige Durchschnittskurven ergeben sich nicht nur bei Synchronisation auf den Handlungsbeginn („response-locked average“); je nach Experiment eignen sich mehrere Ereignisse für die Synchronisation der Durchschnittsbildung, etwa der Startschuss des einzelnen Versuchsdurchlaufs oder ein stets dargebotener Reiz („stimulus-locked average“; Terminologie aus Miller et al, 2011, S. 105, fig. 1).
- 3 Einige Details dazu in Libet et al, 1983, S. 632–634; Libet, 1985, S. 533.
- 4 Siehe Libet et al, 1983, S. 630 (Tabelle 1B), S. 631 (Tabelle 2B, Zeile II); Libet, 1985, S. 533 (Tabelle 1B, Zeile II).
- 5 Siehe Libet et al, 1983, S. 629, 631 (Tabelle 2D); Libet, 1985, S. 533 (Tabelle 1A, Spalte W).
- 6 Siehe Libet et al, 1983, S. 623, 635, 631 (Tabelle 2C, Zeile II); siehe Libet, 1985, S. 529, 532. Siehe auch Libet, 1999 sowie deutsch Libet, 2004, viertes Kapitel.
- 7 Soon et al, 2008; Haynes, 2011; Bode et al, 2011. Bis heute wurden die aufsehenerregenden Ergebnisse dieser Untersuchungen keiner detaillierten wissenschaftstheoretischen Analyse unterzogen. Die dort verwendeten Methoden beschreibt Falkenburg, 2011, S. 147–154.
- 8 Siehe dazu z.B. Bieri, 2005; Beckermann, 2004; Dennett, 2003; Köchy, 2006; Meuter, 2007; Nida-Rümelin, 2005; Pothast, 1987; Wingert, 2004. Die Frontlinien in dieser Debatte folgen zum Teil einem verwirrenden Verlauf. Dafür zwei Beispiele: Erstens entkräftet Pothast eine Reihe kompatibilistischer Argumente, um sich auf die Seite der Freiheitsgegner zu schlagen, und zwar letztlich aufgrund zutiefst moralischer, ja humanistischer Erwägungen (Pothast, 1987, S. 361–422). Und zweitens setze ich mich zwar seit längerem gegen die naturalistischen Spielarten des Kompatibilismus zur Wehr, finde aber einen genuin metaphysischen Kompatibilismus nicht unattraktiv; meiner Ansicht nach sind die ontologischen Kosten für den Kompatibilismus weit höher, als viele Autoren annehmen (O.M., 2007, insbes. fünfter Abschnitt; O.M., 2010, insbes. dritter und neunter Abschnitt). Von allen diesen Feinheiten hängt für die augenblicklichen Zwecke nichts ab.
- 9 Trotz einiger kleiner methodischer Differenzen wurden Libets Ergebnisse gut reproduziert von Haggard et al, 1999. Nichtsdestoweniger wurden spätere raffiniertere Versuche durchgeführt, deren Ergebnisse in entgegengesetzte Richtungen weisen und die ich hier nicht erörtern kann; siehe Trevena et al, 2010. Vergl. dazu Nida-Rümelin, 2010.
- 10 Diese beruhigende Tatsache gilt nicht in gleichem Maße für die neueren technischen Mittel der Freiheitsforschung von Haynes und Mitstreitern, siehe Soon et al, 2008; Haynes, 2011; Bode et al, 2011.
- 11 Für besonders ausgefeilte kompatibilistische Positionen kommt es entscheidend auf den Unterschied zwischen diesen Ebenen an, siehe z.B. Walde, 2006. Der Inkompatibilismus wird dagegen besonders plausibel,

- wenn er sich auf die ontologische Ebene konzentriert. Den Streit zwischen diesen Grundpositionen kann ich hier ausblenden, da ich den Blick für etwas anderes freibekommen möchte.
- 12 Ich nenne stellvertretend für viele Stellen aus Peirces umfangreichem Œuvre nur Peirce, 1934, S. 90 (§ 5.145), S. 112–131 (*Lecture VII*).
 - 13 Auch andere naturwissenschaftliche Schlussweisen sind auf ähnliche Weise fallibel, gelten also nur bis auf weiteres. Das betrifft sogar kausale Beweise per *experimentum crucis*. Zwar lässt sich z.B. aus Newtons berühmtem *experimentum crucis* ein überzeugender Beweis seiner Optik ableiten (O.M., i.E., Teil I); trotzdem kamen später Experimente und darauf aufbauende neue Argumente ans Tageslicht, die Newtons Schluss unterminieren (Rang et al, 2010).
 - 14 Trevena et al, 2002, S. 163–165; Miller et al, 2002, S. 308, siehe dazu Rösler, 2006, S. 173/4, 179 und Pauen, 2004, S. 207.
 - 15 Mit Bedacht drücke ich mich so vorsichtig aus und vermeide es, den fraglichen elektrischen Gehirnprozess mit der Entscheidung zu identifizieren. Ob diese Identifikation angemessen wäre oder nicht, hat keinen Einfluss auf meine Argumentation. Mir geht es um eine formale Frage, die sich vor derartigen Interpretationen klären lässt.
 - 16 Lassen Sie sich nicht vom negativen Vorzeichen beunruhigen. Die wissenschaftliche Einheit für elektrische Potentiale ist so normiert, dass die uns interessierenden Messergebnisse unter Null liegen; trotzdem hat sich die elliptische Rede vom *Anstieg* des Bereitschaftspotentials eingebürgert – streng genommen müssten wir vom Anstieg des Betrags des Bereitschaftspotentials reden. Um der Kürze willen bleibe ich bei der elliptischen Ausdrucksweise. Ebenso will ich eine weitere Komplikation ausblenden: In Wirklichkeit sinkt das Durchschnittspotential kurz vor der Handlung wieder ab; darauf kommt es im folgenden nicht an.
 - 17 Mittels verwandter (hypothetischer) Zahlenverhältnisse illustrieren Miller, Trevena und Rösler dieselbe Sache, siehe Trevena et al, 2002, S. 164; Rösler, 2006, S. 174.
 - 18 Diese logische Diagnose Trevenas und Millers wurde offenbar von fast allen ihren Lesern geteilt, siehe Trevena et al, 2002, S. 309.
 - 19 Im Original: „smearing artifact“, siehe Trevena et al, 2002, S. 163.
 - 20 Ich werde mich im folgenden auf diese *logischen* Behauptungen konzentrieren, also auf Behauptungen darüber, dass das eine nicht aus dem anderen folgt. Insofern ich solche Behauptungen angreife, brauche ich mich nicht mit der Faktenfrage auseinanderzusetzen, ob den Versuchspersonen die einzelnen Bewegungsentscheidungen vor oder nach dem jeweiligen Anstieg des Bereitschaftspotentials bewusst geworden sind. Trevena und Miller sprechen sich (auf der Grundlage weiterer Experimente) dafür aus, dass Libets Schlussfehler keinen Faktenfehler nach sich gezogen hat (Trevena et al, 2002, S. 175, 177/8).
 - 21 Wenn A_i und B_i gegeben sind, so kann man C_i immer so einrichten, dass D den Durchschnitt bildet; $C_i(t) := 3 D(t) - A_i(t) - B_i(t)$. Nicht alle diese Fälle sind neurophysiologisch sinnvoll; aber auch an sinnvollen Fällen herrscht kein Mangel.

- 22 Daraus, dass der Schluss wasserdicht ist, ergibt sich indes nur, dass die postulierten ungestörten Einzelkurven *eine* Erklärung der Durchschnittskurve bieten; ob es die beste Erklärung ist, ergibt sich daraus noch nicht.
- 23 Die Situation ändert sich nicht nennenswert, wenn wir diesen Abstand vergrößern (wodurch sich das Zeitfenster für die Freiheit weiter öffnet) oder wenn wir ihn verkleinern (wodurch es mehr und mehr zugeht und sich im Extremfall von 0 ms fast ganz schließt. In diesem Extremfall wären Entscheidung und Potentialanstiegsbeginn zeitlich identisch.
- 24 Wer sich in Libets verzwicktes Zahlenwerk vertieft, dem wird klar: Es gibt Versuchspersonen, die sich in keinem der 40 Durchläufe einer Serie auch nur ein einziges Mal ihrer Entscheidung vor dem durchschnittlichen Anstieg des Bereitschaftspotentials bewusst wurden (etwa Versuchsperson S. B. in Serie 3, siehe Libet et al, 1983, S. 628, Abbildung 1; S. 630, Tabelle 1, Zeile 3). Das unterscheidet sich eklatant von der Situation, die ich in Abb. 2 anhand dreier Durchläufe illustriert habe und in der c_a deutlich *vor* t_b liegt. (Auf diese Diskrepanz hat meines Wissens bislang niemand in der Literatur zum Verschmierungsartefakt hingewiesen, und dort sind gleichartige Abbildungen gang und gäbe, siehe Trevena et al, 2002, S. 164; Rösler, 2006, S. 174). Um es deutlich zu sagen: Anders als im rein *hypothetischen* Fall der Abb. 2 (die bestimmte arithmetische Verhältnisse illustrieren soll) steht es uns bei der Analyse der *tatsächlichen* Experimente nicht offen, irgendwelche Zeitpunkte der bewussten Entscheidung passend zu postulieren. Denn im Experiment sind diese Zeitpunkte *einzel*n gegeben. (Jedenfalls in den Serien, in denen die Zeitpunkte absolut abgelesen wurden („absolute mode (A)“, Libet et al, 1983, S. 626).) Das bedeutet: Wir können bei Versuchspersonen wie S. B. selbst mithilfe des Verschmierungsartefakts keine *perfekt* freiheitsfreundliche Interpretation konstruieren – keine Interpretation, in der die Versuchsperson bei jedem einzelnen Durchlauf nachweislich frei gewesen wäre. Mindestens eine Ausnahme müssen wir zulassen; diese eine Ausnahme wird von der Ausreißerkurve beschrieben. Die Ausnahme können wir entweder (a) der vollständigen Freiheit zuliebe als singulären Messfehler wegerklären – oder aber (b) als seltene, aber doch vorkommende Ausnahme von der Regel, dass sich Menschen *immer* ihrer Entscheidung bewusst werden, bevor das Bereitschaftspotential ansteigt. (Bedauerlicherweise liefert uns Libet nur für S. B. die *einzelnen* Zeitpunkte der bewussten Entscheidung, siehe Libet et al, 1983, S. 628, Abbildung 1 oben. Er dürfte diese Zahlen repräsentativ gefunden haben, und darum stehen hinter vielen meiner Überlegungen die bekannten Zahlen für S. B. Es wäre attraktiv, weitere Einzelzahlen zu analysieren – aber fürs erste muss ich mich mit den Zahlen begnügen, die vorliegen. Jeff Miller hat mir freundlicherweise zugesagt, weitere Einzelzahlen aus halbwegs gelungenen Reproduktionen des Libet-Experiments zur Verfügung zu stellen; welche Hypothesen sich daraus konstruieren lassen, werde ich bei einer anderen Gelegenheit erörtern; vergl. Fußnote 31).

- 25 Denn der Durchschnitt aus dieser Kurve und 39 Nullkurven liefert Libets D-Kurve – jedenfalls bis zum ersten Knick nach A's Anstiegsbeginn.
- 26 Dürfen wir davon ausgehen, dass die Standardkurven exakt gleich verlaufen? Nein, das wäre überzogen. Aber wenn sie allesamt für die nachweislich *freien* Entscheidungen des Entscheiders einschlägig sein sollen, haben wir wenig Spielraum für ihre Variation. Denn die gemessenen bewussten Zeitpunkte der Entscheidung drängen sich in einem engen Intervall zusammen – jedenfalls bei der Versuchsperson, deren Einzelzahlen vorliegen, siehe vorletzte Fußnote. (Hier wurden 34 aller 40 bewussten Entscheidungen zwischen –215 ms und –42 ms datiert, und die anderen 6 Entscheidungen lagen noch ungünstiger). Wir haben also nur *kurz vor Null* einen gewissen Manövrierspielraum. Doch die Pointe meiner Rechnung liegt darin, dass die freiheitsfreundliche Interpretation (via Verschmierungsartefakt) eine Ausreißerkurve postuliert, die *schon lange vorher* aus dem Ruder läuft. Dafür spielen die kleinen, aber feinen Variationsmöglichkeiten in den letzten 200 ms keine Rolle.
- 27 Wenn wir Glück haben, hat die Versuchsperson in einem der Durchgänge ihre bewusste Entscheidung tatsächlich so früh datiert. Falls nicht (wie bei S. B., siehe vorige Fußnote), so postulieren wir, dass sie sich einmal beim Ablesen vertan hat, das ist Fall (a) aus Fußnote 24. Wer es attraktiver findet, unfreie Ausnahmen zu postulieren (Fall (b)), der müsste einige unwesentliche Details der folgenden Rechnung abändern; an den Größenordnungen ändert sich dadurch wiederum nichts.
- 28 Im Fall (a) aus Fußnoten 24 und 27 wäre die Rechnung nur dann erforderlich, wenn man den ermittelten Durchschnitt unverändert beibehalten wollte, trotz des postulierten Ablesefehlers.
- 29 Es gilt unter Neurophysiologen als unstrittig, dass Libets Experiment keine anderen Resultate geliefert hätte, wenn er mehr als 40 Durchläufe für die Durchschnittsbildung herangezogen hätte (Jeff Miller, mündliche Mitteilung).
- 30 Siehe Fälle (a) und (b) in Fußnote 24.
- 31 Hier ein weiterer Gesichtspunkt für genauere Analysen: Ich habe in meinen bisherigen Rechnungen eine Vereinfachung mitgemacht, von der man sich bei genauerer Analyse lösen müsste: Ich habe immer nur mit dem ermittelten *Durchschnitt* der Zeitpunkte gearbeitet, zu denen der Versuchsperson ihre Entscheidung jeweils bewusst wurde. Anders als bei den gemessenen Potentialen (die infolge des Rauschens für sich allein keine Aussagekraft haben) besagen die einzelnen Zeitpunkte der bewussten Entscheidungen schon für sich allein eine Menge; sie charakterisieren datierbare Ereignisse aus dem mentalen Leben der Versuchspersonen. Insbesondere könnte man sie für die Konstruktion der verschmierten Alternativerklärung heranziehen – wenn man sie hätte. Das ließe so: Man müsste diese Zeitpunkte $c_0, c_1 \dots, c_{39}$ der Reihe nach ordnen und die Konstruktion der postulierten 40 ungestörten Potentialkurven in dieser Reihenfolge anpacken. Die erste Kurve Z_0 müsste wie gehabt zur Ausnutzung des Verschmierungsartefakts sehr früh ansteigen (unter etwaiger Berichtigung des Werts für c_0) – alle anderen Kurven wären exakt auf die Zeitpunkte $c_1 \dots, c_{39}$ abzustim-

men: die Kurve Z_1 begönne erst 50ms nach c_1 anzusteigen, Z_2 begönne damit erst 50ms nach c_2 , usw. Dann hätte die erste Kurve bis zu vierzig Knicke und die letzte Kurve nur einen. Selbstverständlich könnte man die Knicke im nachhinein glätten; darauf kommt es wie gehabt nicht an. Eine solche Rechenübung hätte nur dann Sinn, wenn alle Zahlen einiger repräsentativer Serien vorlägen. Ob die einzigen komplett bekannten Zahlen aus Libets Schriften repräsentativ sind oder nicht (Fußnote 24), können wir nur vermuten.

- 32 Dieser Aufsatz geht auf Diskussionen zurück, die ich mit Hörerinnen und Hörern meiner Vorlesung „Freiheit und Naturwissenschaft“ im Sommersemester 2008 an der HU geführt habe; ich danke den Studierenden für massive, aber konstruktive Kritik. Dank an Matthias Herder für Überprüfung und Korrektur meiner Rechnungen, Erstellung der Abbildungen, Hilfe bei Endredaktion, Mitarbeit beim Literaturverzeichnis, Programmierung und Durchführung des Zufallsexperiments, sowie Dutzende treffender Verbesserungsvorschläge. Zwei anonymen Gutachtern danke ich für wertvolle Hinweise, die zur Verbesserung des Aufsatzes geführt haben.
- 33 <http://www.mathworks.com>. Der Quellcode des Programms kann bei Interesse vom Autor angefordert werden.
- 34 <http://www.philosophie.hu-berlin.de/institut/lehrbereiche/natur/mitarbeiter/zufallsexperiment-verschmierungsartefakt>.

Literatur

- Beckermann, Ansgar, 2004: Schließt biologische Determiniertheit Freiheit aus? In: Hermanni, Friedrich; Koslowski, Peter (Hg.): Der freie und der unfreie Wille – Philosophische und theologische Perspektiven. Paderborn: Fink, S. 19–32.
- Bieri, Peter, 2005: Untergräbt die Regie des Gehirns die Freiheit des Willens? In: Gestrinch, Christof; Wabel, Thomas (Hg.): Freier oder unfreier Wille? Handlungsfreiheit und Schuldfähigkeit im Dialog der Wissenschaften. Berlin: Wichern Verlag, S. 20–36.
- Bode, Stefan; He, Anna Hanxi; Soon, Chun Siong; Trampel, Robert; Turner, Robert; Haynes, John-Dylan, 2011: Tracking the unconscious generation of free decisions using ultra-high field fMRI. In: PLoS One 6 No 6, e21612. [Im Netz unter <http://dx.doi.org/10.1371/journal.pone.0021612>].
- Dennett, Daniel C., 2003: Freedom evolves. New York: Viking.
- Falkenburg, Brigitte, 2011: Mythos Determinismus. Wieviel erklärt uns die Hirnforschung? Berlin: Springer.

- Haggard, Patrick; Eimer, Martin, 1999: On the relation between brain potentials and the awareness of voluntary movements. In: *Experimental Brain Research* 126, S. 128–133.
- Harman, Gilbert H., 1965: The inference to the best explanation. In: *The Philosophical Review* 74 No 1, S. 88–95.
- Haynes, John-Dylan, 2011: Decoding and predicting intentions. In: *Annals of the New York Academy of Sciences* 1224, S. 9–21.
- Köchy, Kristian, 2006: Was kann die Neurobiologie nicht wissen? Bemerkungen zum Rahmen eines Forschungsprogramms. In: Köchy, Kristian; Stederoth, Dirk (Hg.): *Willensfreiheit als interdisziplinäres Problem*. Freiburg: Karl Alber, S. 145–164.
- Kornhuber, Hans Helmut; Deecke, Lüder, 1965: Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. In: *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere* 284, S. 1–17.
- Libet, Benjamin, 1985: Unconscious cerebral initiative and the role of conscious will in voluntary action. In: *The Behavioral and Brain Sciences* 8, S. 529–539.
- Libet, Benjamin, 1999: Do we have a free will? In: *Journal of Consciousness Studies* 6 No 8/9, S. 47–57.
- Libet, Benjamin, 2004: *Mind Time. Wie das Gehirn Bewusstsein produziert*. Aus dem Amerikanischen von Jürgen Schröder. Frankfurt/Main: Suhrkamp. [Erschien zuerst englisch 2004].
- Libet, Benjamin; Gleason, Curtis A.; Wright, Elwood W.; Pearl, Dennis K., 1983: Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. In: *Brain* 106, S. 623–642.
- Meuter, Norbert, 2007: *Natur und Kultur der Freiheit*. In: Heilinger, Jan-Christoph (Hg.): *Naturgeschichte der Freiheit*. Berlin: de Gruyter, S. 405–434.
- Miller, Jeff; Shepherdson, Peter; Trevena, Judy, 2011: Effects of clock monitoring on electroencephalographic activity. Is unconscious movement initiation an artifact of the clock? In: *Psychological Science* 22 No 1, S. 103–109.
- Miller, Jeff; Trevena, Judy Arnel, 2002: Cortical movement preparation and conscious decisions: Averaging artifacts and timing biases. In: *Consciousness and Cognition* 11, S. 308–313.

- Müller, Olaf, 2007: Die Diebe der Freiheit: Libet und die Neurophysiologen vor dem Tribunal der Metaphysik. In: Heilinger, Jan-Christoph (Hg.): Naturgeschichte der Freiheit. Berlin: de Gruyter, S. 335–364. [Im Netz unter <http://nbn-resolving.de/urn:nbn:de:kobv:11-100180617>].
- Müller, Olaf, 2010: Gott, Freiheit und Unsterblichkeit: Drei Postulate der Unvernunft? In: Grajner, Martin; Rami, Adolf (Hg.): Wahrheit, Bedeutung, Existenz. Frankfurt / Main: Ontos, S. 279–315. [Im Netz unter <http://nbn-resolving.de/urn:nbn:de:kobv:11-100196799>].
- Müller, Olaf, i.E.: Newtons Spektrum und Goethes Theorem: Farbe, Licht, Finsternis. Frankfurt/Main: Fischer.
- Nida-Rümelin, Julian, 2005: Über menschliche Freiheit. Stuttgart: Reclam.
- Nida-Rümelin, Julian, 2010: Libet and liberty. Unveröffentlichtes Manuskript eines Vortrags vom 14.5.2010 vor der Konferenz Models of Mind, Rom. [Im Netz unter http://www.julian.nida-ruemelin.de/wp-content/uploads/downloads/2012/07/Libet-and-Liberty_Models-of-Mind-Roma.pdf, zuletzt abgerufen am 11.12.2012].
- Pauen, Michael, 2004: Illusion Freiheit? Mögliche und unmögliche Konsequenzen der Hirnforschung. Frankfurt/Main: Fischer.
- Peirce, Charles Sanders, 1934: Lectures on pragmatism. In: Peirce, Charles Sanders: Collected papers of Charles Sanders Peirce. Volume 5. Pragmatism and pragmaticism. Herausgegeben von Charles Hartshorne und Paul Weiss. Cambridge: Harvard University Press, S. 13–131. [Aus dem Jahr 1903].
- Pothast, Ulrich, 1987: Die Unzulänglichkeit der Freiheitsbeweise. Zu einigen Lehrstücken aus der neueren Geschichte von Philosophie und Recht. Frankfurt/Main: Suhrkamp. [Erschien zuerst 1980].
- Rang, Matthias; Müller, Olaf, 2009: Newton in Grönland. Das umgestülpte experimentum crucis in der Streulichtkammer. In: *philosophia naturalis* 46 No 1, S. 61–114. [Im Netz unter <http://nbn-resolving.de/urn:nbn:de:kobv:11-100187051>].
- Rösler, Frank, 2006: Neuronale Korrelate der Handlungsausführung. Zur Validität der Experimente von Libet (1983). In: Köchy, Kristian; Stederoth, Dirk (Hg.): Willensfreiheit als interdisziplinäres Problem. Freiburg: Karl Alber, S. 165–190.
- Soon, Chun Siong; Brass, Marcel; Heinze, Hans-Jochen; Haynes, John-Dylan, 2008: Unconscious determinants of free decisions in the human brain. In: *Nature Neuroscience* 11 No 5, S. 543–545.

- Trevena, Judy Arnel; Miller, Jeff, 2002: Cortical movement preparation before and after a conscious decision to move. In: *Consciousness and Cognition* 11, S. 162–190.
- Trevena, Judy Arnel; Miller, Jeff, 2010: Brain preparation before a voluntary action: Evidence against unconscious movement initiation. In: *Consciousness and Cognition* 19, S. 447–456.
- Walde, Bettina, 2006: *Willensfreiheit und Hirnforschung. Das Freiheitsmodell des epistemischen Libertarismus*. Paderborn: mentis.
- Wingert, Lutz, 2004: Gründe zählen. Über einige Schwierigkeiten des Bionaturalismus. In: Geyer, Christian (Hg.): *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente*. Frankfurt/Main: Suhrkamp, S. 194–204.