philosophia naturalis

Journal for the Philosophy of Nature

KLOSTERMANN Band/Volume 50/2013 Heft/Issue 1

philosophia naturalis

Journal for the Philosophy of Nature

50 / 2013 / 1

KLOSTERMANN

The articles are indexed in *The Philosopher's Index* and *Mathematical Reviews.*

Abonnenten der Printausgabe können über Ingentaconnect auf die Online-Ausgabe der Zeitschrift zugreifen: *www.ingentaconnect.com/content/klos/philnat*

Zurückliegende Jahrgänge sind mit einer Sperrfrist von fünf Jahren für die Abonnenten von *www.digizeitschriften.de* zugänglich.

Wissenschaftliche Redaktion: Dr. Carsten Seck, Institut für Philosophie, Universität Bonn

Satz: Mirjam Loch, Frankfurt am Main / Druck: KM-Druck, Groß-Umstadt. Gedruckt auf alterungsbeständigem Papier ♾ ISO 9706.

ISSN 0031-8027

Oliver R. Scholz

Wissenschaftstheorie, Erkenntnistheorie und Metaphysik – Klärungen zu einem ungeklärten Verhältnis

Zusammenfassung

Ziel meines Aufsatzes ist es, das ungeklärte Verhältnis zwischen Wissenschaftstheorie, Erkenntnistheorie und Metaphysik zu klären. Dazu muss zunächst etwas über den Charakter der Philosophie gesagt werden (Abschnitt 2). Philosophie ist als eine Disziplin zweiter Ordnung von jeder Einzelwissenschaft verschieden. Insbesondere versucht sie „Was ist F?"-Fragen und „Wie möglich?"-Fragen zu beantworten. Anschließend stelle ich die Ziele und Aufgaben der Erkenntnistheorie (Abschnitt 3) vor und gehe schließlich auf das Verhältnis zwischen Wissenschaftstheorie und Erkenntnistheorie ein; dabei kommt die Metaphysik ins Spiel (Abschnitt 4). Es zeigt sich, dass die Wissenschaftstheorie drei unterschiedliche Projekte umfaßt, die sich ganz folgerichtig aus dem Charakter der Philosophie als einer Disziplin zweiter Stufe ergeben. In Anwendung auf die Wissenschaften ist eine Disziplin zweiter Stufe in drei einander ergänzenden Formen möglich.

Abstract

The paper aims at clarifying the relationship between philosophy of science, epistemology and metaphysics. I begin with a characterization of philosophy (section 2). Philosophy as a second order discipline differs from any of the individual sciences. Typically, it attempts to answer questions of the form "How is it possible that p (given r)?". Next I present the aims and tasks of epistemology (section 3), before, finally, I turn to the relationship between philosophy of science and epistemology. At this point, metaphysics enters the scene (section 4). It turns out that philosophy of science includes three different projects that result from philosophy's being a second order discipline. Applied to the sciences, a second order discipline is possible in three different forms that complement one another.

1. Einleitung[1]

Ausgangspunkt meiner Überlegungen ist eine Irritation. Auf der einen Seite scheint vollkommen klar zu sein, dass Erkenntnistheorie und Wissenschaftstheorie etwas – ja sogar sehr viel – miteinander zu tun haben: Sie sind beide zentrale Disziplinen der Theoretischen Philosophie. Man sollte sogar denken, dass sie miteinander enger zusammenhängen als mit jeweils anderen Disziplinen der Theoretischen Philosophie: Schließlich haben sie es beide mit dem menschlichen Wissen zu tun.

Schlägt man heutige Lehrbücher der Erkenntnis- und Wissenschaftstheorie auf, findet man auf der anderen Seite – von ganz wenigen Ausnahmen abgesehen – keinerlei Auskunft zu ihrem Verhältnis. Schlimmer noch: Typischerweise nehmen die beiden Disziplinen kaum Notiz voneinander. Schon die Inhaltsverzeichnisse von Lehrbüchern der Erkenntnistheorie weisen kaum Überschneidungen mit denen der Wissenschaftstheorie auf. In den einen ist vom Wissensbegriff, von den Erkenntnisquellen, von der Struktur der Rechtfertigung und von Antworten auf den Skeptizismus die Rede. In den anderen geht es um Kriterien der rationalen Hypothesen- und Theorienwahl, um Bestätigung bzw. Bewährung, um Erklärung und Prognose, um Gesetze, Dispositionen und Kausalität, um Raum und Zeit und Raum-Zeit. Die Erkenntnistheoretiker zitieren in aller Regel andere Erkenntnistheoretiker; die Wissenschaftstheoretiker zitieren andere Wissenschaftstheoretiker. Der Rest ist Schweigen.

Diese Sachlage ist um so verblüffender, als in anderen Epochen der älteren und jüngeren Geschichte viele Philosophen Erkenntnis- und Wissenschaftstheorie für identisch oder nahezu identisch gehalten haben. Man denke nur an Aristoteles, an Descartes oder an Kant und die Neukantianer.

Da es stets ratsam ist, den Diskussionsstand möglichst umfassend zu berücksichtigen, habe ich nach der Konsultation meiner privaten Handbibliothek eine systematische Literaturrecherche begonnen. Die Ausbeute war äußerst mager, erschreckend mager. Warum schweigt man sich über eine so nahe liegende Frage aus?

Zu der Frage, warum so wenig zu unserem Thema zu finden ist, möchte ich nur eine Vermutung wagen. Das Problem des Verhältnisses von Erkenntnis- und Wissenschaftstheorie ist ein spezielles Problem der Wissenschaftsklassifikation. Nun halten viele derlei Probleme

für nachrangig, wenn nicht für völlig uninteressant. Fragen der Wissenschaftsklassifikation, so hört man immer wieder, seien von allenfalls pragmatischem oder institutionellem Interesse. Sie interessieren den Bibliothekar, der entscheiden muss, wie er seine Bücher und Regale ordnet, ob er etwa für Erkenntnis- und Wissenschaftstheorie getrennte Standorte vorsieht; oder auch die Leitung einer Universität, die entscheiden muss, ob sie für Erkenntnis- und Wissenschaftstheorie verschiedene Professuren ausschreibt. Wie im folgenden deutlich werden soll, halte ich diese Einstellung für verfehlt: Die Frage ist tatsächlich von großem praktischen und theoretischen Interesse.

Bevor ich auf die Frage des Verhältnisses zurückkomme, werde ich (a) etwas über den Charakter der Philosophie sagen, insbesondere über die für die Philosophie charakteristischen Rationalitätsprinzipien. Anschließend stelle ich (b) kurz die Ziele und Aufgaben der Erkenntnistheorie vor und gehe dann (c) auf das Verhältnis zwischen Wissenschaftstheorie, Erkenntnistheorie und Metaphysik ein. Dabei werde ich auch Kritik üben an traditionellen und gegenwärtigen Konzeptionen dieser Disziplinen.

2. Was ist Philosophie?

Erkenntnistheorie und Wissenschaftstheorie sind philosophische Disziplinen, diese Gemeinsamkeit ist unstrittig. Beginnen wir also mit der Frage: Was ist Philosophie?

Es ist viel Tinte über die Frage vergossen worden, ob die Philosophie selbst eine Wissenschaft ist. Wenn man eine Mehrdeutigkeit in der Frage aufgedeckt hat, lassen sich die resultierenden zwei Fragen leicht beantworten. Mit der Frage, ob die Philosophie eine Wissenschaft ist, kann entweder gemeint sein, ob die Philosophie eine Einzelwissenschaft ist wie andere Einzelwissenschaften, oder, ob die Philosophie wissenschaftlich ist, d.h. sich von allgemeinen wissenschaftlichen Standards leiten läßt. Die erste Frage ist zu verneinen, die zweite zu bejahen.

Betrachten wir, um das zu erläutern, das berühmte Diktum des frühen Wittgenstein: „Die Philosophie ist keine Lehre, sondern eine Tätigkeit." (Wittgenstein, 1989, 4.112) Kurz zuvor hieß es: „Die Philosophie ist keine der Naturwissenschaften. (Das Wort „Philosophie" muß etwas bedeuten, was über oder unter, aber nicht neben den Natur-

wissenschaften steht.)“ (Wittgenstein, 1989, 4.111) Ich kann an diese Bemerkungen anknüpfen, möchte sie aber in einigen Punkten ergänzen und korrigieren. Die zweite Behauptung muss ergänzt werden, damit der entscheidende Punkt deutlich wird: „Die Philosophie ist weder eine der Naturwissenschaften, noch eine der Geisteswissenschaften; sie ist nämlich überhaupt keine Einzelwissenschaft. (Das Wort „Philosophie“ muß etwas bedeuten, was über oder unter, aber nicht neben den Einzelwissenschaften steht.)“ Weder deckt sich der Gegenstandsbereich der Philosophie mit dem einer der Einzelwissenschaften; noch wendet die Philosophie einzelwissenschaftliche Methoden an. Vielmehr verwendet sie eigene Methoden, wie die Begriffsanalyse bzw. Begriffsexplikation, die Methode der Gegenbeispiele, die skeptische Methode, die phänomenologische Epoché, die rationale Rekonstruktion, die Argumentationsanalyse sowie besondere philosophische Argumentationsformen (z. B. transzendentale Argumente, Paradigm case-Argumente).

Des weiteren hatte Wittgenstein gesagt, Philosophie sei „eine Tätigkeit“. Dem ist kaum zu widersprechen; freilich muss näher bestimmt werden, um was für eine Tätigkeit es sich handelt. Schließlich sind auch Häkeln oder Wandern Tätigkeiten. Wittgenstein gibt dazu Hinweise: Es geht ihm zufolge um die „logische Klärung der Gedanken“: „Die Philosophie soll die Gedanken, die sonst, gleichsam, trübe und verschwommen sind, klar machen und scharf abgrenzen.“ (Wittgenstein, 1989, 4.112) Die Philosophen des Wiener Kreises haben sich diesem Verständnis von Philosophie eng angeschlossen – freilich mit einer charakteristischen Einengung auf die Gedanken der Wissenschaft; nach Rudolf Carnap etwa besteht die „Tätigkeit der Philosophie in der Klärung der Begriffe und Sätze der Wissenschaft.“ (Carnap, 1931, 433).

Nun ist die Klärung der Gedanken zweifellos eine wichtige und zentrale philosophische Tätigkeit, aber keineswegs die einzige. Sie ist eine notwendige Vorarbeit, aber nicht schon das ganze Geschäft der Philosophie. Dies wird auch unsere Betrachtung der Erkenntnistheorie und der Wissenschaftstheorie zeigen.

Halten wir als Zwischenfazit fest: Philosophie – oder wenn Sie so wollen – Philosophieren ist eine Verstandestätigkeit, die sich gleichwohl von jeder Einzelwissenschaft grundlegend unterscheidet. Diese Unterschiede haben mit der besonderen Art des Verstehens zu tun, das sie anstrebt, und mit einer kritischen Haltung, die mit diesem Verstehensziel einhergeht. Philosophie ist eine kritische Disziplin zweiter Ordnung, die

auf Resultate früherer Vernunfttätigkeit reflektiert,[2] um diese und ihren Zusammenhang zu verstehen und gegebenenfalls zu korrigieren.

Diese Resultate früherer Verstandestätigkeit brauchen – nota bene – nicht bereits vollentwickelte wissenschaftliche Disziplinen zu sein. In diesem Punkt könnte der Terminus „Disziplin zweiter Ordnung" missverstanden werden. Es kann sich auch um Begriffe, Commonsense-Urteile, Argumentationen oder Theorien handeln.

Nach Platon und Aristoteles beginnt die Philosophie mit dem Staunen, der Verwunderung. Wenn die Verwunderung nicht so stark ist, dass sie sprachlos macht, versucht sie sich typischerweise in Fragen zu artikulieren. Viele denken bei den philosophiespezifischen Fragen zuerst an die „Was-ist-F?"-Fragen, die durch die platonischen Dialoge berühmt geworden sind: Was ist Tugend? Was ist Gerechtigkeit? Was ist Wissen?[3] Mindestens ebenso kennzeichnend für das philosophische Fragen sind Fragen der Form „Wie ist es möglich, dass p?" Während die „Was-ist-F?"-Fragen nach Definitionen von „F" oder weitergehend nach Theorien über F fragen, verlangen die Wie-möglich-Fragen nach philosophischen Erklärungen.[4] Sie sind typischerweise kontrastiv gemeint: „Wie ist es möglich, dass p, gegeben r?" Dabei stehen „p" und „r" für Annahmen. Sowohl bei p als auch bei r handelt es sich um Annahmen, die man prima facie akzeptieren möchte, zwischen denen jedoch eine Spannung besteht derart, dass das Vorliegen von r das Vorliegen von p auszuschließen scheint. In der Tat lassen sich viele der hartnäckigsten philosophischen Probleme in dieser Form darstellen. Drei Beispiele mögen hier genügen:

– Das Problem von Freiheit und Determinismus: Wie ist es möglich, dass wir uns frei entscheiden und frei handeln können, gegeben dass alle Ereignisse in der Welt kausal determiniert sind?
– Das Problem des Skeptizismus: Wie ist es möglich, dass wir etwas wissen, angesichts der bekannten skeptischen Hypothesen (Descartes' Annahme eines uns systematisch täuschenden Dämons, Putnams Hirn-im-Tank-Szenario, o.ä.)?
– Das Theodizee-Problem: Wie ist das große Leid in der Welt möglich, wenn es einen allmächtigen, allwissenden und allgütigen Gott gibt?

Philosophisches Verstehen besteht in vielen Fällen in einem Verstehen, wie es durch eine philosophische Erklärung, d.h. eine rational akzeptierbare Antwort auf eine Frage der Form „Wie ist es möglich, dass p, gegeben r?" hervorgebracht wird.

Zurück zum wissenschaftlichen Charakter der Philosophie: Die Philosophie teilt mit den Einzelwissenschaften die allgemeinen formalen Bedingungen von Wissenschaftlichkeit, die für jede rationale Suche nach Wahrheit und Verstehen gelten.[5] Dazu gehören das Bemühen um sprachliche Klarheit und um strenge intersubjektive Nachprüfbarkeit sowie die Forderung der rationalen Begründung. Im Kern handelt es sich um die Regeln rationaler Argumentation. Auch gibt es in den meisten philosophischen Disziplinen einen Forschungsstand und in den übrigen zumindest einen Diskussionsstand.[6]

Die Philosophie unterscheidet sich von den Einzelwissenschaften zum einen darin, dass sie keinen eingegrenzten spezifischen Gegenstandsbereich hat, sondern, wie es der Volksmund treffend sagt, „über Gott und die Welt" nachdenkt, zum anderen – noch wichtiger – darin, dass sie sich besonderen zusätzlichen Rationalitätsstandards unterstellt. Etwas philosophisch verstehen heißt stets auch es gemäß den philosophiespezifischen Rationalitätsprinzipien zu verstehen.[7] Was damit gemeint ist, wird deutlicher, wenn wir versuchen, zwei von ihnen explizit zu machen, die für das folgende von Bedeutung sind.

Wie oben bereits deutlich wurde, teilt die Philosophie zahlreiche Rationalitätsprinzipien mit den Einzelwissenschaften (eben die allgemeinen Regeln wissenschaftlicher Tätigkeit) und mit unseren alltäglichen Bemühungen um Erkenntnis und Verstehen, nämlich die allgemeinen Regeln theoretischer und praktischer Rationalität. Uns soll es jetzt um die *philosophiespezifischen* Rationalitätsprinzipien gehen.[8]

Philosophisches Verstehen, so meine These, zeichnet sich durch eine besondere kritische Haltung und besondere Standards rationaler Akzeptierbarkeit aus. Es ist nicht ganz leicht diese Rationalitätsprinzipien so zu formulieren, dass sie nicht missverstanden werden. Da ich mit meinen Formulierungen noch nicht wirklich zufrieden bin, werde ich ihnen klärende Bemerkungen beigesellen, um so zu mindest manchen Missverständnissen vorzubeugen.

(PHIL-RAT 1) Keine Meinung ist in der Philosophie allein deswegen rational akzeptierbar, weil sie vom Commonsense akzeptiert wird.

(PHIL-RAT 2) Keine Meinung ist in der Philosophie allein deswegen rational akzeptierbar, weil sie in einer Einzelwissenschaft akzeptiert wird.

Spätestens zu diesem Zeitpunkt vermute ich viele Einzelwissenschaftler bereits weit oben auf der sprichwörtlichen Palme. Auf Vermessenheit und Arroganz lauten gängige Vorwürfe gegen die Philosophen. Wenn Sie Fachwissenschaftler sind und mich dort oben wenigstens noch hören können, möchte ich jetzt Erläuterungen anschließen, die vielleicht manche Woge der Entrüstung bereits glätten können. Die Prinzipien philosophischer Rationalität sind nicht so zu verstehen, als forderten sie die Nichtbeachtung oder gar Verachtung von Commonsense und Einzelwissenschaften. Im Gegenteil: Wir haben allen Grund, den Commonsense und erst recht die Einzelwissenschaften ernst zu nehmen. Das Nicht-Akzeptieren ist mit anderen Worten begründungspflichtig. (In vielen Fällen ist die Beweislast sogar sehr hoch.) Aber die philosophische Rationalität fordert eine aufgeklärte, kritisch-reflektierende Haltung auch diesen Bereichen gegenüber. Weder vortheoretische Intuitionen noch wissenschaftliche Lehrmeinungen und ihre sprachlichen Artikulationen sind für die Philosophie sakrosankte Autoritäten; wenn gute Gründe vorliegen, kann sie von beiden abweichen. Solche Gründe liegen insbesondere dann vor, wenn der Commonsense und die Einzelwissenschaften entweder (a) bereits in sich unklar sind, wenn sie (b) in sich inkohärent sind oder wenn sie (c) sich untereinander widersprechen. Nicht wenige philosophische Probleme haben mit solchen prima facie-Konflikten zwischen dem Commonsense und den Resultaten der Einzelwissenschaften zu tun.

Zur weiteren Besänftigung der Einzelwissenschaftler mag auch die Erinnerung beitragen, dass das Etikett „Philosoph" keineswegs ein Prädikat ist, das die Menschheit ein für alle Mal in zwei stabile Klassen – die Philosophen und die Nicht-Philosophen – einteilt, sondern eine Rolle oder Haltung bezeichnet, die man eine bestimmte Zeit einnehmen und dann wieder verlassen kann, und die vor allem grundsätzlich jeder Person offen steht. Niemand hat also Grund sich zu beklagen. In den revolutionären Phasen einer Wissenschaft beispielsweise werden die Wissenschaftler, die mit dem Zustand ihrer Wissenschaft unzufrieden sind, selbst in stärkerem Maße philosophisch arbeiten und sich dabei von den philosophischen Rationalitätsstandards leiten lassen. (Albert Einstein ist ein gutes Beispiel.) Die Konsequenz, dass die revolutionären Phasen einer Wissenschaft sich allenfalls graduell von der Philosophie unterscheiden, stört mich keineswegs; im Gegenteil: sie scheint mir die Sachlage ganz richtig zu treffen.[9]

3. Was ist Erkenntnistheorie?

Wenden wir uns vor diesem Hintergrund der Erkenntnistheorie und der Wissenschaftstheorie zu. Ich beginne mit der Erkenntnistheorie. Sie ist eine der zentralen Disziplinen der Theoretischen Philosophie. In der Geschichte der philosophischen Terminologie tauchen das Wort „Erkenntnistheorie" und seine Entsprechungen in anderen Sprachen überraschend spät auf; im Deutschen etwa erst im 19. Jahrhundert, ähnliches scheint für den englischen Terminus „epistemology" zu gelten. Einen Begriff von dem Projekt gibt es freilich schon viel länger.

Vergegenwärtigen wir uns dazu klassische Kennzeichnungen der Erkenntnistheorie. Schon Platon stellte die grundlegenden Fragen „Was ist Wissen?" und „Inwiefern ist Wissen mehr wert als bloße wahre Meinung?" John Locke setzte sich in seinem *Essay concerning Human Understanding* das Ziel: „[…] to enquire into the Original, Certainty, and Extent of human Knowledge; together, with the Grounds and Degrees of Belief, Opinion, and Assent […]." (Locke, 1975, 43 (I, 1, 2)) Typisch für die Neuzeit ist, dass jetzt auch nach den Grenzen des Wissens gefragt wird. Immanuel Kant interessierte an der Frage: „Was kann ich wissen?" besonders die Unterfrage: „Was kann ich unabhängig von Erfahrung wissen?" Rudolf Eisler kennzeichnet in seinem „Wörterbuch" die Erkenntnistheorie als „[…] die Wissenschaft vom Wesen und den Prinzipien der Erkenntnis, vom Ursprung, den Quellen, Bedingungen und Voraussetzungen, vom Umfang, von den Grenzen der Erkenntnis" (Eisler, 1927, 389). Und auch in den neuesten Handbüchern finden sich ganz ähnliche Kennzeichnungen: „Epistemology is one of the core areas of philosophy. It is concerned with the nature, sources and limits of knowledge […]" (Klein, 1998, 362).

Die Ziele, Aufgaben und Projekte der Erkenntnistheorie, wie sie gegenwärtig betrieben wird,[10] lassen sich in sechs Gruppen ordnen:

(I) *Begriffsexplikative Aufgaben:* Explikation der zentralen erkenntnistheoretischen Begriffe, insbesondere „Wissen" und „(epistemische) Rechtfertigung".

Zu bemängeln ist, dass viele Erkenntnistheoretiker sich bei den explikativen Aufgaben auf den Begriff des propositionalen Wissens, also: „Wissen, dass p", beschränkt haben, wobei „p" für einen beliebigen Satzinhalt steht. (Beispiele: Er weiß, dass Caesar ermordet wurde. Er weiß, dass

Rom die Hauptstadt von Italien ist.) Dieser Begriff wurde traditionell wie folgt expliziert: *Ein Erkenntnissubjekt S weiß, dass p*, genau dann, wenn gilt: (i) S ist überzeugt, dass p; (ii) p ist wahr; (iii) S ist gerechtfertigt in seiner Überzeugung, dass p. Weitgehende Einigkeit besteht über die folgenden Punkte: Die ersten beiden Bedingungen sind noch nicht hinreichend, denn z. B. glückliches Raten ist noch kein Wissen. Die neuere Diskussion hat gezeigt, dass es sogar gegen die traditionelle Analyse, die auch die dritte Bedingung umfasst, *prima facie* Gegenbeispiele gibt. Dies hat dazu geführt, dass man sehr viel gründlicher über den Begriff der epistemischen Rechtfertigung nachgedacht hat. Meiner Ansicht nach kann man den Geist der traditionellen Analyse retten, wenn man eine Differenzierung im Rechtfertigungsbegriff vornimmt: Man muss zwischen einem subjektiven und einem objektiven Aspekt der Rechtfertigung unterscheiden. Es genügt nicht, dass ich über Gründe für meine Überzeugung verfüge, in deren Lichte ich epistemisch verantwortlich gehandelt habe; die Gründe müssen auch objektiv adäquat sein.[11]

Auch wenn diese revidierte Fassung der klassischen Wissensanalyse angemessen ist, bleibt ein Punkt kritisch anzumerken: Propositionales Wissen ist bei weitem nicht das einzige Ziel unserer kognitiven Bemühungen. Die Erkenntnistheorie sollte sich deshalb auch nicht auf eine Theorie des propositionalen Wissens beschränken, sondern alle kognitiven und epistemischen Desiderate, Fähigkeiten und Leistungen berücksichtigen und im Zusammenhang untersuchen. Dazu gehören neben Wahrheit, Rechtfertigung und Wissen-dass: Wissen-warum, Verstehen, Kohärenz, Systematizität u. a.[12] Diese Erweiterung, die in vollem Gange ist, ist nicht zuletzt von Bedeutung für die adäquate Einbeziehung der Wissenschaften.

(II) *Demarkationsaufgaben:* (a) Allgemein geht es dabei um Umfang und Grenzen menschlichen Wissens. (b) Hinzu kommen wichtige Binnenunterscheidungen, z. B.: Was ist a priori, d.h. erfahrungsunabhängig, wissbar, was ist nur a posteriori wissbar?

(III) *Die Auseinandersetzung mit skeptischen Hypothesen und Argumenten:* Besitzen wir überhaupt Wissen? Ist Wissen überhaupt möglich? Wenn ja, wie ist die Herausforderung durch die skeptischen Hypothesen und Argumente zu beantworten?[13]

(IV) *Das methodologische Projekt:* (a) Wie, d.h. mit welchen Methoden, können wir Wissen und andere epistemische Ziele und Desiderata

erlangen? Und mit Hilfe welcher Kriterien können wir feststellen, ob und inwieweit wir diese Ziele erreicht haben? (b) Wie, d.h. mit welchen Methoden, können wir Irrtum und irrationale Meinungsbildung (Aberglauben, Vorurteile, Wunschdenken etc.) vermeiden?

(V) *Das normative Projekt:* Was sollen wir glauben? Was sind unsere epistemischen Rechte und Pflichten? Was sind intellektuelle bzw. epistemische Tugenden?

(VI) *Das Projekt der Klärung des Wertes von Wissen und Rechtfertigung:* Ist es gut, Wissen bzw. rechtfertigende Gründe zu besitzen? Wenn ja, warum ist es gut? Wozu bzw. wofür ist es gut? Und inwiefern ist Wissen mehr wert als wahre Meinung?

4. Das Verhältnis von Erkenntnistheorie und Wissenschaftstheorie

Was ist nun Wissenschaftstheorie? Welche Ziele und Aufgaben hat sie? Diese Fragen will ich indirekt angehen über die Frage, wie sich Erkenntnistheorie und Wissenschaftstheorie zueinander verhalten.

Als ich nach einem Untertitel für meinen Vortrag gesucht habe, habe ich zunächst mit Vergleichen mit zwischenmenschlichen Beziehungen gespielt: Handelt es sich bei dem seltsamen heutigen Verhältnis von Erkenntnistheorie und Wissenschaftstheorie um zerstrittene Nachbarn, um eine zerrüttete Ehe oder – noch tragischer – um eine uneingestandene Liebe?

Zeitgemäßer ist eine Metaphorik aus einem anderen Bereich; ich meine die große Welt der börsennotierten Aktiengesellschaften. Besonders dramatisch ist ein Szenario, das Sie alle aus den Nachrichten kennen: die Übernahme. Solche Übernahmen werden danach unterschieden, ob sie „freundlich" oder „feindlich" durchgeführt werden. Unter einer „feindlichen" Übernahme, im Englischen „unfriendly" oder gar „hostile takeover", versteht man eine solche, die nicht im Einvernehmen mit der Leitung der Zielgesellschaft durchgeführt wird, sei es weil eine Verständigung nicht erreicht wurde oder – noch unfreundlicher – weil sie gar nicht erst versucht wurde.

Nun sind Lehrstühle und Institute keine börsennotierten Aktiengesellschaften. Dennoch kann es als Analogie erhellend sein, in bezug auf

unsere beiden Disziplinen einmal das „worst case"-Szenario einer feindlichen Übernahme durchzuspielen.

4.1 Erstes Szenario einer feindlichen Übernahme: Warum die Wissenschaftstheorie nicht die Erkenntnistheorie fressen sollte

Unter den Pionieren der Wissenschaftstheorie gab es nicht wenige, die meinten, die gesamte Philosophie laufe auf Wissenschaftstheorie hinaus. Erkenntnistheorie wäre dann – sofern man sie überhaupt gelten lässt – trivialer Weise ein Teil der Wissenschaftstheorie.

So behauptet etwa Rudolf Carnap: „Philosophie ist Theorie der Wissenschaft." und erläutert: „Philosophie ist Wissenschaftslogik, d.h. logische Analyse der Begriffe, Sätze, Beweise, Theorien der Wissenschaft, […]" (Carnap, 1934, 111 f.). Und W.V. Quine bekräftigt noch Jahrzehnte später: „[…] philosophy of science is philosophy enough" (Quine, 1953, 446).[14] Formulieren wir dies als These:

(FÜ-W 1) Philosophie erschöpft sich in Wissenschaftstheorie.

Mit solchen Thesen bringt man sogar die allermeisten Philosophen auf die Palme. Als erstes und wichtigstes fällt auf, dass die gesamte Praktische Philosophie wegfallen würde. Dazu würde sich heute niemand mehr versteigen. Aber auch die Eigentümer der Stammaktien der Erkenntnistheorie sollten ihre Zustimmung verweigern. Erkenntnistheorie erschöpft sich nicht in Wissenschaftstheorie, auch heutzutage nicht. Erinnern wir uns dazu an unsere Liste der erkenntnistheoretischen Ziele, Aufgaben und Projekte.

Der auffälligste Unterschied besteht wohl mit Bezug auf Projekt (III) *Die Auseinandersetzung mit skeptischen Hypothesen und Argumenten.* Während das anti-skeptische Projekt in der Erkenntnistheorie aus den obengenannten Gründen nach wie vor zentral ist, spielt es in der Wissenschaftstheorie, so weit ich sehe, nicht dieselbe fundamentale Rolle;[15] wenigstens gilt dies für den globalen Skeptizismus.

In der allgemeinen Wissenschaftstheorie geht man vielmehr in aller Regel davon aus, dass Wissenschaft nicht nur möglich ist, sondern auch dass es eine ganze Reihe mehr oder weniger hochentwickelter Wissenschaften wirklich gibt. Diese Wissenschaften bilden das Material, das der Wissenschaftstheoretiker beschreibt, rekonstruiert, präzisiert und im einzelnen auch kritisiert. In der speziellen Wissenschaftstheorie,

d. h. in der Metatheorie der Einzelwissenschaften, spielen globale skeptische Hypothesen keine Rolle. Beispielsweise hält sich ein Wissenschaftstheoretiker der Geschichtswissenschaften nicht bei der skeptischen Hypothese auf, dass die Welt erst vor fünf Minuten entstanden sein könnte.

An diesem Punkt bietet sich die Gelegenheit, auf einen der wenigen Vorschläge zum Verhältnis von Erkenntnis- und Wissenschaftstheorie einzugehen, die sich in der Literatur finden. Wie eingangs erwähnt, war die Ausbeute äußerst mager. Ein Vorschlag von Wolfgang Stegmüller, dem Doyen der deutschsprachigen Wissenschaftstheorie, verdient aber in jedem Fall erwähnt zu werden. Er findet sich, gut versteckt, im Band IV, I. Teilband seiner großen Bestandsaufnahme der Wissenschaftstheorie (Stegmüller, 1973, 23):

„Der Wissenschaftstheoretiker stellt die existierenden Wissenschaften nicht in Frage. Vielmehr versucht er deren Rekonstruktion *unter der Voraussetzung, daß eine rationale Rekonstruktion möglich ist*. Der Erkenntnistheoretiker geht dagegen noch einen Schritt weiter. Die Geltungsfrage wird bezüglich der verschiedenen Arten angeblicher wissenschaftlicher Erkenntnisse gestellt."

Kurz zuvor heißt es (ebd.):

„Alle Untersuchungen, die unter der Existenzannahme [gemeint ist: die Annahme der Existenz von Wissenschaften; O.R.S.] laufen, könnte man als wissenschaftstheoretisch im engeren Sinne charakterisieren, diejenigen Untersuchungen hingegen, in welchen die Existenzannahme ihrerseits erst *begründet* werden soll, als erkenntnistheoretisch."

Eine Formulierung Stegmüllers könnte freilich zu Missverständnissen Anlass geben: „Der Wissenschaftstheoretiker stellt die existierenden Wissenschaften nicht in Frage." Der Kontext macht aber bereits klar, wie dies gemeint ist, nämlich im Sinne von: „Der Wissenschaftstheoretiker stellt die Existenz von Wissenschaften nicht in Frage." Damit ist sehr wohl verträglich, dass der Wissenschaftstheoretiker an einzelnen Punkten ihrer Methodologie Kritik übt. Stegmüller redet also nicht einer rein deskriptiven Wissenschaftstheorie das Wort; er betont nur, dass der Wissenschaftstheoretiker von der faktischen Existenz von Wissenschaften ausgeht.[16]

Soweit der auffälligste Unterschied zwischen Wissenschaftstheorie und Erkenntnistheorie. Was die anderen fünf erkenntnistheoretischen Projekte betrifft, so gibt es zwar in der Wissenschaftstheorie mehr oder

weniger enge Parallelprojekte; freilich sind auch hier Unterschiede nicht zu übersehen.

(I) Bei den *begriffsexplikativen Aufgaben* geht es in der Erkenntnistheorie, wie gehört, um die Explikation der zentralen epistemischen und kognitiven Desiderata wie „Wissen", „Rechtfertigung" etc. In der Wissenschaftstheorie. wäre es dagegen wenig sinnvoll, mit der Explikation des zentralen Begriffs „Wissenschaft" selbst zu beginnen.[17] In der Regel stehen zunächst Begriffe für Spezialprobleme wie „Theorie", „Gesetz", „Erklärung", „Voraussage", „Bestätigung" u.ä. im Mittelpunkt.

(II) Auch bei den *Demarkationsaufgaben* sind in der Wissenschaftstheorie andere prominent: Hier geht es um Umfang und Grenzen der Wissenschaften; sowie um gewisse Binnenunterscheidungen, insbesondere zwischen Wissenschaften und Pseudo-Wissenschaften, Wissenschaften und Metaphysik, Wissenschaften und Alltagswissen, Wissenschaften und Künsten oder Technik.

(IV) *Das methodologische Projekt:* Wie, d.h. mit welchen Methoden, können wir wissenschaftliche Erkenntnis erlangen? Wie können wir ein unwissenschaftliches Vorgehen erkennen und vermeiden?

(V) *Das normative Projekt:* Was sollen wir glauben? Dabei sind u.a. Forscher-Tugenden zu untersuchen.[18]

(VI) *Das Projekt der Klärung des Wertes von Wissenschaft:* Ist es gut, Wissenschaften zu besitzen? Wenn ja, warum? Wozu bzw. wofür ist es gut?

Die Strategie, die Erkenntnistheorie zu schlucken, indem man Philosophie insgesamt auf Wissenschaftstheorie reduziert, erscheint somit wenig aussichtsreich.

Betrachten wir eine andere, prima facie vielleicht aussichtsreichere Übernahmestrategie. Ihre Grundidee lautet:

(FÜ-W 2) Nur die Wissenschaften gelangen zu echter Erkenntnis. Deshalb ist nur die Wissenschaftstheorie eine Erkenntnistheorie, die ihren Namen verdient.

Dazu ist folgendes zu sagen: (a) Die Ausdrücke „Wissen" und „Erkenntnis" haben keinen Komparativ. Alltagswissen, etwa Wissen aus einfachen Wahrnehmungen, ist um nichts weniger Wissen als wissenschaftlich gewonnenes Wissen. (b) Außerdem scheint die Übernah-

mestrategie davon auszugehen, dass sich die Erkenntnistheorie nur oder primär mit vor-wissenschaftlicher Erkenntnis befasst. Es trifft aber nicht zu, dass sich die Erkenntnistheorie nur mit Alltagswissen beschäftigt. Zwar entnimmt sie oft Beispiele aus diesem Bereich; aber nicht selten trifft sie Aussagen, die sich auch auf wissenschaftliche Erkenntnisse erstrecken.

Richtig ist allerdings, dass in den modernen Wissenschaften die Erkenntnismöglichkeiten durch instrumentelle, experimentelle und mathematische Mittel[19] sowie durch Simulationen enorm erweitert worden sind. Die Erkenntnistheorie tut gut daran, sich über diese Entwicklungen auf dem laufenden zu halten und die Tauglichkeit ihrer Leitbegriffe und -unterscheidungen in ihrem Lichte zu überprüfen.

4.2 Zweites Szenario einer feindlichen Übernahme: Warum die Erkenntnistheorie nicht die Wissenschaftstheorie fressen sollte

Eine naheliegende, aber zu simple Idee, lässt sich wie folgt formulieren:

(FÜ-E) Wissenschaftstheorie ist die Theorie wissenschaftlicher Erkenntnis. Infolgedessen ist die Wissenschaftstheorie schlicht und einfach ein echter Teil der Erkenntnistheorie.

Hier sollten nun die Aktionäre der Wissenschaftstheorie protestieren: Wissenschaftstheorie, wie sie heute international in Forschung und Lehre betrieben wird, erschöpft sich keineswegs in Theorien wissenschaftlicher Erkenntnis. Zwar schließt die Wissenschaftstheorie wesentlich auch eine Theorie wissenschaftlicher Erkenntnis ein, aber sie geht nicht darin auf. Neben der Theorie wissenschaftlicher Erkenntnis gehören, wie wir gleich sehen werden, mindestens zwei ebenso gewichtige Bestandteile zur Wissenschaftstheorie in ihrer heutigen Gestalt; und das ist nicht nur faktisch so; es ist auch gut so.

4.3 Was also ist Wissenschaftstheorie?

Bevor ich auf die beiden weiteren Teile der Wissenschaftstheorie eingehe, noch ein paar erläuternde Bemerkungen zum ersten Teil. Ein Teil der Wissenschaftstheorie ist (I) bereichsspezifische Erkenntnis-

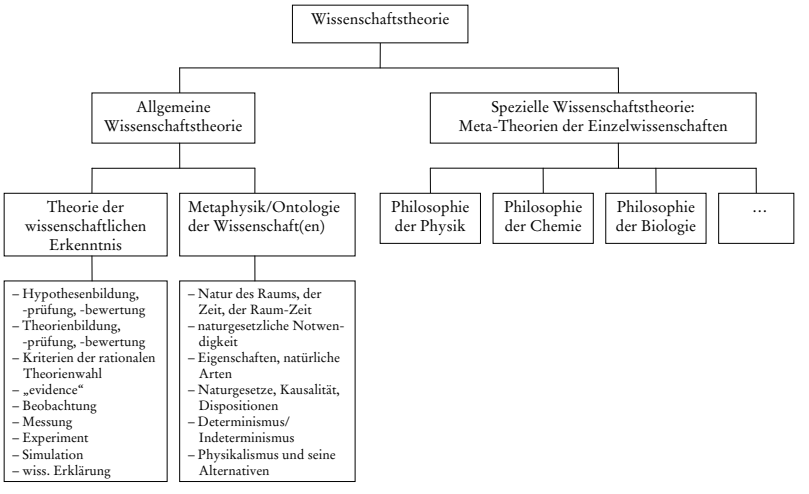
theorie bzw. angewandte Erkenntnistheorie: Es geht dabei insbesondere um Fragen der Hypothesen- und Theorienbildung, -prüfung sowie -bewertung. Da man Theorien benutzt, um Erklärungen zu geben und Prognosen zu machen, rücken auch diese Themen in den Vordergrund.

Daneben beschäftigen sich Wissenschaftstheoretiker aber aus gutem Grund mit weiteren Fragen: Sie untersuchen (II) die metaphysischen und ontologischen Grundlagen der Wissenschaften. Hierher gehören etwa die Diskussionen über Naturgesetze und naturgesetzliche Notwendigkeit, natürliche Arten, Kausalität, Dispositionen, Determinismus und Indeterminismus, über den Physikalismus und seine Alternativen etc.[20]

Die Theorie der wissenschaftlichen Erkenntnis (I) und die Metaphysik der Wissenschaft (II) bilden zusammen die allgemeine Wissenschaftstheorie oder vielleicht besser: *allgemeine Wissenschaftsphilosophie.*

Schließlich betreiben Wissenschaftstheoretiker mit guten Gründen (III) *Meta-Theorie der Einzelwissenschaften* oder kurz: *spezielle Wissenschaftstheorie*. In jeder Einzelwissenschaft gibt es besondere Begriffe, die zu explizieren und dabei in der Regel zu präzisieren sind. Auch hat jede Einzelwissenschaft besondere Methoden entwickelt, deren wissenschaftliche Tauglichkeit zu untersuchen ist. Und schließlich gibt es in jeder Einzelwissenschaft besondere metaphysische und ontologische Fragen. Anfangs stand bei den Meta-Theorien der Einzelwissenschaften die Philosophie der Physik im Vordergrund. Inzwischen gibt es viele Beiträge zur Philosophie der Chemie, der Biologie, der Psychologe, der Ökonomie, der historischen Wissenschaften, der Sozialwissenschaften usw. Da die spezielle Wissenschaftstheorie näher an der Wirklichkeit der Einzelwissenschaften ist, kann sie als Korrektiv der allgemeinen Wissenschaftstheorie dienen.

Insgesamt ergibt sich das folgende komplexe Bild der Wissenschaftstheorie (vgl. Diagramm): Hier sieht man nun auf einen Blick, dass die Erkenntnistheorie die Wissenschaftstheorie schon deshalb nicht fressen kann, weil sie allenfalls auf einen von zwei Teilen der allgemeinen Wissenschaftstheorie sowie auf einen kleinen Teil jeder Meta-Theorie einer Einzelwissenschaft einen legitimen Anspruch erheben könnte. In dem Bereich der Theorie der wissenschaftlichen Erkenntnis sollten Erkenntnis- und Wissenschaftstheoretiker enger zusammenarbeiten, als dies in den letzten Jahrzehnten der Fall gewesen ist.



Ist die Wissenschaftstheorie also ein Mischwesen aus drei heterogenen Teilen, deren Zusammenführung nur historisch zu erklären ist? Vor dem Hintergrund meiner Charakterisierung der Philosophie lässt sich die auf den ersten Blick hybride Gestalt der gegenwärtigen Wissenschaftstheorie leicht verstehen. Ich habe die Philosophie oben als Disziplin zweiter Stufe charakterisiert, die auf Verstandesprodukte erster Stufe (Begriffe, Urteile, Argumentationen, Theorien) reflektiert, um sie und ihren Zusammenhang zu verstehen. Bezogen auf den Untersuchungsgegenstand Wissenschaft läßt sich eine Disziplin zweiter Stufe nun gerade in drei unterschiedlichen Weisen entwickeln: (I) Man kann erstens auf die wissenschaftliche Tätigkeit als kognitiv-epistemische Tätigkeit reflektieren; dann sucht man nach einer *Theorie wissenschaftlicher Erkenntnis*. (II) Man kann zweitens die Gegenstände wissenschaftlicher Theorien auf ihre allgemeinsten Eigenschaften und Strukturen hin untersuchen; dann betreibt man *Metaphysik und Ontologie der Wissenschaften*. (III) Und man kann drittens Einzelwissenschaften zum Gegenstand der philosophischen Untersuchung machen; dann betreibt man spezielle Wissenschaftstheorie und arbeitet an *Meta-Theorien der Einzelwissenschaften*.

5. Zusammenfassung

Ich fasse zusammen: Erkenntnistheorie und Wissenschaftstheorie sind Disziplinen der Theoretischen Philosophie. Philosophie ist als eine Disziplin zweiter Ordnung von jeder Einzelwissenschaft verschieden. Sie reflektiert auf Verstandesprodukte erster Stufe, um sie und ihren Zusammenhang zu verstehen. Insbesondere versucht sie „Was ist F?"-Fragen und „Wie möglich?"-Fragen zu beantworten.

Die traditionelle Erkenntnistheorie konzentrierte sich auf die Fragen „Was ist Wissen?" und „Wie ist Wissen möglich?" Die neuere Erkenntnistheorie hat die Fragestellung zurecht erweitert: Insbesondere bezieht sie andere kognitive und epistemische Desiderata ein, darunter gerade auch solche, die für die Wissenschaften von Bedeutung sind.

In der Wissenschaftstheorie haben sich drei Disziplinen zweiter Stufe ausdifferenziert: (I) die *Theorie wissenschaftlicher Erkenntnis* – in diesem Bereich sollte die Wissenschaftstheorie wieder enger mit der Erkenntnistheorie zusammenarbeiten; (II) die *Metaphysik und Ontologie der Wissenschaften* und (III) die *Meta-Theorien der Einzelwissenschaften*. In diesem Bereich ist die Philosophie der Physik bereits sehr gut entwickelt; andere Meta-Theorien der Einzelwissenschaften laden noch zur Bearbeitung ein.

Die prima facie-Heterogenität wissenschaftstheoretischer Fragestellungen ist demnach keineswegs eine allenfalls historisch zu erklärende Anomalie, sondern ergibt sich ganz natürlich aus dem Charakter der Philosophie als Disziplin zweiter Stufe: In Anwendung auf die Wissenschaften ist eine solche Disziplin zweiter Stufe nämlich in den drei genannten voneinander verschiedenen, aber einander sinnvoll ergänzenden Formen möglich.

Anmerkungen

1 Frühere Fassungen wurden an den Universitäten Münster, Düsseldorf, Marburg, Leipzig und Dortmund zur Diskussion gestellt. Für Hinweise und Anregungen danke ich Thomas Bartelborth, Andreas Bartels, Dieter Birnbacher, Axel Bühler, Gerhard Ernst, Brigitte Falkenburg, Andreas Hüttemann, Peter Janich, Marie Kaiser, Markus Schrenk, Gerhard Schurz, Wolfgang Spohn, Christian Suhm und Christian Weidemann. Die Endfassung wurde im Rahmen der DFG-Forschergruppe „Kausalität,

Gesetze, Dispositionen, und Erklärungen am Schnittpunkt von Wissenschaften und Metaphysik" fertiggestellt.

2 Soweit stimme ich mit Rosenberg (1984) überein. Ich möchte jedoch betonen, dass sich philosophische Aussagen durchaus auf die Wirklichkeit beziehen können und nicht bloß auf unser Denken über die Wirklichkeit. Philosophen reflektieren unser Denken und Sprechen über die Wirklichkeit mit; dies hindert sie jedoch nicht daran, (wahre oder falsche) Aussagen über die Wirklichkeit zu machen.

3 Dazu ausführlicher Scholz (1999, 83–86).

4 Vgl. dazu Nozick (1981, 8 ff.) und Scholz (1999, 87–90).

5 Vgl. z. B. Stegmüller (1973, 5 ff.)

6 Institutionell schlagen sich diese Merkmale der Wissenschaftlichkeit z. B. in der Praxis der Begutachtung von Forschungsprojekten und von Beiträgen für philosophische Zeitschriften sowie in der Praxis der Bewertung von akademischen Qualifikationsarbeiten nieder.

7 Dazu ausführlicher Scholz (1999, bes. 91 ff.)

8 Einzeln genommen sind die unten beispielhaft angeführten Rationalitätsprinzipien noch nicht philosophiespezifisch. Meine Behauptung ist, daß (PHIL-RAT 1) und (PHIL-RAT 2) in Verbindung mit weiteren Prinzipien *zusammengenommen* philosophiespezifisch sind.

9 Dies wirft die Frage nach der historischen Variabilität philosophischer und einzelwissenschaftlicher Rationalitätsstandards auf; dazu Scholz (1999, 92f.)

10 Vgl. Williams (2001, bes. 1 f.)

11 Williams (2001, 22 f.) im Anschluß an Fogelin (1994, 27 f.). Williams hat den zweiten Teil von Fogelins dritter Bedingung zurecht abgeschwächt.

12 Vgl. Alston (2005, 39–47).

13 Dieses Projekt wirkt auf Nicht-Philosophen vermutlich besonders befremdlich. Warum sollte uns das Problem des philosophischen Skeptizismus interessieren? Dieses Problem ist von Interesse nicht etwa, weil es so viele radikale Skeptiker gäbe, die man bekehren müßte; auch nicht, weil es sich um ein brennendes praktisches Problem handeln würde; sondern weil wir durch die Untersuchung der skeptischen Argumente etwas über Wissen und Rechtfertigung lernen (methodologischer Grund) und weil wir mit den Antworten auf die skeptische Herausforderung eine Übersicht über alle grundlegenden philosophischen Theorien des Wissens und der Rechtfertigung gewinnen (metaphilosophischer Grund).

14 Nachgedruckt in: Quine (1976, 151).

15 Das heißt nicht, dass skeptische Überlegungen gar keine Rolle spielen (Stichwörter: Unterbestimmtheit und Unbestimmtheit); vgl. Humphreys (2004, 551).

16 Was wiederum nicht bedeutet, dass er etwa die Existenz jeder Wissenschaft anerkennen muß, von der irgendwann einmal behauptet worden ist, es gebe sie. Die Präsupposition des Wissenschaftstheoretikers ist vielmehr eine schwächere: nämlich, dass es überhaupt Wissenschaften gibt, die man rekonstruieren und untersuchen kann.

17 Dies betont auch Stegmüller (1973, 5).

18 Dazu Charpa (2001, Kapitel 2).
19 Dies betont Humphreys (2004, 549).
20 Historisch und institutionell ist die Wissenschaftstheorie auch ein Nachfolger dessen, was früher Naturphilosophie hieß.

Literatur

Alston, William P., 2005: Beyond „Justification". Dimensions of Cognitive Evaluation. Ithaca/London: Cornell University Press.

Carnap, Rudolf, 1931: Die physikalische Sprache als Universalsprache der Wissenschaft. In: Erkenntnis 2, S. 432–465.

Carnap, Rudolf, 1934: Über den Charakter der philosophischen Probleme. In: ders., 2004: Scheinprobleme in der Philosophie und andere metaphysikkritische Schriften. Herausgegeben, eingeleitet und mit Anmerkungen versehen von Thomas Mormann. Hamburg: Felix Meiner, S. 111,127.

Charpa, Ulrich, 2001: Wissen und Handeln: Grundzüge einer Forschungstheorie. Stuttgart/Weimar: J.B. Metzler.

Eisler, Rudolf 1927-30: Wörterbuch der philosophischen Begriffe. 4., völlig neubearbeitete Auflage. Berlin: Mittler & Sohn.

Fogelin, Robert, 1994: Pyrrhonian Reflections on Knowledge and Justification. Oxford: Oxford University Press.

Humphreys, Paul, 2004: Scientific Knowledge. In: Niiniluoto, Ilkka et al. (Hg.): Handbook of Epistemology. Dordrecht: Kluwer Academic Publishers, S. 549–569.

Klein, Peter, 1998: Epistemology. In: Craig, Edward (Hg.): Routledge Encyclopedia of Philosophy. Band 3, London/New York: Routledge, S. 362–365.

Locke, John, 1975: Essay concerning Human Understanding, hg. v. P. Nidditch. Oxford: Clarendon.

Nozick, Robert, 1981: Philosophical Explanations. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Quine, W.V., 1953: Mr. Strawson on Logical Theory. In: Mind 62, S. 433–451; nachgedruckt in: ders., 1976: Ways of Paradox and Other Essays. Cambridge, Mass./London: Harvard University Press, S. 137–157.

Rosenberg, Jay F., 1984: The Practice of Philosophy. Englewood Cliffs, N.J.: Prentice-Hall.

Scholz, Oliver R., 1999: Was heißt: etwas in der Philosophie verstehen? In: Raatzsch, Richard (Hg.): Philosophieren über Philosophie. Leipzig: Leipziger Universitätsverlag, S. 75–95.

Stegmüller, Wolfgang, 1973: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band IV: Personelle und Statistische Wahrscheinlichkeit. Erster Halbband. Berlin, Heidelberg, New York: Springer-Verlag.

Williams, Michael, 2001: Problems of Knowledge. Oxford: Oxford University Press.

Wittgenstein, Ludwig, 1989: Logisch-philosophische Abhandlung/ Tractatus logico-philosophicus. Kritische Edition. Herausgegeben von Brian McGuinness und Joachim Schulte, Frankfurt am Main: Suhrkamp.

Frank Hofmann / Ferdinand Pöhlmann

Seeing oneself through the eyes of others. Beckermann on self-consciousness

Abstract

Ansgar Beckermann's account of self-consciousness can be seen as an attempt to locate the origin of self-conscious states in social cognition. It is assumed that in order to acquire self-consciousness, a cognitive system has to 'see itself through the eyes of the others'. This account, however, is doomed to failure, for principled reasons. It cannot provide a satisfactory explanation of the special, identification-free reference of first-person thoughts and, thus, fails to explain crucial features of *de-se* attitudes. In addition, Beckermann's account exhibits various other shortcomings.

Zusammenfassung

Ansgar Beckermanns Theorie zur Erklärung von Selbstbewusstsein kann als exemplarischer Versuch verstanden werden, die Wurzeln selbstbewusster Zustände in sozialer Kognition zu suchen. Dabei wird angenommen, dass eine notwendige Bedingung zur Entwicklung von Selbstbewusstsein darin besteht, dass sich kognitive Wesen in sozialer Interaktion ‚mit den Augen eines anderen sehen' lernen. Dieser Ansatz scheitert aber aus prinzipiellen Gründen, da er die besondere, identifikationsfreie Referenz von Ich-Gedanken nicht erklären kann und damit eine Erklärung wesentlicher Züge von *de-se*-Zuständen schuldig bleibt. Zudem ist Beckermanns Theorie im Besonderen durch einige weitere, grundlegende Unzulänglichkeiten gekennzeichnet.

1. Introduction

Our primary goal in this paper is to describe and criticize a certain approach to self-consciousness which takes self-consciousness to originate in social cognition. As a paradigmatic and particularly clear example of this approach we take Ansgar Beckermann's specific account of self-consciousness (Beckermann, 2008; 2003). According to Beckermann, the ability to think of oneself as oneself, constitutive of self-con-

sciousness, arises from the task of representing the contents of others' mental states – as (sometimes) concerning oneself.

The first part of our paper summarizes Beckermann's account of self-consciousness (section 2). The rest of the paper is dedicated to a critique of this account. Some not too serious problems with the proposal are briefly mentioned and put to one side (section 3). These problems concern certain aspects of Beckermann's proposal which are of a quite general nature, like the notions of computational and causal role and the conceptual and indexical character of first-person states. Even if Beckermann could deal successfully with these problems, however, serious problems concerning directly the nature of self-consciousness would remain. These problems constitute the main difficulty for Beckermann's account, and we will discuss them in detail (section 4). The main difficulty is a generic problem for any 'social' account of self-consciousness: it cannot explain the peculiar *de-se* character of self-conscious states, since it cannot explain the peculiar reference of the first-person concept that is immune to error through misidentification.

2. Beckermann's account of *de-se* states

Let us begin with an overview of Beckermann's theory of self-consciousness which can count as a paradigmatic example of the approach towards an explanation of self-consciousness via social cognition.[1] To illustrate his account, Beckermann tells a story of a cognitive system, called 'Al'. Although one could get the impression that Beckermann takes this story to be a real account of the ontogenetic development of self-representations and self-consciousness, we think one should take it merely as a metaphor or illustration which is meant to describe various forms of self-knowledge and their natures, functions, and limitations. Accordingly, we will describe his story not as a series of developmental steps of one organism, but we will carve out and present the central claims of the view standing behind the story.[2]

Beckermann starts by noting that only entities that form *representations* of their environment are capable of having self-representations. This means that simpler organisms that admittedly can experience their environment and behave in it but do not form representations,[3] are *per se* not capable of forming self-representations. An example would be a

system whose behavior rests exclusively on stimulus-response-mechanisms.[4]

Organisms that form representations of their environment are termed 'cognitive systems' by Beckermann. Drawing on John Perry's work,[5] he claims that cognitive systems represent objects in their vicinity as standing in certain spatial and other relations to themselves due to the agent-relative roles these objects play with regard to them.[6] This means that cognitive systems do not represent the objects in their environment in some kind of universal coordinate system as, e.g., the AI-program SHRDLU does. Instead, they represent objects in an egocentric frame of reference. One might think that this mode of representation presupposes some kind of self-representation, but Beckermann explicitly denies this claim by stating that in order to egocentrically represent objects around oneself one merely has to represent the property of standing-in-a-certain-relation-to-oneself. No explicit self-representation is contained.[7] The relevant properties are agent-relative properties, and one can represent them '*en bloc*', as it were. The same holds true of the case of representations of the cognitive system itself: it is possible to represent internal states of the representing system without explicitly representing the system as such. According to Beckermann (and Perry, too), the reason for this is simply the fact that the kinds of representations that figured hitherto in the discussion – egocentric/agent-relative representations of the environment and proprioceptive representations – share the property that every token of them always concerns one and the same cognitive system, namely, the one that these representations belong to. It will never be the case that a particular cognitive system represents objects in its environment as standing in a relation to a cognitive system other than itself when engaged in egocentric perception; equally, it will never be the case that a cognitive system represents internal states of some other cognitive system when engaged in proprioception. Nevertheless, such representations amount to a kind of self-knowledge in that they implicitly purport to carry information about the spatial relations the representing system bears to the represented objects and about the internal states that are actually instantiated in the representing system. For that reason, Perry calls such knowledge 'agent-relative knowledge'.

By now, we have an organism which is capable of perceiving, and behaving in, its environment by means of its agent-relative knowledge. No explicit self-representation – i.e., no representation of the representing

system itself – is needed for these purposes. But the world does not only contain dead objects but *other cognitive systems* as well. Hence, it is very important for each cognitive system to know what the others are going to do, and whether they are friendly or hostile against oneself, in order to act appropriately. This is why cognitive systems not only represent objects and their properties but other cognitive systems and their mental states as well. In other words, cognitive systems are capable of *meta-representations*, i.e., representations of representations. The next step is a crucial one: in order to represent representations of others which have as their content the representing system itself, the representing system has to represent itself. That's where a self-representation comes into play. This self-representation is the core representation of a fully self-conscious system. Henceforth, it is not only used to represent others' representations of oneself, but also to represent one's own physical and mental properties and to develop a body schema (Beckermann, 2003, 183 f.).

But what exactly makes this self-representation a *self-conscious* representation of oneself, i.e., a *de se*-representation? All we have so far is a representation of a particular cognitive system that happens to be of the cognitive system itself. Now Beckermann claims that exactly those self-representations count as *de se* which are *equivalent* to the agent-relative representations the organism uses to orient in and interact with its environment. Equivalence in this case means *having the same computational role* (Beckermann, 2003, 186). This has two decisive consequences. Firstly, proprioceptive input leads not only to agent-relative representations but to explicit self-representations as well. Secondly, explicit self-representations inherit two features of agent-relative representations concerning actions. The first feature is that in particular situations, representations with agent-relative content immediately lead to certain actions, typically. Representations that contain explicit reference to the representing system are then supposed to have the same immediate impact on action. For example, when a ball is flying in one's direction, one will immediately try to catch it or to avoid being hit by it, typically. This will be the case no matter whether one represents merely the agent-relative roles of the ball or whether one explicitly engages in self-conscious representation.

The second feature is that both types of representations are connected with certain bodily movements that constitute certain types of actions. That is, to represent in an agent-relative manner some object of one's

vicinity means directly to know how to move one's body in order to interact with that object. This feature is also supposed to be transferred to the corresponding self-conscious representations.

With these two features inherited from agent-relative representations, self-representations play a special role among all representations of a cognitive system, according to Beckermann. The special role lies exactly in the action-related consequences those self-conscious representations have. As an illustration Beckermann cites Perry's famous example of a supermarket shopper who follows a sugar trail, trying to tell the owner of the damaged sugar package that he is making a mess and not realizing that he himself is the one with the damaged package. When he finally realizes his fault, his actions concerning 'the owner of the damaged sugar package' change significantly (Perry, 1993, 33).

To sum up Beckermann's account, one can say that *de-se* representations have the following distinguishing marks: they are used to represent others' representations of oneself and one's own physical and mental states, and they have an immediate significance for action due to their equivalence with agent-relative representations. In Beckermann's words: "*[D]e se* beliefs […] are characterized by the specific causal role they play within the cognitive system of a person and with regard to her or his actions" (Beckermann, 2003, 186).

3. Problems with Beckermann's account

In this section we will briefly note several problems for Beckermann's account. These concern certain aspects that are usually associated with self-conscious states in the philosophical debate. We will mention four problems: the issue of indexicality, a problem concerning causal-computational role and Beckermann's thesis of 'equivalence', the issue of the conceptual character of self-consciousness, and the issue of the identity of the first-person concept (over time). We think that these four issues raise serious questions and problems, but we put them to one side. The main purpose of this section is to prepare the ground for presenting the real, fundamental difficulty, by setting it apart from the four issues just mentioned. The fundamental difficulty with Beckermann's account (and, indeed, any 'social' account of self-consciousness) will be presented in the next section.

A first question concerns the *indexicality* of the first-person concept. Beckermann relies heavily on ideas of John Perry concerning different types of 'self-related knowledge.' He incorporates in his story the notions of knowledge of the person one happens to be, agent-relative knowledge and self-attached knowledge, the fundamental triad in Perry's work. But regarding another central claim of Perry's, namely the essential indexicality of the first-person concept, Beckermann does not say a word. Perry claims that some indexical expressions, namely, those that express 'locating beliefs' about who one is, which time it is, and where one is, are not substitutable by other (non-indexical) expressions without loss of explanatory force (Perry, 1993). It's not clear whether Beckermann regards this feature of 'I' as non-essential as, e.g., Ruth Millikan does (Millikan, 1990), or whether he thinks that in his tale about the genesis of self-representations the indexicality is somehow incorporated into his account. The first interpretation is supported by the fact that he claims that all objects, including the representing system itself, are represented by internal 'names,' which by definition are not indexical expressions, we take it. But the second interpretation seems plausible as well, since Beckermann claims that Al would express representations that are about itself by the indexical 'I' (Beckermann, 2003, 184).

A second issue concerns causal-computational role and Beckermann's 'equivalence thesis'. A central element in Beckermann's account is the notion of a 'role' a representation has. On the one hand, there is the special *computational role* that the self-representation has (Beckermann, 2003, 186; 2008, 77 f.). On the other hand, there is its special *causal role* (Beckermann, 2003, 186f.; 2008, 79 ff.). The causal role consists in the special, unmediated action-relevance that agent-relative representations have and that is inherited by the self-representation via its equivalence with agent-relative representations. This sounds quite right, and it is widely accepted that *de se*-representations have this kind of action-relevance. But the problem with Beckermann's account regarding this point is twofold. Firstly, he gives no explanation of the unmediated action-relevance of agent-relative representations, he only states this supposed fact by re-telling some examples from Perry. These examples in turn rely on the unargued assumption that Fregean modes of presentation can be interpreted as causal roles. Even if this is the case, neither Beckermann nor Perry provides any explanation of it. The second problem lies in the postulated equivalence between agent-relative representations and

self-representations. Even if we take it for granted that agent-relative representations have this special causal role, Beckermann does not put forward anything which could make it intelligible that self-representations inherit this role – he merely postulates this. To give some evidence how this might happen, Beckermann should have presented (within his story) a somewhat more detailed account of how the equivalence between agent-relative representations and self-representations evolves. But he only states that it does evolve.

It seems, though, that the *causal role* is dependent on the *computational role* a representation has in the cognitive architecture of a cognitive system. Beckermann states that the causal role is a 'consequence' of the equivalence between agent-relative and self-representations, and he writes that "'being equivalent' here mean[s] simply 'having the same computational role'" (Beckermann, 2003, 186).

So, perhaps we just have to look at the computational roles of these representations. Unfortunately, the situation here is the same, if not worse. The computational role is explained by Beckermann in terms of 'modes of presentation' or Fregean contents: representations with the same mode of presentation/Fregean content correspond to the same computational role (Beckermann, 2003, 185). Hence, self-representations have a special mode of presentation, which Beckermann calls "EGO-mode of presentation" (Beckermann, 2003, 185). But Beckermann does not tell us more about it, except that it is a "very special way" (Beckermann, 2003, 185) in which one is given to oneself. Furthermore, Beckermann states that the computational role of self-representations has *two* special features. The second feature is the peculiar *causal role* we already encountered. The first feature is that the proprioceptive input is not only represented in terms of agent-relative representations but also explicitly in terms of self-representations. Hence, it seems that the computational role of self-representations is simply special because proprioceptive input is related to them and they have a close connection to action. But if we take Beckermann literally, even these features are not peculiar to self-representations, since agent-relative representations have exactly the same properties. For, the equivalence between the two ensures that they have the same computational role.

A third problem with Beckermann's account concerns the *conceptual character* of self-representation. Beckermann writes that only "self-knowledge in a strong sense" requires that the cognitive system develops

a concept of itself (Beckermann, 2008, 75). This seems to imply that the self-representation is a concept. But Beckermann does not say anything about the consequences that being a concept could have for the self-representation. Instead, he claims that conceptual self-representations and agent-relative representations, which form a kind of self-knowledge that is not dependent on a concept of oneself, are very similar in that they have the same computational roles (cf. Beckermann (2008, 75)). Does this mean that there exist non-conceptual and conceptual forms of self-consciousness side by side? Or is only the second kind a kind of self-consciousness? Does the difference between the two only consist in that the second one refers explicitly to the representing system while the first one does so only tacitly or implicitly? If that is the only difference, wouldn't that mean that conceptual and non-conceptual representations of particular things are only different in their degree of explicitness? Beckermann does not say anything that could give us an answer to these questions.

Another point relating to the issue of conceptual character is the question of the *identity* of the self-representation. Beckermann claims that the first-person concept develops from the social interaction and the need for meta-representations of others' mental states coming with it. Hence, at the beginning we have a self-representation that has a specific role among the representations of a cognitive system, namely, to represent oneself as one figures in mental states of others. But later, it will have the computational role mentioned above. It seems that important features of the first-person concept have changed, and one could wonder whether the first-person concept before the development is the same concept as the concept after it. Because one could argue that concepts are (partly) defined by their computational role, it seems at least doubtful that this is the case. Unless Beckermann gives an argument in favor of the identity of the relevant concept over time, the story might seem incoherent.[8]

4. The fundamental difficulty with Beckermann's theory

Aside from the problems just mentioned we think that the fundamental difficulty with Beckermann's account – and with the entire 'social approach' to self-consciousness – resides in the fact that *it cannot explain*

the peculiar de-se *character of self-conscious representations*. The means available to Beckermann are not suitable for the task of explaining the *de-se* character of self-representations. (This problem, we submit, is generic to the entire 'social approach' to self-consciousness, and not just a problem for Beckermann's specific version.) One cannot explain the *de-se* character of self-conscious states by reference to a role in social metacognition. This is what we would like to argue in the remainder of this paper.

The fundamental difficulty relates to two important features of self-consciousness: its *reference* and its *immunity*. Let us explain. The first-person concept refers, we take it, *pace* Anscombe. It is self-referential, in the sense of referring to the subject. For example, Al's first-person concept refers to Al. (Self-referentiality in this sense is simply reflexivity, and does not require or involve any *de-se* mode of presentation.) But the first-person concept refers in a special way. It refers in a way which allows for representations which are immune to error through misidentification with respect to the first person. For the sake of brevity, let us call this feature 'immunity'. Immunity in this sense lies at the heart of the special *de-se* character of self-conscious states. Indeed, one can take it as the crucial criterion for self-consciousness. The debate about self-consciousness has focused on immunity ever since Strawsons's, Shoemaker's, and Evans' works. We will follow their lead here. So let us suppose that what needs to be explained about the first-person concept is its reference and immunity.[9]

The crucial question now is whether Beckermann's account can explain the reference and immunity of the first-person concept. In the following we would like to argue that the answer is negative. Beckermann's account lacks the resources for such an explanation. We take this as a sufficient reason for concluding that his account cannot explain self-consciousness, since reference and immunity are the crucial features of self-consciousness in need of explanation.

Let us begin by taking a closer look at *immunity*. Roughly, a first-person representation is immune (with respect to the first-person position)[10] just in case it is impossible that it is false because, and only because, the subject represents of someone that something is true of that person, but fails to identify that person with herself.[11] For example, the first-person representation whose content can be expressed by the utterance 'I see a pink elephant' can be false because the elephant I am

currently seeing is in fact grey, but it cannot be false because it is in fact you who sees the elephant but not me (at least, normally). Hence, this representation is immune to an error on the ground that I misidentify someone seeing an elephant with me. Whether the representation is immune or not is not determined by the object the judgment refers to, but by the kind of information that grounds the judgment (that is, the way in which the judgment is arrived at).[12] The best characterization of immunity is given by *identification-freedom*. Because I do not identify myself, there is no possibility of mis-identification. This is why I cannot misidentify someone else as me. Immune first-person representations are identification-free. Reference to myself is not mediated by any detection of identifying properties, i.e., it is not mediated by any identification of myself.[13]

Immunity can be spelled out in detail in various different ways. But what we will argue will not depend on which of these various more precise statements of immunity one favors. The crucial feature of immunity, for our purposes, is identification-freedom. This is, more or less, Evans' understanding of the phenomenon.[14] And from now on, we will rely on this basic understanding. Our argument will be independent of any further details.[15]

Let us now consider the issue of *reference*, and the explanation of reference, of the first-person concept. For many philosophers, the first-person concept is referential, i.e., each token refers to a certain individual (namely, the one which is exercising the first-person concept on that occasion). Famously, Elisabeth Anscombe has denied that the first-person concept refers.[16] But it seems fair to say that her view is rather an implausible position, and there are not many who have followed her. Now, let us suppose that the first-person concept is a referential representation. (And clearly, Beckermann agrees.) This raises the question how we can explain its reference. Following Reichenbach, we can say that the token-indexical rule *describes* the reference of the first-person concept: any token of the first-person concept, occurring within the thought t, refers to the thinker of this thought t. This, however, does not provide an *explanation* of why a token of the first-person concept refers to the subject to which it in fact refers. And one can wonder whether it is not an important theoretical task to illuminate and explain how the first-person concept refers.

If we take together these two features of the first-person concept,

its reference and its immunity, it seems quite clear that in principle, an explanation of reference could be given by recourse to the causal-computational role – and, indeed, an explanation which is perfectly in line with immunity. The basic idea for such an explanation is the special causal-computational role that the first-person concept has in relation to egocentric perception and proprioception. The first-person concept is tied *immediately* (i.e., without any mediation by identifying properties) to proprioceptive experiences that represent the subject's bodily states. If, for example, Al feels pain in his left knee, Al is inclined to immediately form the first-person representation 'Al feels pain in his left knee.' (This is so if Al's internal name 'Al' really is a first-person representation.) Similarly, 'Al' is immediately linked to egocentric contents in Al's perceptual experience. If Al's perceptual state represents a tree-in-front-of-Al, then Al is inclined to immediately form the first-person representation 'A tree is in front of Al.' Proprioception and egocentric perception provide information about Al to which Al's first-person concept is sensitive *without the help of any mediating identification.* Now, plausibly, because the information is always about Al, this concept refers to Al. As Beckermann emphasizes, proprioception and egocentric perception always provide information about one and the same object, namely, the cognitive system itself (Al). And this is why no identification is required. If – perhaps *per impossibile* – proprioception could sometimes concern some other cognitive system, identification would become necessary. Only because there is no such variation in the object of proprioception, identification is superfluous. We have a kind of (structural, non-accidental) 'informational constancy' which makes identification superfluous. (Similarly, egocentric content always relates things to one and the same cognitive system, Al.) As Récanati nicely puts it: "The subject himself does not need to be explicitly represented, since the representation can only be about him and his situation."[17] The concept that Al acquires on the basis of proprioception and egocentric perception allows for immune self-representations, since it refers to Al and does so without any identification of Al (i.e., without the need for any uniquely identifying descriptive content).

We can generalize the result. Whenever a system has a set of representational states that always concern the system itself – which exhibits informational constancy –, it seems possible to 'introduce' a concept of the system which allows for immune self-representation. By its causal-

computational role, this concept is sensitive to these representational states, and it can refer without identification.

By now, surely we do not yet have a full-blown explanation of reference and immunity. But at least, we have an idea and a sketch of such an explanation. And it does not seem hopeless to think that the searched-for explanation could be given along these lines. So we have reason to assume that this is the right track for explaining reference and immunity.

This raises a problem for Beckermann's account. According to Beckermann, the origin of Al's first-person concept is social metacognition. The primary job of the first-person concept is to represent Al as occurring in the contents of others' representations. Now, it may be the case that the first-person concept performs this job. But does this help to explain its crucial features, reference and immunity? On reflection, the answer is negative. It seems hopeless to try to understand reference and immunity by looking at the role of the first-person concept in social metacognition. The reason for this is simply the fact that *there is no informational constancy* to be found here – in contrast to the just-mentioned explanation in terms of proprioception and egocentric perception. Others do not always represent Al, they represent other conspecifics as well. Probably, they will represent Al only in a minor fraction of all cases. So there is not even any approximation of informational constancy. Al has to find out whether another cognitive system represents Al or some other, third cognitive system. It would be wildly incorrect to assume (by default) that others always represent Al. Therefore, the explanatory idea just mentioned cannot be transferred to, or mirrored within, the social metacognition account. Something else needs to be provided as an explanation of reference and immunity, and it is hard to see what could do the job. At least, Beckermann does not provide it, and we cannot even see any hint in his account.

Beckermann holds that Al's internal name of Al comes to acquire a certain causal-computational role (the one he tries to describe by speaking of 'equivalence'). The alleged 'equivalence' between Al's internal name and states with agent-relative content consists essentially in a certain causal-computational role of this name, and it is an important element in his account. At the same time, Al's internal name of Al is used, by Al, in order to represent Al as occurring in the representational contents of others' states. So the internal name plays two important roles at the same time. Now, however, the question is which role explains what.

And given the difference with respect to informational constancy just pointed out, it seems clear that the situation is quite asymmetric. The causal-computational role explains reference and immunity, the role in social metacognition does not. If this is so, we have to conclude that the origin of the first-person representation lies in the system's own representational states exhibiting informational constancy (proprioception and egocentric perception), not in social metacognition.

The following diagnosis seems plausible. It is not an accident that the connection to proprioception and egocentric perception occurs in Beckermann's theory. This part is doing the explanatory work, for the explanation of reference and immunity. Once we have that work done, the first-person concept can be recruited for some other task, such as the task of social metacognition and theory of mind. But Beckermann commits a mistake if he locates the origin of self-consciousness in social metacognition. It is another and distinct part of his overall theory which explains self-consciousness – or, at least, could provide the basic material for such an explanatory account. The origin of self-consciousness lies in whatever accounts for reference and immunity, if it lies anywhere at all. Al can begin "to see himself through the eyes of others", but only if, and since, Al already possesses self-consciousness (Beckermann, 2003, 184).

Finally, let us take a look at a possible argument in defense of Beckermann. Beckermann might suggest that his goal was to explain self-consciousness in so far as it is necessary for dealing with a certain task; for other tasks, representations with agent-relative contents are sufficient. Therefore, self-consciousness has its 'origin' in social metacognition. Or so Beckermann might claim.

This argument fails, however, and it fails for two reasons. First of all, what is claimed within this argument is not correct, viz., that a self-conscious representation of Al is only necessary when it comes to social metacognition. And secondly, even if this were correct, it would not improve the situation with respect to the question of how we can explain the *de-se* character of the first-person concept. An explanation of immunity and reference would still be lacking. We would still not understand how Al's representation of Al could be self-conscious. Let us argue for these two points in the following.

Firstly, the argument in defense of Beckermann contains a false claim. The claim is that a self-conscious representation of Al is only necessary

when it comes to social metacognition. To see why this claim is false, one has to consider the issue of the *form* or *format* of representation, i.e., of the distinction between conceptual and non-conceptual content. Arguably, proprioception and egocentric perception have non-conceptual content, whereas thoughts and other propositional attitudes possess conceptual content. We take it as an empirically well-established fact about human cognition that perception (including proprioception and imagery) is processed in a way which is quite different from the way in which conceptual categorizations are processed. The best explanation for this is the assumption that they differ in the form or format of representation. If this is so, the need for acquiring a first-person *concept* is already in place when it comes to information about the system's own bodily state (proprioception) and the agent-relative roles of objects in the system's environment (egocentric perception). The proprioceptive and perceptual states are not suitable for thought, since they do not have the requisite kind of form or character – they are not concept involving. Therefore, if one accepts a distinction between conceptual and non-conceptual representation and assigns non-conceptual content to proprioception and egocentric perception, Beckermann's argument fails. Even if all the information is already contained in these non-conceptual states, a first-person concept is needed in order to bring it into the realm of thinking (with all its inferential capacities and processing). Al not only wants to (proprioceptively) *perceive* the pain in his left knee. Al also wants to be able to *think* that there is a pain in his left knee. Therefore, Al needs a first-person concept. Without such a concept Al's thinking would be 'blind' to the information contained in his perceptual states.[18]

Beckermann could resist this counter-argument by rejecting the distinction between conceptual and non-conceptual representation. (We believe that this would be a rather desperate move.) But even then his argument would not succeed. For, consider egocentric contents in perception. These perceptual states represent agent-relative properties of objects in Al's environment. For example, a red apple is represented as red and as being-one-meter-in-front-of-Al. The spatial feature of the apple is represented *en bloc*, as it were. But certainly, Al not only wants to represent apples as having spatial locations relative to Al. He also wants – and needs – to represent apples as standing in the very same spatial relations to other objects. Therefore, Al needs a representation

of the spatial relation being-one-meter-apart-from, and not only of the impure spatial property of being-one-meter-in-front-of-Al. And then he needs an explicit representation of Al in order to apply the former representation of the spatial relation, if he wants to represent Al as standing in this relation to some other object. Of course, whether Al 'needs' a certain representation depends on the tasks Al is supposed to solve. So 'needing' a kind of representation is relative. But it seems quite clear that the 'need' for a representation of spatial relations (and not just of impure, agent-relative spatial properties) is quite strong, since it allows for a *much more general* application which is useful for a potentially unlimited number of cases. (Of course, many inferential relations will become accessible only by having the complex representation, comprising two singular representations of the objects and a representation of the spatial relation. Only then will many logical relations become 'visible'.)[19]

If, on the other hand, Beckermann wanted to insist that the 'need' for a self-representation that comes from the task(s) of social metacognition is strict and absolute, we have to respond that, on reflection, this is not really true. In principle, one could represent others' mental states concerning oneself in an *en-bloc* fashion, in the same way in which egocentric representations represent the spatial facts in an *en-bloc* fashion. For example, Al would represent some other cognitive system as having the property of representing-Al-as-friendly. Of course, such a way of representing social mental facts would be vastly impractical, perhaps to the point of being no longer computationally manageable. (The range of application would not be general, as one can put it.) But in principle such a way of representing these facts is possible. So even there the 'need' is not strict or absolute.

Now, let us move on to our second point. We submit that even if the claim that we have just criticized were correct, this would not improve the situation concerning the explanation of the *de-se* character of self-conscious representations. Suppose that Al needed an explicit representation of Al for the task of social metacognition. This would not show, however, how this representation gets the two crucial features of reference and immunity. It could still be the case that what explains these features is something quite different from the role in social metacognition; and in particular, it could still be the case that what explains these features is a causal-computational role vis à vis proprioception and

egocentric perception. Therefore, we have to conclude that Beckermann does not explain self-consciousness by means of social metacognition – even if the claim that we argued against above were correct. The explanation of self-consciousness may still reside in something quite different from social metacognition.[20]

Notes

1 Another example for this kind of approach is Musholt (2012). Tugendhat (1979) is sometimes interpreted as belonging to this approach, but he has rejected this interpretation (Tugendhat, 2005). Of course, the approach goes back to, or is inspired by, George Herbert Mead's philosophy, and it is manifest in Jürgen Habermas' work, in one way or another.

2 To give just one example of a question that would have to be addressed in a genuine, serious account of the ontogenetic development of self-consciousness: How can an organism that has absolutely no conception of itself come to represent that another subject is looking at *it*, or wants to interact with *it*? At least some rudimentary form of self-representation seems to be required. In general, Beckermann's story is rather a theory of the nature and function of self-consciousness than a developmental theory.

3 Beckermann identifies representations as "more or less stable internal structures, that stand for some aspects of the environment, even if they are not currently experienced." Beckermann (2008, 69). (Our translation.)

4 Beckermann gives as examples the crab that Churchland uses to illustrate eye-arm-coordination in Churchland (1986, 284), and the program SHRDLU, designed by Terry Winograd. Cf. Beckermann (2008, 68) and Beckermann (2003, 176), respectively.

5 Especially Perry (1998).

6 In Perry's terms, agent-relative roles of objects determine epistemic and pragmatic methods that are appropriate to get to know of or act on them, respectively. See Perry (1998, 84), and Perry (2002, 197ff.) (Here, his terminology has changed from 'agent-relative *roles*' to 'agent-relative *relations*'.)

7 This claim, although not questioned by Beckermann, is not entirely uncontroversial. For there are accounts that presume some form of self-representation even in egocentric representations of the environment, see e.g. Schellenberg (2007) and, less clearly though, Brewer (1992).

8 Originalism about concept identity, as recently proposed by Sainsbury, Tye (2011), would be a view of concepts that is favorable to Beckermann's theory. Originalism allows for changes in semantic features and/or computational roles, without loss of identity of the concept. Other views of concepts that individuate concepts semantically (such as Fodor's representational/computational theory) seem to be incompatible with semantic changes.

9 Bermúdez (2011) takes immunity (in this sense) as definitive of self-consciousness. We take it that immunity is a necessary condition, not a sufficient condition.
10 Strictly speaking, immunity is always relative to a certain position. For the sake of brevity, we will not always mention this qualification in the following. The context should make things clear enough.
11 Shoemaker develops this property with reference to Wittgenstein in Shoemaker (1968).
12 Cf. Gertler (2011, 216); Evans (1982, 218 f.). With regard to the first-person concept this characterization implies that not every judgment involving it is immune to such an error. And in general, other judgments concerning particular things (*de re*-judgments) can be immune in that sense, too. See Gertler (2011, 216) and Evans (1982, 219) for examples.
13 Here we want to point out that we distinguish between referring to oneself and identifying oneself. So referring to oneself is not *per se* an identification of oneself. Identification requires the detection of some property, or cluster of properties, sufficiently rich for determining the referent. Reference to oneself is possible without the detection of such a property, or cluster of properties. This is one of the lessons we take from Shoemaker. Note, furthermore, that the property, or cluster of properties, used for identification can contain indexical-demonstrative elements. (Récanati, 2007, uses 'identification' and 'articulation' more or less synonymously. Cf. ibid., 147, e.g. In this sense, then, reference is also to be distinguished from articulation.)
14 Cf. Evans (1982, 180 f.).
15 Perhaps, one should distinguish between 'absolute' and 'circumstantial immunity', as Shoemaker (1968) does. (Récanati thinks that Shoemaker is confused here. Cf. Récanati (2007, ch. 20.). And perhaps, immunity can be extended to the predicative position, as Bar-On (2004) suggest. Since we are interested only in the first-person concept, we will ignore any such extension. For further refinements and discussions of immunity see, for example, Pryor (1999), Coliva (2006), Récanati (2007). For our purposes, the basic understanding of immunity as identification-free reference and self-knowledge is sufficient.
16 Cf. Anscombe (1975).
17 Récanati (2007, 146). Here, Récanati echoes Perry's talk of 'implicit' and 'explicit representation'. Perry also speaks of the subject's being an 'unarticulated constituent' of the representation. Cf. Perry (1986).
18 The distinction between conceptual and nonconceptual content has by now become very wide-spread. It can be drawn and explained in various different ways, e.g., by reference to maplike vs. sentence-like representation (Tye) or analog vs. digital representation (Dretske), etc. It does not matter for the present purposes which of these further developments or explanations one favors, all we need is the distinction itself. It does also not matter whether the distinction concerns the contents or the representations (representational vehicles) or both.
19 Just to mention another 'need' for an explicit self-representation, the use

of spatial cognitive maps seems to require an element suitable for marking the system's own position (real or imagined) within its map.

20 Many thanks for valuable comments and points of criticism to Ansgar Beckermann, Hanspeter Mallot, Mark May, Wolfgang Röhrich, Peter Schulte, and Hong Yu Wong.

References

Anscombe, Gertrude Elizabeth Margaret, 1975: The first person. In: Guttenplan, Samuel D. (ed.): Mind and language. Oxford: Clarendon Press, pp. 45–64.

Bar-On, Dorit, 2004: Speaking My Mind: Expression and Self-Knowledge. Oxford: Clarendon Press.

Beckermann, Ansgar, 2003: Self-Consciousness in cognitive systems. In: Kanzian, Christian (ed.): Persons: An interdisciplinary approach. Proceedings of the 25th International Wittgenstein Symposium 11th to 17th August Kirchberg am Wechsel (Austria). Kirchberg am Wechsel: Österreichische Ludwig-Wittgenstein-Gesellschaft, pp. 175–188.

Beckermann, Ansgar, 2008: Gehirn, Ich, Freiheit: Neurowissenschaften und Menschenbild. Paderborn: Mentis.

Bermúdez, José, 2011: Bodily awareness and self-consciousness. In: Gallagher, Shaun (ed.): The Oxford Handbook of the Self. Oxford: Oxford University Press, pp. 157–179.

Brewer, Bill, 1992: Self-location and agency. In: Mind 101, pp. 17–34.

Churchland, Paul M., 1986: Some Reductive Strategies in Cognitive Neurobiology. In: Mind XCV, pp. 279–309.

Coliva, Annalisa, 2006: Error through misidentification. Some varities. In: The Journal of Philosophy 103, pp. 403–425.

Evans, Gareth, 1982: The varieties of reference. Edited by John Henry McDowell. Oxford, New York: Clarendon Press; Oxford University Press.

Gertler, Brie, 2011: Self-knowledge. New York: Routledge.

Millikan, Ruth Garrett, 1990: The Myth of the Essential Indexical. In: Noûs 24, pp. 723–734.

Musholt, Kristina, 2012: Self-consciousness and intersubjectivity. In: Grazer Philosophische Studien 84, pp. 75–101.

Perry, John, 1986: Thought without representation. In: Proceedings of the Aristotelian Society 137, pp. 137–152.

Perry, John, 1993: The problem of the essential indexical. In: Perry, John (ed.): The problem of the essential indexical: And other essays. New York: Oxford University Press, pp. 33–52.
Perry, John, 1998: Myself and I. In: Stamm, Marcelo (ed.): Philosophie in synthetischer Absicht: Synthesis in mind. Stuttgart: Klett-Cotta, pp. 83–103.
Perry, John, 2002: Identity, personal identity, and the self. Indianapolis: Hackett.
Pryor, James, 1999: Immunity to error through misidentification. In: Philosophical Topics 26, pp. 271–304.
Récanati, François, 2007: Perspectival Thought: A Plea for (Moderate) Relativism. Oxford: Oxford University Press.
Sainsbury, R. Mark; Tye, Michael, 2011: An Originalist Theory of Concepts. In: Aristotelian Society Supplementary Volume 85, pp. 101–124.
Schellenberg, Susanna, 2007: Action and Self-Location in Perception. In: Mind 116, pp. 603–632.
Shoemaker, Sydney S., 1968: Self-Reference and Self-Awareness. In: The Journal of Philosophy 65, pp. 555–567.
Tugendhat, Ernst, 1979: Selbstbewusstsein und Selbstbestimmung. Frankfurt a. M.: Suhrkamp.
Tugendhat, Ernst, 2005: Über Selbstbewusstsein: Einige Missverständnisse. In: Grundmann, Thomas et al. (eds.): Anatomie der Subjektivität. Frankfurt a. M.: Suhrkamp, pp. 247–254.

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-

philosophia naturalis 50/2013/1

natives are *ad hoc* and extremely improbable (as can be seen by constructing them explicitely). 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 störfeurigen 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 Augenblinzelns 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 wissenschafts*geschichtliche* 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 Gegengründe 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 Kompabilität abgeschwächt werden? Und vertrüge 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 welchen 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. 350 ms 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, Standardabwei-

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 hereinkommen, 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 Nullinie, 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 $-10\,\mu V$.[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 $t_\varnothing$ des Anstiegsbeginns dieser drei Potentialkurven liegt bei:

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

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 $t_\varnothing$ 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 $c_Ø$ 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 A_3, B_3 und C_3 bzw. A_4, B_4 und C_4 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 störfeurigen 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 ermessen 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, … P_{39} (in denen es

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

$$D(t) = \frac{\sum_{0}^{39} P_i(t)}{40}. \tag{10}$$

Und sie bildet auch nicht nur den Durchschnitt aus vierzig gleichartigen Kurven D_0, D_1, … 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:

$$D(t) = \frac{\sum_{0}^{39} D_i(t)}{40}, \text{ mit } D_i(t) \approx D(t). \tag{11}$$

Vielmehr bildet die Durchschnittskurve auch den Durchschnitt von vierzig Kurven Z_0, Z_1, … 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, … 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 µV auf –10 µV.

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

[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 $Z_1(t)$, … $Z_{39}(t)$. Der Durchschnitt aus allen diesen Kurven soll D(t) ergeben. Das führt zu folgendem Postulat:

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

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 … 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

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

$$(16)\quad A(t)=\begin{cases} 0\,\mu V \text{ für } t \text{ vor } -550\,ms; \\ -400/550\frac{\mu V}{ms}\,(t+550\,ms) \text{ für } t \text{ zw. } -550\,ms \text{ u. dem Knick.}^{25} \end{cases}$$

Wo liegt der Knick dieser Kurve? Genau bei dem Zeitpunkt, zu dem die kongruenten Standardkurven $Z_1(t)$, … $Z_{39}(t)$ 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 50 ms vor dem Anstiegsbeginn der zugehörigen Potentialkurve A getroffen wurde, also zum Zeitpunkt –600 ms.[27] Nun lag der Durchschnittszeitpunkt der Entscheidungen bei –200 ms. Nehmen wir an, dass er in den 39 standardgemäß verlaufenden Durchgängen $Z_1(t)$, … $Z_{39}(t)$ immer zum gleichen Zeitpunkt t' stattfand. Aus diesen Annahmen ergibt sich:

$$(17)\qquad -200\,ms = (39\,t' + -600\,ms)/40, \text{ also}$$

$$(18)\qquad t' = -7400\,ms/39 = -189{,}7\,ms.$$

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 10 ms nach hinten.[28] Da ich verlange, dass die 39 ungestörten Standardkurven immer genau 50 ms nach der jeweiligen bewussten Entscheidung anzusteigen beginnen, steht nun ein wichtiger Parameter dieser 39 Kurven fest. Sie beginnen ihren Anstieg bei –140 ms. Hier ihre Form:

$$(19)\qquad Z(t)=\begin{cases} \beta(t + 140\,ms) \text{ für } t \text{ später als } -140\,ms; \\ 0\,\mu V \text{ sonst.} \end{cases}$$

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:

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

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 µV, siehe (14)). 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 …, 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) $A(t) = 288/140\,\frac{\mu V}{ms}\,(t + 140\,ms) - 298\,\mu V$ für t später als $-140\,ms$.

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 $-10\,\mu V/550\,ms = -0{,}018\,\frac{\mu V}{ms}$ liegt und dass die Ausreißerkurve eine Steigung von $288/140\,\frac{\mu V}{ms}$ hat. Also gilt:

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

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

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

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

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

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

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 µV), 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:

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

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

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:

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

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 durchmischtes Zufallssignal. Darin überlagen 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 halbierter 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.

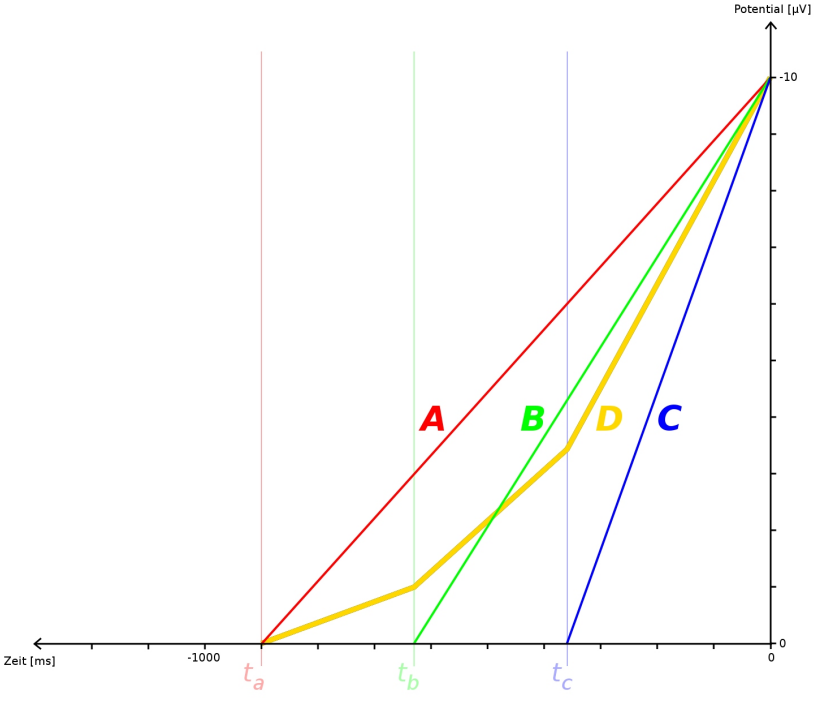


Abb. 1: Durchschnitt dreier Einzelkurven. Drei Potentialkurven A (rot), B (grün), C (blau) verlaufen zunächst auf der Nullinie (nicht farbig eingezeichnet) und beginnen bei drei verschiedenen Zeitpunkten t_a, t_b bzw. t_c linear anzusteigen. Die Durchschnittskurve D (gelb) beginnt ihren Anstieg zu dem 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.

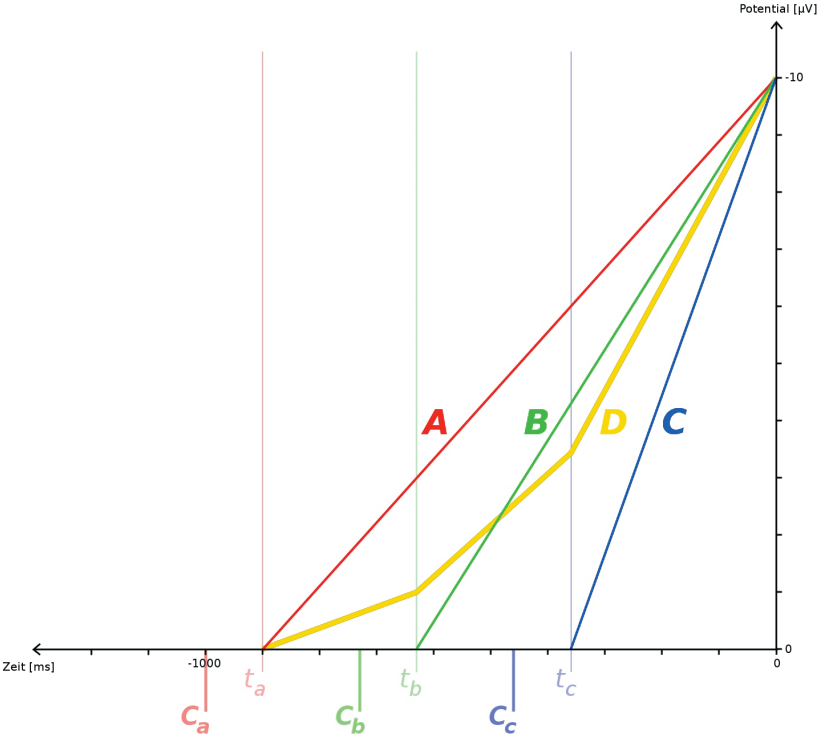


Abb. 2: Frühe (fiktive) Zeitpunkte bewusster Willensentscheidungen. Die neurophysiologischen Verhältnisse aus Abbildung 1 lassen sich mit erstaunlich frühen Zeitpunkten c_a, c_b bzw. c_c bewusster Willensentscheidungen kombinieren. Hier liegt z. B. jede Willensentscheidung 100 ms vor dem zugehörigen 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 *Durchschnitts*kurve 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).

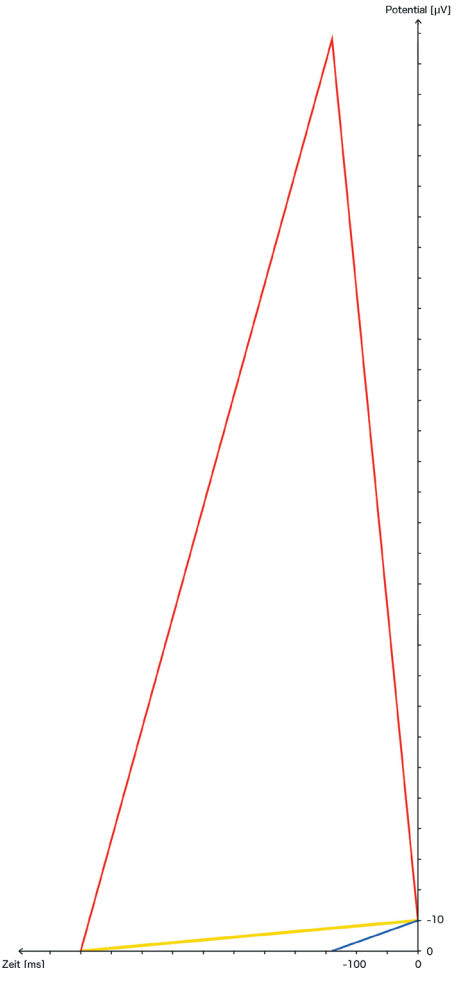


Abb. 3: Ausreißerkurve. Um die tatsächlichen Messungen Libets mit der Möglichkeit von Freiheit zu vereinbaren, kann man passende neurophysiologische Verhältnisse postulieren: Libets gemessene Durchschnittskurve (gelb) aus vierzig Versuchen kann sehr wohl durch (sich vierzigfach gegenseitig neutralisierendes) 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.

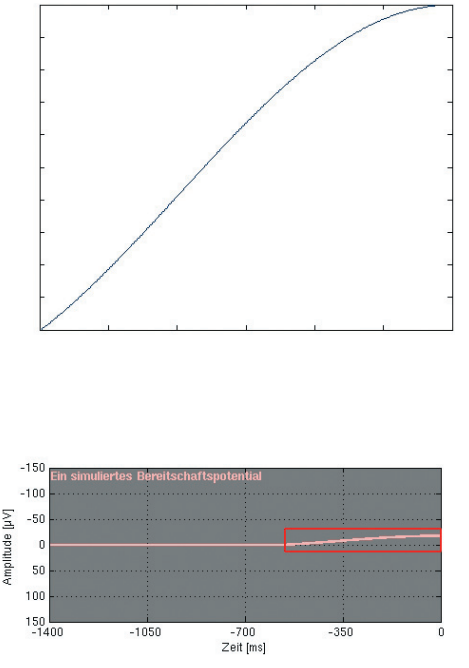


Abb. 4: Grundlagen fürs Zufallsexperiment. Im Zufallsexperiment haben wir keine linearen Anstiege benutzt (wie in Abbildungen 1 bis 3), sondern geeignete Ausschnitte aus Sinuskurven (oben), und zwar à la Libet parametrisiert (unten im eingerahmten rechten Teil der Abbildung). Bevor die Kurve unten sinusförmig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

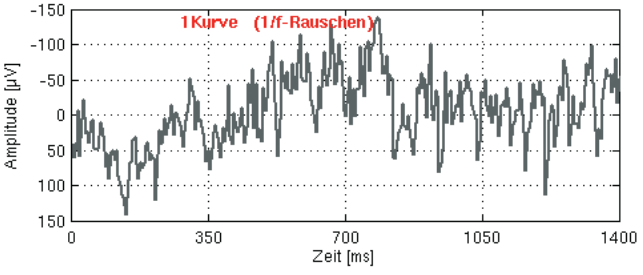


Abb. 5: Rosa Rauschen (1/f-Rauschen). In diesem Zufallssignal (dessen Ausschläge weit über Libets Durchschnittskurve liegen) überlagern sich diverse Schwingungen der unterschiedlichsten Frequenzen, Amplituden und Phasen. Enthaltene Schwingungen mit hohen Amplituden haben öfter niedrige Frequenzen und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

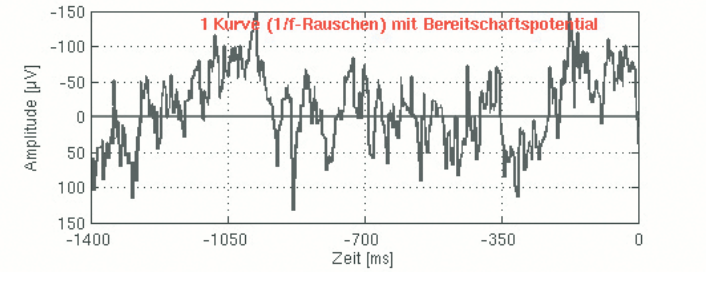


Abb. 6: Modell für eine Einzelmessung Libets. Hier ist das Rosa Rauschen (Abbildung 5) zu einem libet-artigen Anstieg des Bereitschaftspotentials (Abbildung 4 unten) addiert. Da das Rauschen weit höhere Ausschläge hat als das ungestörte Bereitschaftspotential, übertönt es das uns interessierende Signal.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

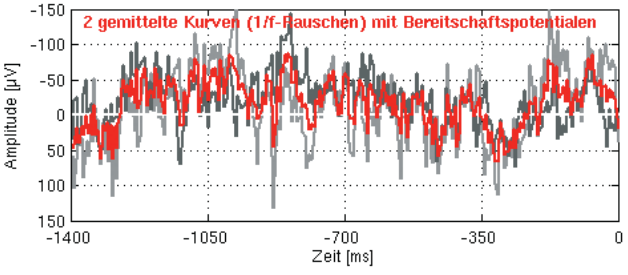


Abb. 7: Durchschnitt zweier Einzelmessungen. Das gewünschte Signal ist zwar noch nicht sichtbar. Aber wie man sieht, schlägt die Durchschnittskurve (rot) der Tendenz nach nicht mehr so stark nach oben und unten aus wie die Einzelkurven (hellgrau, dunkelgrau). Der Grund: Die zufälligen Signale neutralisieren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

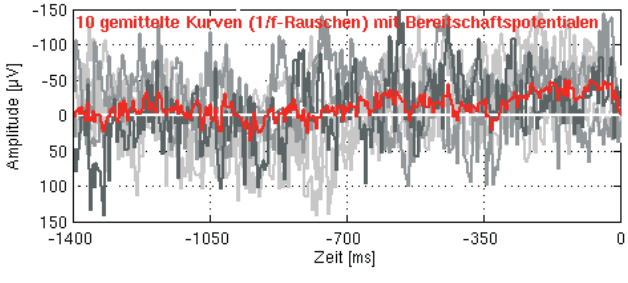


Abb. 8: Durchschnitt aus zehn Einzelmessungen. Das geübte Auge erkennt in der roten Durchschnittskurve bereits eine ansteigende Tendenz des Bereitschaftspotentials; richtig deutlich ist diese Tendenz noch nicht. Insbesondere lässt sich nicht gut erkennen, ob der Anstieg erst bei –350 ms beginnt oder 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.

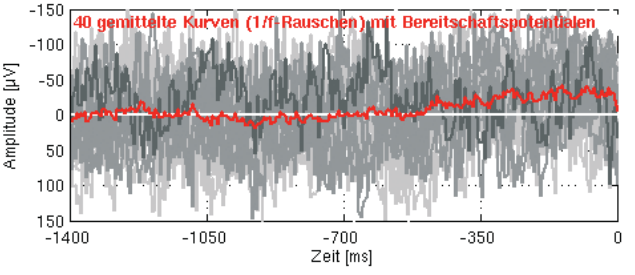


Abb. 9: Durchschnitt aus vierzig Einzelmessungen. Das gewünschte Signal lässt sich bereits gut aus der roten Durchschnittskurve erkennen. Die Einzelmessungen (deren letzte dunkelgrau und deren Vorgängerinnen immer blasser hellgrau gezeigt werden) neutralisieren sich in ihren hohen Ausschlägen, und sie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

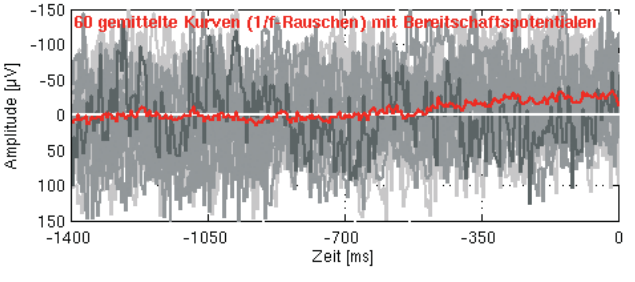


Abb. 10: Durchschnitt aus sechzig Einzelmessungen. Das bei Abbildung 9 erwähnte Zittern geht unmerklich zurück, die Durchschnittskurve nähert sich noch etwas deutlicher an das uns interessierende Signal an. Es ist vernünftig zu postulieren, dass allen sechzig Einzelmessungen jeweils eine ungestörte Kurve 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 Durchschnittbildung, 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 später 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 *einzeln* 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 liefe so: Man müsste diese Zeitpunkte c_0, c_1 … , 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 … , c_{39} abzustim-

men: die Kurve Z_1 begönne erst 50 ms nach c_1 anzusteigen, Z_2 begönne damit erst 50 ms 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: Gestrich, 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.

Simon Friederich

Interpreting Heisenberg interpreting quantum states

Abstract

The paper investigates possible readings of the later Heisenberg's remarks on the nature of quantum states. It discusses, in particular, whether Heisenberg should be seen as a proponent of the epistemic conception of states – the view that quantum states are not descriptions of quantum systems but rather reflect the state assigning observers' epistemic relations to these systems. On the one hand, it seems plausible that Heisenberg subscribes to that view, given how he defends the notorious "collapse of the wave function" by relating it to a sudden change in the epistemic situation of the observer registering a measured result. On the other hand, his remarks on quantum probabilities as "potentia" or "objective tendencies" are difficult to reconcile with such a reading. The accounts that are attributed to Heisenberg by the different possible readings considered are subjected to closer scrutiny; at the same time, their respective virtues and problems are discussed.

Zusammenfassung

Diese Arbeit untersucht mögliche Lesarten von Heisenbergs späten Bemerkungen über die Natur von Quantenzuständen. Insbesondere wird die Frage diskutiert, ob Heisenberg als Vertreter der epistemischen Auffassung von Quantenzuständen gelten kann – der Idee, dass Quantenzustände nicht Beschreibungen der objektiven Eigenschaften von Quantensystemen sind sondern die epistemischen Beziehungen von Beobachtern zu diesen Systemen widerspiegeln. Einerseits erscheint es plausibel, dass Heisenberg dieser Sichtweise zustimmt, wenn man sich ansieht, wie er den berüchtigten „Kollaps der Wellenfunktion" verteidigt, indem er ihn mit einer plötzlichen Änderung in der Kenntnis des ein Messergebnis registrierenden Beobachters in Zusammenhang bringt. Andererseits sind seine Bemerkungen über quantenmechanische Wahrscheinlichkeiten als „Potentia" oder „objektive Tendenzen" mit einer solchen Lesart nur schwer in Einklang zu bringen. Die Positionen, die Heisenberg den verschiedenen möglichen Lesarten zufolge vertritt, werden eingehenden Untersuchungen unterzogen; gleichzeitig werden ihre jeweiligen Vorzüge und Probleme diskutiert.

philosophia naturalis 50/2013/1

1. Heisenberg and the epistemic conception of quantum states

The epistemic conception of quantum states is the idea that quantum states are not descriptions of the properties of quantum systems but rather reflect the state assigning observers' epistemic relations to these systems. Although this view has its early roots in the works of Copenhagen adherents Peierls and, maybe (as discussed in this paper), Heisenberg, it has attracted an increasing amount of attention only in recent years. What makes it attractive is that it promises to provide an elegant dissolution of the notorious paradoxes of measurement and quantum non-locality without relying on any technical notions such as hidden variables, branching worlds or spontaneous collapse that are designed just for the purpose of interpretation. Today, there are several different competing attempts of spelling out the epistemic conception of states in detail, its most discussed version probably being Quantum Bayesianism, as developed by Fuchs, Caves, and Schack.[1]

While there has been a considerable amount of work to develop a version of the epistemic account of quantum states that is satisfying from a conceptual point of view, there has been no, or only very little, effort to trace back the historical origins of this idea. The present paper is an attempt to partially fill in this gap by considering the question of whether in the writings of the later Heisenberg one can find an account of quantum states which is an early – perhaps the first – clear-cut exposition of the epistemic conception of states. The question may sound as if it were very easy to answer, but, as the present paper will show, this is not the case. Furthermore, there is clearly no consensus on this matter among those who have proposed interpretations of Heisenberg's philosophical writings: Some take it for granted that Heisenberg is an adherent of the epistemic conception of quantum states, whereas others do not even consider it as a possibility. Marchildon, for instance, when he claims that the epistemic conception of states "goes back at least to Heisenberg" (Marchildon, 2004, 1454), evidently presupposes that Heisenberg is a proponent of it, whereas, for instance, Shimony, ascribes to Heisenberg the "profound and radical" thesis "that the state of a physical object is a collection of potentialities" (Shimony, 1983, 214f.) – a view which, as will be shown, is in strong tension with the epistemic conception of states, to say the least.

While this lack of unanimity may be surprising at first sight, it becomes understandable if one takes a closer look at the passages in Heisenberg's later writings that are relevant for the understanding of his view of quantum states. As will be discussed in later sections of this paper, the main conclusions one is likely to draw as regards his account of quantum states strongly depend on which of his remarks on the subject matter one regards as central. It is therefore small wonder that interpreters focusing on different aspects of his remarks on quantum states have ascribed to him accounts of the nature of quantum states that are completely at odds with each other.

The present paper aims at a clarification of these issues by investigating more closely the different readings of Heisenberg's remarks on quantum states that seem possible and coherent. In order to clarify the meaning of the epistemic conception of quantum states, Section 2 contains a brief exposition and motivation of that view. Section 3 focuses on those of Heisenberg's remarks on quantum states which seem to suggest that, indeed, he is a proponent of the epistemic conception of states. Section 4, in contrast, considers remarks which are difficult to reconcile with such a reading and may prompt one to look for alternatives. Examples of alternative readings are given, and their respective problems and advantages are assessed from both an exegetical and a systematic perspective. Section 5 reconsiders the possibility of interpreting Heisenberg as defending a version of the epistemic conception of states, this time by taking into account the challenge posed by the passages that seem to render such a reading difficult. In Section 6, an entirely different perspective on Heisenberg's remarks on quantum states is presented, based on considerations about the more general aims of philosophical writings proposed by Mara Beller. According to that perspective, Heisenberg is an – as regards philosophical matters – *opportunistic* thinker, whose remarks on the interpretation of quantum mechanics are mainly motivated by dialectical purposes and do not lead to a coherent foundational perspective. The paper closes with a brief concluding remark in Section 7 without ultimately deciding which of the proposed readings of Heisenberg's remarks on quantum states is correct. Its more modest main aim is to spell out the interpretive options and to improve our understanding of the systematic issues involved.

2. The epistemic conception of quantum states and the measurement problem

The epistemic conception of states, as already remarked, is the idea that quantum states are not descriptions of the properties of quantum systems but rather reflect the assigning observers' epistemic relations to these systems. What is involved in adopting this view has been clarified by its proponent Rudolf Peierls, a PhD student of Heisenberg in the late 1920s and an important figure of 20th century physics in his own right. The quantum state, according to Peierls, "represents our *knowledge* of the system we are trying to describe", and an important consequence of this, as he argues, is that the states assigned by different observers "may differ as the nature and amount of knowledge may differ" (Peierls, 1991, 19). This latter aspect provides us with a very useful criterion to decide which views can count as varieties of the epistemic conception of states and which not: If quantum states are supposed to reflect the observer's epistemic relations to the quantum systems states are assigned to, then, since different observers may in general know different things about the values of observables of a given system, the states assigned by different observers may legitimately be different, depending on the different observers' differing epistemic conditions.

The epistemic conception of quantum states is incompatible with the idea that there is such a thing as an agent-independent "true" quantum state of a quantum system – a quantum state it "is in" –, for if such a state existed, one would have to assign this state to the system in order to assign correctly, and assigning any other state would be wrong. If indeed the quantum states assigned by different observers having different knowledge of a quantum system are supposed to reflect their epistemic relations to the system, these states may evidently be different. This contrasts with the situation in classical mechanics, where it seems natural to regard the points in phase space as the *ontic* states classical systems *are in*. In that context, only that point in phase space correctly describes a classical system which is accurately representing its position and momentum. The status of quantum states, on the epistemic conception of quantum states, is closer to that of probability densities defined over phase space than to points in phase space themselves. Probability densities play a fundamental role in classical statistical mechanics, where which probability density one assigns to some system depends

on what information about the values of its macroscopic variables one takes into account. The main difference between probability densities in classical statistical mechanics and quantum states as interpreted by the epistemic conception of states is that there exist extremal probability densities expressing complete information about the ontic state of the system (in form of a Dirac-Delta function over phase space) in the classical case, whereas no quantum state is assumed to correspond to a quantum system's ontic state (inasmuch as such a state is assumed at all[2]) in an analogous sense. To sum up, only accounts that reject the notion of an observer-independent *true* quantum state of a quantum system – a quantum state it *is in* – can be versions of the epistemic conception of quantum states.

The main motivation for taking the epistemic conception of states seriously is that it neatly dissolves the notorious paradoxes of measurement and non-locality. Heisenberg is naturally read as addressing the measurement problem in his remarks on measurement in quantum mechanics by endorsing its dissolution based on the epistemic conception of states, but he does not comment on issues connected to quantum non-locality. I shall therefore focus on the measurement problem rather than on quantum non-locality in what follows.[3]

The measurement problem arises from the fact that, if one thinks of quantum states as states quantum systems "are in" and assumes that the time-evolution of these states always follows the Schrödinger equation, measurements rarely have outcomes.[4] In practice, physicists avoid this difficulty by resorting to von Neumann's notorious projection postulate – the "collapse of the wave function" – which states that the density matrix corresponding to the state of the system after measurement is obtained by projecting the pre-measurement state onto the subspace of eigenstates to the observable measured with eigenvalue corresponding to the measured value of that observable.

The projection postulate, however, is strongly disliked by many researchers in the foundations of quantum mechanics, and many of the most prominent interpretations of the theory attempt to present a coherent picture of the world as described by quantum mechanics that avoids it. According to Laura Ruetsche, for instance, measurement collapse is "a Humean miracle, a violation of the law of nature expressed by the Schrödinger equation" (Ruetsche, 2002, 209), and many others would no doubt agree. Such criticism of the projection postulate, how-

ever, is compelling and natural only if one thinks of the state and its time-evolution as descriptions of what actually happens to the quantum system to which the state is assigned. The intuitive motivation for the projection postulate is very different and relates to the need of readjusting the state after measurement in order to make it compatible with what one knows of the values of observables of the system after having registered the measured result. Measurement collapse looks completely natural if one regards it as reflecting a sudden change in the state-assigning observer's epistemic situation, not as a sudden change in the properties of the measured system itself. Consequently, by adopting a version of the epistemic conception of states, saying that the state reflects the assigning agent's epistemic relation to the system the state is assigned to, measurement collapse can be justified in an elegant way, and the measurement problem is avoided.

A natural question in this context is *in which sense* quantum states might be said to reflect the observers' epistemic relations to the quantum systems. There are two fundamentally different types of possible answers to this question. According to the first type of answer, quantum states encode information about underlying "ontic" states, which are conceived as configurations of fundamental parameters that might as well be called "hidden variables". Rob Spekkens advocates this view and motivates it in an intriguing way by showing that many of the most characteristic features of quantum theories can be reproduced in a toy theory that is based on the epistemic conception of states, supplemented with hidden variables in that sense (cf. Spekkens (2007)). Heisenberg, however, strongly criticizes hidden variables interpretations of different sorts and repeatedly argues for the completeness and definiteness of quantum mechanics[5], so his position is certainly not an epistemic account of states in terms of hidden variables. To obtain a more plausible candidate for Heisenberg's account of quantum states, the question of in which sense quantum states might be said to reflect the observers' epistemic relations to quantum systems must be answered in a different way, namely, by saying that quantum states reflect the observers' expectations as to what the results of future measurements might be.[6] Accounts that rely in one way or another on this strategy of spelling out the epistemic conception of states are based on the hope that the notorious paradoxes of measurement and non-locality can be evaded without specifying underlying any "ontic" states at all, simply by adopting a

certain perspective on the nature of quantum states. In the next section, we will consider whether this perspective is endorsed in Heisenberg's later writings on the interpretation of quantum mechanics.

3. Heisenberg and the epistemic conception of quantum states: a straightforward case?

As already announced in the introductory section, the question of whether Heisenberg should be read as defending the epistemic conception of quantum states is not easy to answer. In analogy to Bohr's remarks on the interpretation of quantum mechanics, Heisenberg's philosophical writings on quantum mechanics have invited strikingly different interpretations from different interpreters. At least in part, this is due to the fact that his texts on the interpretation of quantum mechanics are directed not only at physicists (or philosophers having a background in physics), but also at a wider audience of interested laymen. His remarks are intended not only to present an interpretation of quantum mechanics for those who are knowledgeable about the underlying mathematical formalism, but also to provide an overview of elementary features of that formalism for those who know nothing about it at all. Trying to express himself in a manner that is comprehensible also to non-physicists, Heisenberg presents his thoughts in a non-technical and sometimes ambiguous way. As will become clear in what follows, his remarks on the nature of quantum states are a case in point in that they allow for a variety of different, mutually incompatible, readings. The discussion will be based on a rather small number of passages I shall quote, considering various possible readings of these, even though there are certainly many more passages in his later writings that might be relevant with respect to the points at issue. The main reason for proceeding in this way is that the other passages seem to be at least as susceptible to different, mutually incompatible readings as those that I shall discuss. It appears to me more useful to focus thoroughly on a rather narrow class of statements which are directly relevant to the present topic than on a larger number of diverse remarks which allow for at least as many different interpretations as well.

The main reason for ascribing to Heisenberg an epistemic account of quantum states is that, when he writes that "[s]ince through the

observation our knowledge of the system has changed discontinuously, its mathematical representation also has undergone the discontinuous change and we speak of a 'quantum jump'" (Heisenberg, 1958, 28), he relates measurement collapse to a sudden change in the epistemic situation of the observer. Since such an account of measurement collapse seems to presuppose the epistemic conception of states, the quote just given provides evidence that Heisenberg subscribes to that view. The way how he introduces the notion of a quantum state ("probability function") to his readers can be taken to confirm such a reading:[7]

> The probability function represents a mixture of two different things, partly a fact and partly our knowledge of that fact. It represents a fact insofar as it assigns at the initial time the probability unity (i.e. complete certainty) to the initial situation: the electron moving with the observed velocity at the observed position; "observed" means observed with the accuracy of the experiment. It represents our knowledge insofar as another observer could perhaps know the position of the electron more accurately. The error in the experiment does – at least to some extent – not represent a property of the electron but a deficiency in our knowledge of the electron. Also this deficiency of knowledge is expressed in the probability function. (Heisenberg, 1958, 19)

Quantum states, according to this statement, depend on the accuracy of the state assigning observers' knowledge of the systems states are assigned to, and in that sense they have an important *subjective* element. At the same time, they also have an *objective* element – "a fact", as Heisenberg writes – that manifests itself in the assignment of probability one to the values of observables that one knows after having measured them. The fact that Heisenberg acknowledges such an *objective* element of quantum states does not by itself pose any problem for the reading of him as endorsing a version of the epistemic conception of states. A natural way of interpreting his claim that the quantum state "represents a fact insofar as it assigns at the initial time the probability unity (i.e. complete certainty) to the initial situation" is to read it as claiming that the values of observables are objective facts about quantum system and that an observer's assignment of a quantum state to a quantum system must be compatible with knowledge about such facts in that the state assigned must ascribe probability one to the values of observables that the observer knows to obtain. If, for instance, an observer knows that the z-component of the spin of a certain electron has the value 1/2, the

state he assigns to the system must be an eigenstate of the operator S_z with eigenvalue 1/2. That this state assignment must be in accordance with the observer's knowledge of the value of the spin observable is quite naturally regarded as an "objective element" of the state assigned. This does *not* contradict the epistemic conception of states as long as one does not make the further claim that the state assigned has to be identified with the *true* state of the system – the one it "is in". According to this reading of the passage quoted above, the values of observables are *objective* features of physical reality, whereas the probabilities ascribed to possible measurement outcomes have at least some *subjective* aspect in that the states from which they are derived are different for observers having different epistemic relations to the system.[8]

Heisenberg claims explicitly that different observers may legitimately assign different states to the same system, namely, with respect to cases where an observer has knowledge of the values of observables of a system (an electron, in his example) which is objectively inferior to the knowledge of another observer who "know[s] the position of the electron more accurately." In the type of situation considered here, an observer who is less well informed about the value of an observable of the system than another may assign a different state than the latter without necessarily making any sort of mistake. Heisenberg does not seem to consider the state assignment of the observer who is less well informed as somehow "wrong" or "incorrect", even though the observer who knows more about the values of observables is certainly more likely to make successful predictions about the results of future measurements of the system.

Let us now consider a slightly different type of situation where different observers have different knowledge of the values of observables of a quantum system in such a way that none of them knows strictly *more* about the values of observables than the others. What can we say about Heisenberg's position with regard to this more general case? Assuming that he holds that two observers may legitimately assign two different states if one of them knows strictly less about the values of its observables than the other, we can safely conclude that he would hold the same with respect to the type of situation where none of them knows strictly less (or more) than the other. Together with the observation that, in the passage quoted above, he does not speak about the "true" quantum state of a system (or its "true" probability function) these considerations can

be taken to suggest that Heisenberg does indeed endorse a version of the epistemic conception of quantum states.

In the following section, I shall turn to other passages in his writings that bear on the interpretation of quantum states, which, unlike the ones considered so far, pose considerable difficulties for interpreting him as endorsing an epistemic account of quantum states.

4. Probabilities as objective tendencies: the crux with the "subjective element"

The most important difficulty for reading Heisenberg as defending a version of the epistemic conception of states is that he holds that quantum states contain "statements about possibilities or better tendencies …, and these statements are objective". Here is a context in which he advances this claim:[9]

> The probability function combines objective and subjective elements. It contains statements about possibilities or better tendencies ("potentia" in Aristotelian philosophy), and these statements are completely objective, they do not depend on any observer; and it contains statements about our knowledge of the system, which of course are subjective in so far as they may be different for different observers. In ideal cases the subjective element in the probability function may be practically negligible as compared with the objective one. The physicists then speak of a "pure case". (Heisenberg, 1958, 27)

By "statements about possibilities or better tendencies" that are contained in quantum states Heisenberg seems to mean statements about the probabilities of the values of observables that are derived from quantum states via the Born rule. When he argues that these statements are "completely objective" and that they "do not dependent on any observer", he seems to endorse an account of quantum probabilities as objective probabilities that is very much in spirit with what is nowadays called the "propensity interpretation" of probability.[10] Accounts of probability that are based on this interpretation conceive of probability as measuring an object's objective tendency – or *disposition* – to display a certain property. Propensity views of probability conceive of probabilities as *objective* in a very strong sense in that they conceive them as (objective) properties of the objects (or events) to which they are ascri-

bed. At the opposite spectrum of accounts of probability are *subjectivist* ("Bayesian") accounts. These interpret probabilities as degrees of belief of the agents assigning probabilities, for which questions of truth or correctness as regards probability assignments do not make sense. Whereas the propensity view regards probabilities as agent-independent features of the things themselves, probabilities as degrees of belief may differ from agent to agent, without any of them being in error or making any mistake. In order to be compatible with the epistemic conception of quantum states, accounts of quantum probabilities must make at least some small concession to a subjectivist interpretation of quantum probabilities as it has to allow that quantum probabilities may differ for different observers.

Heisenberg's claims on quantum probabilities as "objective tendencies" have indeed been interpreted as evidence that he subscribes to a propensity view of quantum probabilities. Henry P. Stapp, for instance, claims that "[Heisenberg] propos[es], in effect, that certain probabilities defined by the theory be interpreted as the 'objective tendencies', or propensities, for corresponding *actual events* to occur" (Stapp, 2004, 42). According to Stapp, the idea of interpreting quantum probabilities as propensities is a metaphysically significant step in that it "carries quantum theory far beyond the ontologically neutral stance of the strictly orthodox interpretation" (Stapp, 2004, 42). In a similar vein, Abner Shimony attributes to Heisenberg a "profound and radical thesis", namely, "that the state of a physical object is a collection of potentialities" (Shimony, 1983, 214f.). According to these interpretations of Heisenberg's remarks on quantum probabilities his views are a "foray into metaphysics" (Camilleri, 2009, 153), as Kristian Camilleri notes, which goes far beyond the mere attempt of clarifying the notions of state and probability as they are used in quantum mechanics.

However, an immediate objection to reading Heisenberg as defending a propensity interpretation of quantum probabilities is that it seems to be incompatible with his claim that measurement collapse reflects the fact that "our knowledge of the system has changed discontinuously." This account of measurement collapse, as already discussed, presupposes that the state subjected to collapse reflects the epistemic situation of the observer and is therefore at odds with the idea that the state describes the system's propensities of displaying certain values of observables. As

we have seen in Section 2, the main motivation for adopting the epistemic conception of states is that it permits to interpret measurement collapse in an organic way as reflecting a sudden change in an observer's knowledge of the system.

Those who read Heisenberg as endorsing a propensity view of quantum probabilities may respond to this challenge by pointing to the fact that he classifies as objective only those probabilities assigned to quantum systems in what he calls "pure cases". It is only in these "pure cases" that the "subjective element" of quantum states discussed in the previous section, as he claims, becomes "practically negligible". A "pure case", as he uses this expression, is present whenever a system can be considered in complete isolation from its environment, at least to a high degree of precision, so that it need not be regarded as part of a larger, composite system, together with others. In this situation, as Heisenberg claims, it can be described by a *pure* quantum state (a *Hilbert space vector*[11]), so that we may think of "pure cases" as those where an observer assigns a pure state to a system whose interactions with the environment are negligibly small. Since, according to Heisenberg, the "subjective element" of quantum states is absent in cases that meet this condition, it follows that he regards this element as restricted to situations where a system is not in isolation from its environment, which means that it cannot be well described by means of a pure quantum state.[12]

The idea that pure states, but not mixed states, correspond to the objective properties of quantum systems is familiar from the well-known "ignorance interpretation of mixed states", and it is tempting to read the passage quoted at the beginning of this section as endorsing that view. The ignorance interpretation of mixed states is the view that each quantum state is described by exactly one *pure* quantum state, and that mixed states should be assigned only by those observers who do not know the exact pure state of the system they are assigning a quantum state to. Thus, on the ignorance interpretation of mixed states a mixed state reflects the assigning observer's ignorance of the actual pure quantum state the system is in. These pure states, in contrast to the mixed states, describe the objective probabilities of quantum systems to display certain properties at a given time. In view of the fact that Heisenberg characterises quantum probabilities as "objective tendencies", arguing that the "subjective element" of quantum states is absent in "pure cases", one might think that the position he defends is a ver-

sion of the ignorance interpretation of mixed states. However, as will become clear in a minute, this can hardly be the case.

To see why, we have to recall the distinction between "properly" and "improperly" mixed states, originally introduced by d'Espagnat (1976). The first of these two, the "proper mixtures", are mixed states that are assigned to quantum systems in cases where the state preparation procedure does not narrow down possible states to assign to a unique pure state. It is only in these cases that the ignorance interpretation of mixed state makes sense in that it claims that the assignment of a mixed state reflects the observer's ignorance about the pure state the system really is in. In contrast to mixed states of this type, there are also the so-called "improper mixtures" to which the ignorance interpretation of mixed states cannot be applied. "Improper mixtures" are mixed states which one obtains for systems that are subsystems of larger many-component systems by performing the trace operation over the degrees of freedom of the other subsystems. If we denote by ρ the density matrix assigned to the combined (many-component) system and label the subsystems by the index i, the reduced state of the subsystem j is given by $\rho_j = \mathrm{Tr}_{i\neq j}(\rho)$, where the trace $\mathrm{Tr}_{i\neq j}$ is taken over the degrees of freedom of the other subsystems $i\neq j$. In the generic case, the state ρ assigned to the combined system cannot be written as a (tensor) product state of the reduced density matrices ρ_i of its subsystems. This is equivalent to saying that the state ρ will normally be an *entangled* state and that the reduced states ρ_i will have the form of mixed states, which means that for them the inequality $\mathrm{Tr}(\rho_i^2)<1$ holds. However, mixed states that have been obtained in this way as reduced density matrices via performing the trace cannot be given an ignorance interpretation in the sense of reflecting ignorance about the "true" pure states the subsystems i are in. The reason for this is that the state ρ assigned to the combined system in general does not have the form of a mixture of states which are products of pure states for the subsystems i in such a way that the reduced density matrices ρ_i are mixtures of these pure states with the coefficients used in the decomposition of the state ρ of the combined system. Consequently, the reduced density matrices ρ_i cannot be given an interpretation as reflecting ignorance about any pure states the subsystems i are in, so the ignorance interpretation of mixed states applies only to proper, not to improper, mixtures.

The ignorance interpretation of mixed states that conceives of the assignment of mixed states as reflecting ignorance about the true pure

states quantum systems are in, as we see, applies only to "properly" mixed states, not to "improper mixtures". Heisenberg's remarks on mixed states as expressing some sort of "ignorance", however, do not refer to systems where the state preparation procedure does not narrow down possible states to assign to a unique pure state, but to systems that are not in isolation from their environment and must therefore be considered as subsystems of a larger system. This concerns, in particular, quantum systems that are measured by means of a measuring apparatus with which they interact. Concerning these systems Heisenberg claims that "it is very important to realize that our object [being measured] has to be in contact with the other part of the world, namely, the experimental arrangement, the measuring rod, etc., before or at least in the moment of observation" (Heisenberg, 1958, 27; cf. 1955, 27). According to the considerations presented before, the only state one can assign to the system being measured when it is coupled in this way to the measuring apparatus is a mixed stated given as a reduced density matrix. Heisenberg explicitly ascribes a "subjective element of incomplete knowledge" to this state when he writes that "[a]fter this [measurement] interaction has taken place, the probability function contains the objective element of tendency and the subjective element of incomplete knowledge, even if it has been a 'pure case' before" (Heisenberg 1958, 28).[13] This "incomplete knowledge", however, as we have seen, cannot be incomplete knowledge about the true pure state the measured system is in, so, we might ask, what kind of incomplete knowledge can it be? Heisenberg's answer to this question is that it is incomplete knowledge about the microscopic details of the rest of the world to which the measured system, via the measurement device, must be coupled. The "subjective element" which, according to himself, is present in (improperly) mixed states relates to an ignorance on behalf of the observer in that it reflects her "uncertainties of the microscopic structure of the whole world" (Heisenberg 1958, 27). This statement makes it clear that Heisenberg is not endorsing the conventional ignorance interpretation of *properly* mixed states in terms of ignorance about underlying pure states, but another, more idiosyncratic, ignorance interpretation of *improperly* mixed states, holding that these reflect ignorance of an altogether different sort, namely, ignorance of the microscopic details of the rest of the world. Is this a reasonable type of an "ignorance interpretation" of mixed states?

There are several aspects to this question. First, we should note that Heisenberg has indeed good reasons not to endorse the conventional ignorance interpretation of mixed states as reflecting incomplete knowledge of the true pure state the system is in, for this interpretation, as is widely known, runs straight into the measurement problem. The conventional ignorance interpretation of mixed states *would* have a chance of solving the measurement problem only in case the post-measurement state of the measured system – or that of the measured system together with the measuring apparatus – would be a mixed state in form of a proper mixture of pure states associated one-to-one to the possible outcomes of measurement. However, this is not the case, for the state of the measured system (or that of the measured system together with the apparatus, if the environment is included) is an improper mixture for which, as we have seen, no ignorance interpretation in terms of underlying pure states can be given. Consequently, Heisenberg's strategy to characterise improper, rather than proper, mixtures as reflecting incomplete knowledge is in a sense well-motivated.

The problem with his supposed strategy to spell out ignorance in terms of "uncertainties of the microscopic structure of the whole world", however, is that it does not allow for an elegant dissolution of the measurement problem either. If the states of the systems being measured always *were* proper mixtures, the interpretation of mixed states in terms of ignorance about underlying pure states could justify measurement collapse by relating it to a change in knowledge about the underlying pure state. As we have seen, however, the state of the measured system[14], taken by itself, must be an improper, rather than a proper, mixture, so that it cannot be interpreted as reflecting ignorance of that sort. Heisenberg's idea that mixed states may reflect ignorance of an altogether different sort, namely, ignorance about "uncertainties of the microscopic structure of the whole world" does not help us solving the measurement problem in an analogous fashion. A measurement normally does not provide us with new information about the "microscopic structure" of the rest of the world, and even if it did, it remains unclear why this information should necessitate a sudden change of state for the system being measured in exactly the way prescribed by von Neumann's projection postulate. Heisenberg's defence of measurement collapse was that "through the observation our knowledge of the system has changed discontinuously." However, it is dubious why a change

in our "uncertainties of the microscopic structure" of the rest of the world should affect the state of the system being measured and force it to undergo a "discontinuous change".

As we have seen before, the problem with ascribing to Heisenberg a propensity view of quantum probabilities is that such a view threatens to be incompatible with his defence of measurement collapse as reflecting a sudden change in the knowledge of the observer. To avoid this problem, we considered Heisenberg's claim that mixed quantum states – *improper* mixtures, as I have argued – have an epistemic aspect in that they reflect ignorance in form of "uncertainties of the microscopic structure of the world." The result we have obtained is discouraging, for it remains completely unclear in which sense uncertainties of that sort should be relevant for a sudden change of the state of a quantum system being measured. The most optimistic conclusion that can possibly be drawn at this stage is that the problem of interpretation we encounter here is just due to the fact that Heisenberg does not spell out his ideas in greater detail. However, we should also consider alternative readings.

One possible reaction to the problem of combining Heisenberg's remarks on quantum probabilities as objective tendencies with his account of measurement collapse is to simply ignore the remarks on collapse, for instance by interpreting them as mere rhetorical moves to make the apparent abruptness and unruliness of collapse seem more acceptable.[15] An interpretive reading along these lines is developed by Henry P. Stapp, who claims that according to Heisenberg collapse is a physical process that from time to time interrupts unitary time-evolution as governed by the Schrödinger equation in an irregular manner. According to this supposed view of Heisenberg, "[t]he fundamental dynamical process of nature is no longer one single uniform process, as it is in classical physics. It consists rather of two different processes" (Stapp, 2004, 41). The smooth, unitary time-evolution and the discontinuous state-change of collapse are both regarded as physical processes, and the features of reality that correspond to the sudden state-changes of collapse are certain randomly occurring "actual events". Quantum mechanical probabilities are objective in this view in that they measure observer-independent tendencies of the corresponding actual events to occur. The picture of reality that emerges from this interpretation exhibits certain similarities to spontaneous localisation theories such as the model developed by Ghirardi, Rimini, and Weber ("GRW"), which

modifies standard quantum mechanics by complementing time-evolution according to the Schrödinger equation with randomly interspersed spontaneous localisation processes. In comparison to GRW-theory, Heisenberg's position as construed by Stapp has the disadvantage that it does not give a quantitative account of the frequency and dynamics of how these processes of spontaneous localisation occur.

Stapp's interpretation of Heisenberg's remarks on quantum probabilities, as we see, is unsatisfying for systematic reasons, but, more importantly, it is unconvincing as a reading of Heisenberg on exegetical grounds. It identifies the transition from the realm of "potentiality" to that of "actuality" in form of an "actual event" with the discontinuous change of state in measurement collapse, whereas Heisenberg categorically claims that "the transition from the 'possible' to the 'actual'" (Heisenberg, 1958, 28; cf. 1955, 28) has nothing to do with that change which results from the "discontinuous change of our knowledge in the instant of registration" (Heisenberg, 1958, 29). According to Heisenberg, the transition from the possible to the actual coincides not with measurement collapse, but rather with the "physical act of observation", namely, the "interaction of the object with the measuring device" (Heisenberg, 1958, 28f.), which, as he suggests, must be sharply distinguished from collapse. Stapp's account, aside from ignoring Heisenberg's claim on collapse as reflecting a sudden change of knowledge of the observer, thus fails for both systematic and exegetical reasons. In the next section, I consider a completely different reading of Heisenberg's remarks of quantum states that takes his account of the "the transition from the 'possible' to the 'actual'" just quoted as its starting point and reconsiders the idea that he should be read as endorsing a version of the epistemic conception of states.

5. The epistemic conception reconsidered

Kristian Camilleri has recently proposed a novel interpretation of Heisenberg's later writings according to which the dichotomy between the "actual" and the "possible" should be conceived as a contrast between two different *forms of description* rather than a contrast between two different metaphysical categories. According to this reading, "the transition from the 'possible' to the 'actual'" should not be construed as a

physical process occurring in nature, but rather as a change in the language we use at different stages in our description of the measurement process. In particular, as Camilleri contends, this transition has nothing to do with measurement collapse:

> The 'transition from the potential to the actual' is therefore completely misunderstood if it is interpreted as a collapse of a physically real extended wave-packet in space. Nor should it be interpreted in Berkeley's sense: *esse est percipi*. Rather, we must understand the 'actual' and the 'possible' as two modes of description, both of which employ the language of time and space at some level. The transition from potentiality (a quantum-mechanical description) to actuality (a classical space-time description) must be understood as a transition from one *mode of description* to another. The two modes of description – the possible and the actual – are deemed complementary in Heisenberg's, though not in Bohr's, sense of the term. (Camilleri, 2009, 170)

Camilleri backs up his interpretation by means of a large number of passages in Heisenberg's later writings on the topic of language which he reads as endorsing an account which, according to Camilleri, "could be termed a 'quasi-transcendental conception of language'" (Camilleri, 2009, 152).[16] Without going into the details of Camilleri's reading of Heisenberg on the role of language in quantum mechanics, I take it as a starting point for an interpretation of Heisenberg's remarks on quantum states as containing statements about "objective tendencies" that avoids the propensity interpretation of probabilities and fits well with the epistemic conception of states.

Let us start from Camilleri's reading of Heisenberg's remarks that interprets the distinction between actuality and potentiality corresponds as a distinction between two modes of description, namely, a "classical space-time description" and a "quantum-mechanical description". Does this distinction go well with the epistemic conception of states that regards quantum states as reflecting the state assigning agents' epistemic conditions? I shall argue that it does, at least if the epistemic conception of states is spelled out in a certain way, namely, as saying that quantum states reflect the observers' knowledge of the values of observables.[17] If the epistemic conception of states is spelled out in that way, one might indeed somewhat schematically describe it as recognising fundamentally different "actuality" and "potentiality" modes of description in quantum mechanical practice.

The "actuality mode" is exemplified by sentences which report an

observer's knowledge of the values of observables such as in the form "The value of the observable A of the system s lies in the set Δ." Sentences of that form are referred to as "non-quantum magnitude claims" (NQMCs) by Richard Healey , and I adopt this terminology in what follows. In accordance with Camilleri's characterisation of the "actuality mode of description" NQMCs can be regarded as "classical space-time descriptions" in that they describe, up to a certain accuracy, the properties of a physical system by means of the vocabulary of classical physics, without employing any probabilistic notions.[19] The "potentiality mode of description", on the other hand, is given by the assignment of quantum states to quantum systems, together with the derivation of probabilities for the values of observables from the states via the Born Rule. This is certainly not an "actuality mode of description" from the perspective of the epistemic conception of states, which denies that assignments of quantum states do ever report any facts about quantum systems. From the perspective of the epistemic conception of states, quantum states are not really *descriptions* of quantum systems at all, at least not in a more narrow sense of "description", where two descriptions of the same object cannot both be correct if they are mutually incompatible the sense in which assignments of different probabilities to one and the same value of a certain observable are. However, one may want to call the assignment of quantum states to quantum systems a "mode of description" in a wider sense of the word "description", and in that case one will certainly have to call it a "potentiality mode" rather than an "actuality mode" in that it uses probabilistic notions and, as already remarked, does not describe any "actual" facts pertaining to the objects.

From the perspective of such an epistemic account of states, the transition from "potentiality" to "actuality" amounts to a switch in ones "mode of description" of a system treated quantum mechanically. Assume that you have assigned a quantum state to a system ("potentiality mode"), thereby making some predictions as to what, for instance, the results of future measurements will be, and that, subsequently, on the basis of incoming measurement data, you are able to formulate an "NQMC", that is, a statement about the values of observables measured ("actuality mode"). The new information of the values of observables obtained makes it necessary to adjust the state assigned to the system by means of the discontinuous change that is measurement collapse, but

this change, in contrast to the one described before, occurs on the level of time-evolution of quantum states alone. In other words, it appears completely within the sphere of the "potentiality mode of description", in agreement with Heisenberg's claim that collapse has nothing to do with the transition from potentiality to actuality but merely signals a change in the representation of our knowledge.

Some adherents of the epistemic conception of states may not like to characterise the different uses of language in quantum mechanics in terms of the notions of "potentiality" and "actuality", for these are metaphysically charged notions that may cause serious misunderstanding. Nevertheless, as we have seen, versions of the epistemic conception of states that regard quantum states as reflecting the observer's epistemic situation in the form of reflecting knowledge of the values of observables seem to go well with Camilleri's interpretation of Heisenberg's distinction between the 'actual' and the 'potential' as two contrasting modes of speech. What remains to be shown is that it is compatible with Heisenberg's claim that the probabilities derived in "pure cases" are "completely objective" and that the "subjective element" can be neglected in situations where pure states are assigned. This claim is in tension with the epistemic conception of states according to which different observers having different knowledge of the same quantum system may always assign different states to it. This "subjective element" of observer dependence characterises *all* quantum states, both pure and mixed, according to the epistemic conception of states.

The question of how to address the status of pure states is a very general challenge for the epistemic conception of states that arises not only in connection with the interpretation of Heisenberg's remarks on quantum states. It can be formulated as follows: Whenever an observer has knowledge of the values of observables of a quantum system that narrows down possible states to assign to a unique pure state, this knowledge about the values of observables cannot be improved any further due to the uncertainty principle. Take the case of an observer who measures, say, the z-direction S_z of electron spin by means of a Stern-Gerlach device.[20] Having observed an electron leaving the device on an "up"-trajectory, she knows that the value of S_z is 1/2. Only the assignment of the state $|+_z>$, the eigenstate of S_z to the eigenvalue 1/2, is compatible with that piece of knowledge. If she proceeds to measure another direction of spin, say S_x , her knowledge of the value of S_z will get lost

according to the uncertainty principle. Her knowledge of the values of observables, in this case spin in various directions, may be *changed* due to measurement, but it cannot be *improved* in that for every bit of information she obtains of the value of spin in a certain direction she loses information of the value of spin in another direction. Any assignment of state that is made on the basis of knowledge which is not improvable in that sense can be seen as *objectively privileged* as compared to all other assignments that observers may make on the basis of different knowledge of the values of observables. It is natural to conclude from this that a state assignment which is "privileged" in that sense is an assignment of the *true* quantum state of the system – the one it "is in". This conclusion is incompatible with the epistemic conception of states, and the line of thought just presented therefore provides a general challenge for the epistemic conception of states that is relevant not only for the question of how to interpret Heisenberg's remarks.

Fortunately – from the perspective of the epistemic conception of states – the challenge just formulated can be answered. To see how, one should first note that the epistemic conception of states can allow that a state that is assigned by an observer whose knowledge of the values of observables cannot by improved any further has a special and privileged status. This, however, does not necessarily mean that this state can count as the true state of the system in the sense of a state the system is in, for one need not conclude that this state corresponds to a property which the system has independently of someone being there who happens to assign it. The proponent of the epistemic conception of states may well regard the expression "state assigned by those whose knowledge about the values of observables cannot be further improved" as having a referent only if an agent happens to be there who really has such excellent knowledge. Sometimes there might indeed be agents having knowledge of the values of observables which narrows down possible states to assign to a uniquely determined pure state, but the assumption that even when there are no agents having such knowledge, there exists some state which would have to be assigned by anyone assigning a state to the system is at odds with the principles of the epistemic conception of states itself. Consequently, even if one concedes that from time to time agents may have knowledge of the values of observables which cannot be further improved, this does not mean that one is also committed to the view that for any system there exists some state it is in.

This line of thought is not found in Heisenberg's writings, but we can speculate that he might have accepted it, to combine an epistemic account of quantum states with the view that the probabilities derived from quantum states in what he calls "pure cases" are objective. Heisenberg's habit of calling these probabilities "objective tendencies" is misleading on this reading, for this wording invites an interpretation of him as endorsing a propensity view of quantum probabilities with which the epistemic conception of states is incompatible. From the perspective of the epistemic conception of states, the probabilities derived from pure states are not *objective* simpliciter, but, nevertheless, *objectively privileged* over any other probabilities that different observers might have assigned on the basis of poorer knowledge of the values of observables. This can be seen as a sense in which, in accordance with Heisenberg's claims, the probabilities ascribed to the values of observables in "pure cases", that is, the probabilities derived from pure states, can be regarded as "objective" to a higher degree than others without obtaining any conflict with the epistemic conception of states.

On the one hand, the reading of Heisenberg according to which he holds that the probabilities derived from pure states are "objective" in the (weak) sense of "objectively privileged" just specified may seem not as natural as the reading discussed in the previous section according to which he defends a propensity view of quantum probabilities. On the other hand, it has the virtue of being compatible with the epistemic conception of states and, therefore, with his defence of measurement collapse as deriving from the instantaneous change in the knowledge of the observer when registering the measured result. The two readings developed in this and the previous section are very different, and even if my sympathies are slightly in favour of the reading proposed in this section, I still find it difficult to say which of them is more likely correct. In the following section, the last of this paper, I present yet another reading of Heisenberg's remarks on quantum states, based on the idea that there might just be no coherent account of the nature of quantum states and probabilities in his later writings at all.

6. Heisenberg as a philosophical opportunist?

As we have seen in the previous sections while discussing possible readings of Heisenberg's remarks on the nature of quantum states, the interpretation of these remarks is an intricate task. In this section, I consider as a possible reaction to the interpretive problems we have encountered the idea that finding a coherent account of quantum states in Heisenberg's later writings is so difficult simply because there is no such account to be found at all.

A reading that incorporates this idea might be based on a general approach to the interpretation of Heisenberg's philosophical remarks on quantum mechanics developed by Mara Beller. Beller sees Heisenberg as chiefly concerned with establishing and defending his own reputation as a physicist whose discoveries are of uttermost relevance not only for physics itself, but also for age-old debates outside physics in philosophy. She claims that Heisenberg "wanted to be the new Kant" (Beller, 1999, 195) in that he intended to make a lasting contribution to the debate about causality that goes back to Hume and Kant by arguing "that 'there seems to be the strongest evidence' for the 'final' renunciation of causality and objectivity" (Beller, 1999, 196).[21] Despite his strong claims on the irrevocable "inevitability of acausality" Heisenberg's claims by means of which he intends to support his philosophical conclusions are, according to Beller, "built on shaky circular arguments, on intuitively appealing but incorrect statements, [and] on metaphorical allusions" (Beller, 1999, 196). Beller holds that philosophical ideas such as that of the indispensability of classical concepts are endorsed by Heisenberg only in a "local and opportunistic fashion" (Beller 1999, 199), that is, they are used for rhetorical purposes only, not as well-weighed contributions to substantial philosophical debates. According to Beller, Heisenberg's celebrated use of the positivist "principle of elimination of unobservables" (Beller, 1999, 52) in his seminal 1925 paper "Quantum-theoretical re-interpretation of kinematic and mechanical relations" (Heisenberg, 1925)[22] can be regarded as an early case in point, since, as she concludes from his correspondence during the period when he began writing the paper, "epistemological considerations were far from Heisenberg's mind during his first attempts to tackle the problem" (Beller, 1999, 53 f.).

Without necessarily accepting all of Beller's negative judgements about Heisenberg's contributions to the interpretation of quantum mechanics,

we can take up her idea that the endorsement of philosophical ideas in his philosophical writings may sometimes be "local and opportunistic" and apply them to the reading of his remarks on the nature of quantum states and quantum probabilities, which Beller does not seem to consider explicitly. From the perspective of her account, it seems natural to read Heisenberg's apparent endorsement of the epistemic conception of states in conjunction with his claims about quantum probabilities as "objective tendencies" mainly as a rhetorical move that should not be mistaken for a serious philosophical account. According to this perspective, the fact that he seems to adopt both the epistemic conception of states and the propensity interpretation of quantum probabilities without demonstrating how these conflicting views may be reconciled is seen as an indication that his primary aim is to be able to avail himself of arguments against his critics that are based on either of them. The dialectical advantages he obtains through this move are the following:

An important criticism of quantum mechanics and its standard interpretation, brought forward, for instance, by Schrödinger (1952), concerns the notion of a "quantum jump", which is central to the "orthodox view" defended by Heisenberg. Historically, the notion of a quantum jump emerged as that of an unruly transition between two energy levels, and it survived in von Neumann's axiomatic codification of the theory in form of wave function collapse that occurs whenever the system is measured. The epistemic conception of states provides Heisenberg with a forceful defence of collapse, which we have encountered many times in the present paper: The sudden change of the quantum state in wave function collapse is not seen as a real physical processes but merely reflects a change in the knowledge of the observer assigning the state to the system. Assuming that the state represents features of the observer's epistemic situation, collapse becomes natural and innocuous. Heisenberg can answer Schrödinger's objection by claiming that "[w]hen the old adage 'Natura non facit saltus' is used as a criticism of quantum theory, we can reply that certainly our knowledge can change suddenly and that this fact satisfies the use of the term 'quantum jump'" (Heisenberg, 1958, 28).

Although this reply against critics is certainly very useful for Heisenberg to defend the standard account of the measurement process including collapse, it is not without risks in that by declaring the state subjected to collapse as reflecting subjective information it threatens to invite

the charge of being "subjectivist", "instrumentalist", "operationalist", or whatever pejoratively laden form of what today is called "anti-realism" one might think of. From the perspective of the "opportunistic" reading of Heisenberg presented in this section, he declares adherence to an account of quantum probabilities as "objective tendencies" exactly for the purpose of defending himself against this type of charge. The clue of this move is twofold: On the one hand, by presenting an account of quantum probabilities as objective tendencies Heisenberg can point to an at least seemingly clear-cut and precise sense in which quantum mechanics *is* objective against those accusing him of subjectivism. On the other hand, this move makes it possible for him to start criticising his own critics by suggesting that their objections arises from a naïve, perhaps nostalgic, desire for classical *determinism*[23], which prevents them from recognising the alleged real locus of objectivity in quantum mechanics in the objectivity of quantum mechanical probabilities. By combining an interpretation of quantum probabilities as "objective tendencies" with the epistemic conception of states Heisenberg manages to turn the dialectical situation in the dispute between him and his critics in favour of himself. His opponents may be busy at this stage with rejecting the challenge of naïvely presupposing a deterministic classical world view and therefore may fail to detect the inconsistency (according to this reading) of the package of views which Heisenberg offers. Thus, according to this reading, Heisenberg adopts the epistemic conception of states together with an account of quantum probabilities as objective tendencies mainly for rhetorical purposes, not because he is able to combine them in a consistent and unified view. This is certainly not a flattering reading, but it clearly has some advantages and does not seem easy to reject.

7. Concluding Remark

It seems difficult to decide which of the readings of Heisenberg's remarks on quantum states developed in the previous sections does best justice to his remarks on quantum states. The reading proposed in the last section, inspired by Beller's view of Heisenberg as a philosophical opportunist, goes perhaps best with the formulations he chooses, and an attractive option might be to accept it without necessarily agreeing

with the dismissive overtones of Beller's remarks on Heisenberg as a philosopher of physics. In view of the fact that the most difficult and paradoxical issues in the foundations of quantum mechanics are widely regarded as unresolved as of today, Heisenberg's failure (according to this reading) to deliver a coherent account may be seen as a harmless and easily excusable shortcoming, which pales in comparison to the variety and richness of his ideas on foundational issues. More recent accounts such as those developed by Stapp, Mermin, Fuchs, Healey and others, mentioned in the previous sections, may be seen as modern attempts to combine crucial insights expressed in Heisenberg's writings into more coherent accounts of the foundations of quantum mechanics. My (rather modest) aim in this paper has been to clarify which readings of Heisenberg's remarks on quantum states can be given, which are their most important problems, and which solutions to them are coherent. Arguably, answering these questions is no less a legitimate goal than to determine what Heisenberg's actual view of quantum states really was.

Acknowledgements

I am very grateful to Koray Karaca and Gregor Schiemann for very helpful comments on an earlier version. Furthermore, I would like to thank two anonymous referees of *Philosophia Naturalis* for their useful reports and suggestions on how to improve the paper.

Notes

1 For studies defending or developing versions of the epistemic conception of states and views in a similar spirit, see Fuchs and Peres (2000), Mermin (2003), Caves et al. (2002a; 2002b), Fuchs (2002), Pitowsky (2003), Caves et al. (2007), Spekkens (2007), Fuchs and Schack (2010), Friederich (2011).

2 See the distinction between epistemic accounts of quantum states which are based on hidden variables interpretations and others which are not, discussed at the end of this section.

3 See Section 2 of Friederich (2011) for a discussion of how the measurement problem and the problem of quantum non-locality are dissolved by the epistemic conception of states and Section 3 of Fuchs (2002) for more detailed considerations in favour of the epistemic conception of quantum states related to quantum non-locality, based on Einstein's remarks on entangled states.

4 This is at least what happens if one assumes, as usual, the so-called eigenstate-eigenvalue link which says that for a system in a state ψ an observable A has a definite value a if and only if ψ is an eigenstate of (the operator corresponding to) A with eigenvalue a.
5 For his criticism of alternatives to the "Copenhagen interpretation", in particular of hidden variable approaches, see Chapter VIII, "Criticism and Counterproposals to the Copenhagen Interpretation of quantum theory", of Heisenberg (1958).
6 Chris Fuchs, an important proponent of this approach of spelling out the epistemic conception of states, prefers to talk in a more general way of "experimental interventions into nature" instead of "future measurements", see p. 7 of Fuchs (2002).
7 See Heisenberg (1955, 27) for a less precise precursor of the formulation in the following quote.
8 In some epistemic accounts of quantum states quantum probabilities are argued to be of a more objective than subjective character. Healey's pragmatist interpretation of quantum theory (Healey, 2012) for instance, which is at least very close to the epistemic conception in spirit, conceives of probabilities as objective, and similar comments apply to the epistemic account of states developed in Friederich (2011), where, however, the status of quantum probabilities is not considered explicitly.
9 See, for instance, Heisenberg (1984b, 240; 1958, 145–155) for similar remarks on his "Aristotelian" notion of "potentia". Schiemann (2008, 57) gives an argument of why Heisenberg's notion of "potentia" is quite far from the original concept in Aristotle's works.
10 For an illuminating classification of propensity interpretations of probability which also serves as an introduction to the subject see Gillies (2000).
11 See Heisenberg (1955, 27), for the German original see Heisenberg (1984a, 447).
12 Pure states contrast with mixed states which, in ordinary Hilbert space quantum mechanics, are states that cannot be written in the form of Hilbert space vectors ψ, but only as density matrices ρ. For density matrices ρ corresponding to mixed states one has $Tr(\rho^2) < 1$, for those corresponding to pure states $Tr(\rho^2) = 1$.
13 The same point is made in Heisenberg (1955, 27–28).
14 Referring to "the" state of the measured system I presuppose an ontic, rather than an epistemic account of states. This is justified in the context of the present section, where I focus on readings of Heisenberg as endorsing an ontic account of states.
15 For a reading that adopts this idea and develops it further, see Section 6 of this paper.
16 Heisenberg expresses his views of the overarching importance of language, for instance, when he writes: "Language is, as it were, a net spread between people, a net in which our thoughts and knowledge are inextricably enmeshed" (Heisenberg, 1971, 138; cf. 1958, ch. X: "Language and Reality in Modern Physics").

17 For a version of the epistemic conception of states based on this idea, see Friederich (2011).
18 See Healey (2012, sect. 3).
19 This characterisation remains a little unfortunate, however, because knowledge of the values of non-classical observables such as spin can be formulated in an exactly analogous manner.
20 I focus on the spin degree of freedom, where the associated Hilbert space has only (complex) dimension 2, to make the argument as perspicuous as possible. Completely analogous considerations apply for the orbital degrees of freedom of whatever particle.
21 The references in the quote are to Heisenberg (1934, 17).
22 For an English translation see Heisenberg (1967a).
23 This is evident, for instance, in how Heisenberg deals with alternative interpretations of quantum mechanics, portraying them in a very unfavourable light, see Chapter VIII of Heisenberg (1958).

References

Beller, Mara, 1999: Quantum Dialogue: The Making of a Revolution. Chicago: University of Chicago Press.

Bub, Jeffrey, 2007: Quantum Probabilities as Degrees of Belief. In: Studies in History and Philosophy of Modern Physics 38, pp. 232–254.

Camilleri, Kristian, 2009: Heisenberg and the Interpretation of Quantum Mechanics: The Physicist as a Philosopher. Cambridge: Cambridge University Press.

Caves, Carlton M.; Fuchs, Christopher A.; Schack, Rüdiger, 2002: Quantum Probabilities as Bayesian Probabilities. In: Physical Review A 65, p. 022305.

Caves, Carlton M.; Fuchs, Christopher A.; Schack, Rüdiger, 2007: Subjective Probability and Quantum Certainty. In: Studies in History and Philosophy of Modern Physics 38, pp. 255–274.

d'Espagnat, Bernard, 1976: Conceptual Foundations of Quantum Mechanics. 2nd edition. Reading, Mass.: Addison-Wesley.

Friederich, Simon, 2011: How to spell out the Epistemic Conception of Quantum States. In: Studies in History and Philosophy of Modern Physics 42, pp. 149–157.

Fuchs, Christopher A., 2002: Quantum Mechanics as Quantum Information (and only a little more). In: Khrennikov, A. (ed.): Quantum Theory: Reconsideration of Foundations. Växjö: Växjö University Press, pp. 463–543.

Fuchs, Christopher A.; Peres, Asher, 2000: Quantum Theory needs no 'Interpretation'. In: Physics Today 53, pp. 70–71.
Fuchs, Christopher A.; Schack, Rüdiger, 2010: A Quantum-Bayesian Route to Quantum-State Space. In: Foundations of Physics 41, pp. 345–356.
Gillies, Donald, 2000: Varieties of Propensity. In: British Journal for the Philosophy of Science 51, pp. 807–835.
Healey, Richard, 2012: Quantum Theory: a Pragmatist Approach. In: British Journal for the Philosophy of Science 63, pp. 729–771.
Heisenberg, Werner, 1925: Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. In: Zeitschrift für Physik 33, pp. 841–860. Transl. as Heisenberg (1967a).
Heisenberg, Werner, 1934: Wandlungen in den Grundlagen der exacten Naturwissenschaft in jüngster Zeit. In: Naturwissenschaften 40, pp. 669–675. Transl. as: Recent changes in the foundations of the exact sciences. In: Heisenberg (1979, 11–26).
Heisenberg, Werner, 1955: The Development of the Interpretation of Quantum Theory. In: Pauli, Wolfgang; Rosenfeld, Leon; Weisskopf, Victor (eds.): Niels Bohr and the Development of Physics. Essays dedicated to Niels Bohr on the occasion of his seventieth birthday. New York: McGraw Hill, pp. 12–29.
Heisenberg, Werner, 1958: Physics and Philosophy: The Revolution in Modern Science. London: George Allen & Unwin. Repr. New York: Harper 2007 (page numbers referring to this edition).
Heisenberg, Werner, 1967a: Quantum-Theoretical re-Interpretation of Kinematic and Mechanical Relations. In: van der Waerden, B. L. (ed.): Sources of Quantum Mechanics. Amsterdam: North-Holland, pp. 261–276.
Heisenberg, Werner, 1971: Physics and Beyond. New York: Harper & Row.
Heisenberg, Werner, 1979: Philosophical Problems of Quantum Physics. Woodbridge, Conn.: Ox Bow. Repr. of: Philosophical Problems of Nuclear Science. New York: Pantheon 1952.
Heisenberg, Werner, 1984a: Gesammelte Werke. Series C: Philosophical and Popular Writings. Vol. I: Physik und Erkenntnis 1927–1955 (ed. by Blum, W.; Dürr, H.; Rechenberg, H.). Munich: Piper.
Heisenberg, Werner, 1984b: Gesammelte Werke, Series C: Philosophical and popular writings. Vol. II: Physik und Erkenntnis 1956–1968 (ed. by Blum, W.; Dürr, H.; Rechenberg, H.). Munich: Piper.

Marchildon, Louis, 2004: Why should we interpret Quantum Mechanics? In: Foundations of Physics 34, pp. 1453–1466.
Mermin, N. David, 2003): Copenhagen Computation. In: Studies in History and Philosophy of Modern Physics 34, pp. 511–522.
Peierls, Rudolf, 1991: In defence of 'Measurement'. In: Physics World (January) 1991, pp. 19–20.
Pitowsky, Itamar, 2003: Betting on the Outcomes of Measurements: A Bayesian Theory of Quantum Probability. In: Studies in History and Philosophy of Modern Physics 34, pp. 395–414.
Ruetsche, Laura, 2002: Interpreting Quantum Theories. In: Machamer, P.; Silberstein, M. (eds.): The Blackwell Guide to the Philosophy of Science. Oxford: Blackwell, pp. 199–226.
Schiemann, Gregor, 2008: Werner Heisenberg. München: C. H. Beck.
Schrödinger, Erwin, 1952: Are There Quantum Jumps? Part I. In: The British Journal for the Philosophy of Science 3, pp. 109–123.
Shimony, Abner, 1983: Reflections on the Philosophy of Bohr, Heisenberg and Schrödinger. In: Cohen, R. S.; Laudan, L.; Riedel, D. (eds.): Physics, Philosophy and Psychoanalysis. Dordrecht: Springer, pp. 209–221.
Spekkens, Robert W., 2007: Evidence for the Epistemic View of Quantum States: A Toy Theory. In: Physical Review A 75, p. 032110.
Stapp, Henry P., 2004: Mind, Matter and Quantum Mechanics. 2nd edition. Berlin/Heidelberg/New York: Springer.

Marco Giovanelli

Leibniz-Äquivalenz vs. Einstein-Äquivalenz

Was man von der Logisch-Empiristischen (Fehl-)Interpretation des Punkt-Koinzidenz-Arguments lernen kann

Zusammenfassung

Die Entdeckung, dass Einsteins berühmtes Punkt-Koinzidenz-Argument zur allgemeinen Kovarianz tatsächlich eine Reaktion auf die Lochbetrachtung war, hat in den vergangenen 30 Jahren zu einer intensiven philosophischen Debatte geführt. Auch wenn die philosophischen Konsequenzen äußerst kontrovers gesehen werden, stimmen die Protogonisten doch darin überein, das Argument als Ausdruck von Leibniz-Äquivalenz, mithin als eine moderne Version von Leibniz berühmten Ununterscheidbarkeitsargumenten gegen Newtons absoluten Raum aufzufassen. Ziel des Aufsatzes ist es zu zeigen, dass der Bezug zu Leibniz, wenn auch auf den ersten Blick plausibel, tatsächlich in vielerlei Hinsicht irreführend ist. Insbesondere wird dahingehend argumentiert, dass die Logischen Empiristen ein signifikantes historisches Beispiel für einen Versuch darstellen, das Punkt-Koinzidenz Argument als ein Ununterscheidbarkeitsargument im Sinne von Leibniz, ähnlich denen im 19. Jahrhundert von Helmholtz, Hausdorff und Poincaré vorgebrachten, zu deuten. Dieser Deutung der Allgemeinen Relativitätstheorie gelingt es aber nicht, das eigentlich philosophisch Neue von Einsteins Theorie plausibel zu interpretieren. Wenn Einsteins Punkt-Koinzidenz/Lochargument als ein Ununterscheidbarkeitsargument angesehen werden soll, kann dies kein Argument à la Leibniz sein. Vielmehr hat Einstein ein neuartiges Ununterscheidbarkeitsargument eingeführt, das vielleicht besser als ‚Einstein-Äquivalenz' charakterisiert werden sollte. Durch Aufnahme und Weiterentwicklung einiger Ideen von Weyl wird gezeigt, dass Leibniz' Argumente zwar das Konzept der ‚Symmetrie' in die Wissenschaftsgeschichte eingebracht haben, Einsteins Argument aber etwas antizipiert hat, was heute gewöhnlich ‚Eichfreiheit' genannt wird. Wird im ersten Fall die Ununterscheidbarkeit durch ein zu wenig an mathematischer Struktur erzeugt, so ist sie im zweiten Fall gerade die Folge eines Überschusses an mathematischer Struktur.

Abstract

The discovery that Einstein's celebrated argument for general covariance, the 'point-coincidence argument', was actually a response to the 'hole argument' has generated an intense philosophical debate in the last thirty years. Even if the philosophical consequences of Einstein's argument turned out to be highly controversial, the protagonists of such a debate seem to agree on considering Einstein's argument as an expression of 'Leibniz equivalence', a modern version of Leibniz's celebrated indiscernibility arguments against Newton's absolute space. The paper attempts to show that the reference to Leibniz, however plausible at first sight, is actually in many respects misleading. In particular it is claimed that the Logical Empiricists offer a significant historical example of an attempt to interpret the point-coincidence argument as an indiscernibility argument in the sense of Leibniz, similar to those used in 19th century by Helmholtz, Hausdorff or Poincaré. However the logical empiricist account of General Relativity clearly failed to grasp the philosophical novelty of Einstein's theory. Thus, if Einstein's point coincidence/hole argument can be regarded as an indiscernibility argument, it cannot be an indiscernibility argument in the sense of Leibniz. Einstein rather introduced a new form of indiscernibility argument, which might be better described as an expression of 'Einstein-equivalence'. Developing some ideas of Weyl it is argued that, whereas Leibniz's arguments introduced the notion of 'symmetry' in the history of science, Einstein's argument seems to anticipate what we now call 'gauge freedom'. If in the first case indiscernibility arises from a lack of mathematical structure, in the second case it is a consequence of a surplus of mathematical structure.

1. Einleitung

In der Fachliteratur ist es mittlerweile üblich geworden, Einsteins sogenanntes Punkt-Koinzidenz-Argument, seine Antwort auf die berühmte ‚Lochbetrachtung' (inzwischen auch im deutschsprachigen Raum Loch-Argument genannt), als Ausdruck von Leibniz-Äquivalenz (Earman und Norton, 1987) zu betrachten. In dieser Gedankenlinie würde Einsteins Loch-Argument eindeutig dem Leibnizschen Verschiebungsargument gegen den Newtonschen absoluten Raum ähneln. Insbesondere in Einsteins Argument würde das, was wir inzwischen als „Diffeomorphismen"[1] bezeichnen, die gleiche Rolle wie die „Verschiebungen"[2] in Leibniz' Argument gegen Clarke spielen. In beiden Fällen können ‚Welten', die durch bestimmte Transformationen aufeinander abgebildet werden, als *dieselbe* ‚Welt' bezeichnet werden.

In einem bahnbrechenden Beitrag entwickelten John Earman und John D. Norton (1987) diesen Zusammenhang ausführlich. Die Protagonisten der umfangreichen Debatte, die von diesem Aufsatz ausgelöst wurde,[3] befürworten die Analogie zwischen Loch-Argument/Verschiebungsargument entschieden oder nehmen sie als ein polemisches Ziel auf.[4] Aus dieser Sicht sollte das Punkt-Koinzidenz-Argument als ein „neues Leibnizsches Argument" (Bartels, 1994; 1996) betrachtet werden.[5]

Ziel dieses Aufsatzes ist es aufzuzeigen, dass die Bezugnahme auf Leibniz in diesem Zusammenhang in vielerlei Hinsicht irreführend ist. Es geht dabei weniger um historische Genauigkeit, als um die philosophische Würdigung der grundlegenden Neuheit von Einsteins Ununterscheidbarkeitsargument im Vergleich zu denen von Leibniz. Ich behaupte, dass es möglich ist, diesen Umstand mittels einer historisch-kritischen Rekonstruktion der Rolle der Leibnizschen Ununterscheidbarkeitsargumente in der logisch-empiristischen Interpretation der Allgemeinen Relativitätstheorie zu beleuchten.

Indem sie dem gewichtigen Interpretationsansatz von Moritz Schlick folgten, haben sich die Logischen Empiristen explizit bemüht, das Punkt-Koinzidenz-Argument als eine Unterscheidbarkeit Leibnizscher Art zu verstehen, gleich den Argumenten, die in der philosophischen Debatte über Geometrie im 19. Jahrhundert weit verbreitet waren. In einer solchen Interpretation spielen beliebige stetige Verformungen der Welt, die die Nachbarschaft der Punkte erhalten, gerade die Rolle der Verschiebungen in Leibniz' ursprünglichem Argument. Allerdings vermochten die Logischen Empiristen durch diese Strategie keine plausible philosophische Interpretation der allgemeinen Relativitätstheorie zu liefern.[6] Deswegen könnte eine historische Analyse der Gründe und Ursprünge dieser philosophisch tiefgründigen und physikalisch höchst informierten, aber letztes Ende gescheiterten Interpretation zeigen, dass der Vergleich zwischen Leibniz' Ununterscheidbarkeitsargumenten und Einsteins Punkt-Koinzidenz-Argument/Lochbetrachtung irreführend ist. M.E. könnte dies zu einem wichtigen systematischen Ergebnis führen: *Leibniz und Einstein formulierten zwei verschiedene Arten von Ununterscheidbarkeitsargumenten* oder, anders ausgedrückt, Einstein führte eine neue Form von Ununterscheidbarkeitsargumenten ein, die sich von der klassischen, von Leibniz in der Korrespondenz mit Clarke formulierten, unterscheidet.

In dieser Hinsicht wird das wissenschaftliche und philosophische Werk Hermann Weyls eine fundamentale Rolle für diesen Aufsatz spielen: Weyl bietet m. E. nicht nur den Schlüssel zum Verständnis der Bedeutung von Ununterscheidbarkeitsargumenten à la Leibniz, sondern er hat auch einen wichtigen Hinweis für das Verständnis ihres Unterschieds zu Ununterscheidbarkeitsargumenten á la Einstein gegeben. Obwohl die Literatur zur ‚Lochbetrachtung' sehr umfangreich ist, glaube ich, dass diese Intuition Weyls bisher nicht berücksichtigt wurde.

Dieser historisch-kritischen Rekonstruktion folgend, glaube ich, dass es möglich ist aufzuzeigen, dass Einstein eine neue Form von Ununterscheidbarkeitsargumenten entwickelt hat, die mit Leibniz' ursprünglichen Ununterscheidbarkeitsargumenten keinesfalls verwechselt werden darf, wie es jedoch üblicherweise in der Fachliteratur geschieht. In einer ersten Annäherung können wir folgendes festhalten: Die ‚*Leibniz-Ununterscheidbarkeit*' ist die Folge von anscheinend physikalischen Unterschieden, die ihren Ausdruck im mathematischen Apparat der Theorie nicht finden können; das, was wir als ‚*Einstein-Ununterscheidbarkeit*' bezeichnen können, ist dagegen die Folge von anscheinend mathematischen Unterschieden, denen in der physikalischen Realität nichts korrespondiert. Die These, zu der meine Darlegung führen soll, ist also, dass eine Verwechslung dieser beiden Formen der Ununterscheidbarkeit unter dem Ausdruck ‚Leibniz-Äquivalenz' faktisch eine Untergrabung der bahnbrechenden Wichtigkeit des berühmten Einsteinschen Arguments bedeutet. Der Ausdruck ‚Einstein-Äquivalenz' könnte dagegen Einsteins Verdienst in gebührendem Maße würdigen.

2. Leibniz' Ununterscheidbarkeitsargumente: Ein Weylscher Ansatz

Hermann Weyl war vielleicht der erste, der die Tiefe der „philosophische[n] Wendung" (Weyl, 1952, 127) erkannte, die Leibniz dem einfachen geometrischen Begriff der geometrischen ‚Ähnlichkeit' gab. In einem Buch über Gruppentheorie schrieb er diesbezüglich: „Leibnitz [sic] declared: two figures are similar or equivalent if they cannot be distinguished from each other when each is considered by itself […] they have every imaginable property of objective meaning in common, in spite of being individually different" (Weyl, 1939, 15). Nach Weyl

hat Leibniz deswegen „the true general meaning of similitude" (ebd.) getroffen.

Weyls Interpretationsansatz kann eine Bestätigung in Leibniz' philosophischen Reflexionen über die Geometrie finden, deren Wichtigkeit jüngst in der Fachliteratur hervorgehoben wurde.[7] Leibniz' berühmter Definition der Ähnlichkeit zufolge sind zwei Figuren ähnlich, wenn sie separat betrachtet ununterscheidbar sind. Ähnliche Figuren (die die gleiche Form, aber eventuell unterschiedliche Größe haben) sind ‚dieselbe' Figur für den Geometer. Ihr Unterschied kann nicht ‚begrifflich' ausgedrückt werden, sondern enthüllt sich lediglich durch einen ‚anschaulichen Vergleich' (Couturat, 1902/1961, 412), also durch das, was Leibniz comperceptio nennt.[8]

Es ist möglich aufzuzeigen, dass die berühmten Leibnizschen Gedankenexperimente gerade dazu dienen, Idealfälle zu zeigen, bei denen die Möglichkeit solch eines Vergleiches fiktional ausgeschlossen ist. Wenn „auf irgendeine Weise nun Gott alles ändern würde, indem er die Proportionen aufrecht erhält, verlieren wir jedes Maß und können nicht wissen, um wie viel sich die Sachen verändert haben" (GM, VII, 276). In gängiger Terminologie, die auf Leibniz zurückgeht, wären das ursprüngliche und das transformierte Universum ‚ununterscheidbar'. Auf diese Weise ist dieses Leibnizsche *Nocturnal-doubling*-Gedankenexperiment einfach das Gegenstück seiner Definition der Ähnlichkeit.[9]

Leibniz war zutiefst von der Überlegenheit seiner eigenen Definition von Ähnlichkeit überzeugt, sodass er sich wiederholt um eine entsprechende phänomenologische Auffassung des *Kongruenz*begriffs bemühte.[10] Zwei Figuren sind kongruent, wenn sie erst unterschieden werden können, wenn gleichzeitig ein drittes Objekt wahrgenommen wird. Leibniz gab sich mit einer solchen Definition von Kongruenz nicht zufrieden (vgl. De Risi, 2007, 143 f.). Es ist jedoch plausibel, dass die berühmten Argumente, die Leibniz in seinem Briefwechsel mit Clarke gebrauchte – der Tausch von Osten und Westen oder das Verrücken aller Objekte um drei Fuß nach Osten – als Gegenstück einer solchen phänomenologischen Definition von Kongruenz betrachtet werden können, genau so – wie wir angenommen haben – wie das *Nocturnal-doubling*-Gedankenexperiment das Gegenstück zu der Leibnizschen Definition von Ähnlichkeit darstellt. Auch in diesem Fall besitzt die Euklidische Geometrie kein Mittel, die ursprüngliche von der transformierten Situation zu unterscheiden.

Weyl (1952) zu Folge erfassen Leibniz' Definitionen von Ähnlichkeit und Kongruenz intuitiv die fundamentale Idee, die dem modernen Begriff von „Automorphismus" zu Grunde liegt. In der Geometrie werden zwei Figuren als dieselbe bezeichnet, „wenn die eine durch einen Automorphismus in die andere überführt werden kann" (Weyl, 1927, 79), d.h. durch eine strukturerhaltende Abbildung des Raumes auf sich selbst: „Dies ist nun unsere Deutung der Leibnizschen Definition ähnlicher Figuren als solche, die nicht auseinander gehalten werden können, wenn sie jede für sich betrachtet werden" (Weyl, 1990, 98). Im Allgemeinen sind die Transformationen, auf die Leibniz sich bezieht, nicht willkürlich; sie lassen gewisse „objektiv[e] räumlich[e] Beziehungen ungeändert" (Weyl, 1930, 9). Welche Welten als ununterscheidbar gelten können, hängt von der Art der geometrischen Struktur ab, mit der wir es zu tun haben. Insbesondere Leibniz' Argumente setzen voraus, dass der Raum euklidisch ist, d.h., *sibi congruus* und auch *sibi similis* (LH, XXXIV, 1:14, Bl. 23 retro; De Risi, 2005).

Es ist also schwer zu verstehen, in welchem Sinne Leibniz' Argumente Newtons ‚absoluten Raum' angreifen sollten, da ein solcher Raum gerade mit denselben euklidischen Automorphismen des Leibnizschen ausgestattet ist.[11] Newtons Anliegen war nicht, den absoluten Ort im Raum festzustellen, sondern zu bestimmen, was es heißt, „am gleichen Ort (zu verschiedenen Zeiten)" zu sein, d.h. einen absoluten Unterschied zwischen Ruhe und Bewegung festzustellen. Leibniz' Verteidigung des Prinzips der Relativität aller Bewegung durch eine Erweiterung von Galileis (1632) Schiff-Gedankenexperiment auf die ganze Welt konnte dagegen den Unterschied zwischen Ruhe und Kreisbewegung nicht erklären, wie sie Newton (1687) in seinem Eimerexperiment gelang.

Auch in diesem Fall war Weyl (1927) wahrscheinlich der erste das zu betonen, was heute gängig geworden ist:[12] die Leibniz-Newton Auseinandersetzung kann erst verstanden werden, wenn man nicht die Automorphismen des Raumes, sondern jene der Raumzeit in Betracht zieht. Die Raumzeit der klassischen Mechanik, ‚Galileis Raumzeit', wie man sie oft nennt, besitzt eine „affingeometrische Struktur […] welche nicht den Unterschied zwischen Ruhe und Bewegung festlegt, sondern die gleichförmige Translation von allen andern Bewegungen absondert" (Weyl, 1922b, 57). In einer solche Struktur sind die Geraden „unter allen Linien objektiv ausgezeichnet, hingegen können aus der Schar aller Geraden die ‚vertikalen' nur durch Konvention, die sich auf individuel-

le Aufweisung stützt, herausgehoben werden" (Weyl, 1927, 70). ‚Leibniz' Raumzeit', in der alle Bewegung für relativ erklärt wird, enthält nicht genug Struktur um eine solche Trägheitsstruktur zu bestimmen, sie kann gerade Inertialbahnen nicht von den krummen Weltlinien der beschleunigten Bewegungen unterscheiden; dagegen enthält ‚Newtons Raumzeit' zu viel Struktur, indem sie verlangt, eine vertikale Weltlinie, einen Körper in Ruhe, objektiv zu bestimmen.

Weyls interpretativer Ansatz, obwohl in vielerlei Hinsicht viel zu vereinfacht, scheint trotzdem brauchbar zu sein, um die Voraussetzung, die Leibniz' Ununterscheidbarkeitsargumenten zu Grunde liegt, zu verstehen.[13] Solche Argumente funktionieren erst, wenn man im Voraus festgestellt hat, welche die relevante geometrische Struktur der Welt ist: *Zwei Welten, die auseinander durch eine strukturerhaltende Transformation, oder einen Automorphismus hervorgehen, sind geometrisch als ‚dieselbe' Welt anzusehen.* Man hat nämlich keine begrifflichen Ressourcen zur Verfügung, um den anschaulichen individuellen Unterschied zwischen Urbild und Abbild festzustellen, da alle geometrische Struktur, die im Urbild zu finden war, im Abbild wiedergefunden werden kann. Damit hat Leibniz den Begriff des ‚Automorphismus', oder, wie Weyl (1952) sagte, den Symmetriebegriff in die Geschichte der Naturwissenschaft eingeführt.

In der Automorphismengruppe, die die geometrische Struktur erhält, offenbart sich so der genaue ‚Charakter der Homogenität' des Raumes oder der Zeit. Durch eine solche Homogenität stellen sich Raum und Zeit dem materiellen Weltinhalt als „Formen der Erscheinungen gegenüber"; in ihrer Homogenität erweisen sie sich als „Prinzip […] der Individuation", indem sie die Existenz „verschiedener Individuen" ermöglichen, die doch „in allen ihren Beschaffenheiten einander gleichen" (Weyl, 1930, 8).

Leibniz' Argumente scheinen also nicht fähig nachzuweisen, dass der Raum die bloße Konsequenz der Relationen zwischen Körpern ist und nicht, wie Newton angeblich behauptet habe, eine Substanz, die eine von Körpern selbständige Existenz besitze. Weyl scheint interessanterweise an vielen Stellen zu betonen, dass dies gerade das ist, was Kant vage geahnt hatte, als er Raum und Zeit als ‚Formen' der Erscheinungen betrachtete:[14] Der Gegensatz zwischen *Relationalismus* und *Substantialismus* war alles in allem von Anfang an ein ‚Scheinproblem'.[15] Kant scheint damit gesehen zu haben, dass der relevante Gegensatz jener zwi-

schen Form und Inhalt ist: „Der Raum besitzt gemäß der Geometrie eine gewisse innere Struktur, unabhängig von dem materiellen Körper, der ihn erfüllt" (Weyl, 1920/21, 130). An diesem Standard muss selbstverständlich die radikale Neuheit der allgemeinen Relativitätstheorie gemessen werden: „Einstein erblickt", in Weyls Worten, dass in einer solchen Struktur „ein physikalisches Zustandsfeld von der gleichen Realität wie etwa das elektromagnetische Feld" (Weyl, 1920/21, 130) ist.

3. Leibnizsche Ununterscheidbarkeitsargumente in der Debatte des 19. Jahrhunderts über die Grundlagen der Geometrie

Weyls Ansatz, wenn es ihm auch sicherlich an einer präzisen Formulierung mangelt, liefert m. E. einen guten Hinweis, um eine Philosophie der Leibnizschen Ununterscheidbarkeitsargumente zu entwerfen. Dies wird besonders deutlich, wenn man die Geschichte von Ununterscheidbarkeitsargumenten in der Debatte des 19. Jahrhunderts über die Grundlagen der Geometrie in Betracht zieht. Die Protagonisten einer solchen Debatte, wie Hermann von Helmholtz, Henri Poincaré oder auch der junge Felix Hausdorff, verallgemeinerten Leibniz' Gedankenexperimente (auch wenn Leibniz selten explizit erwähnt wird): Zwei Welten werden nicht nur ununterscheidbar sein, wenn sie kongruent oder ähnlich sind, sondern sogar, wenn sie durch kompliziertere Verzerrungen und schließlich durch jegliche stetige Umformung überhaupt ineinander verformt werden. Mit anderen Worten kann man hier die Tendenz beobachten, Abbildungen des Raumes auf sich selbst (Automorphismen) zu untersuchen, die zunehmend schwächere Stufen geometrischer Struktur erhalten.

Diese Tendenz scheint sich zuerst mit dem Studium der invarianten Eigenschaften von Figuren unter Zentralprojektionen zu zeigen. Unter anderem im Werk von Poncelet (1822) und Steiner (1832) wurde die scheinbar unabdingbare Kopplung von Geometrie und ‚Maßbegriff' aufgehoben. Die Grundfrage der Geometrie, wie sie Ferdinand August Möbius (1827) in der Vorrede zum *Barycentrischen Calcul* formuliert, ist das Studium von „geometrische[n] Verwandtschaften" unter Figuren: Gleichheit, Ähnlichkeit, Affinität, Kollineation, bis zu „elementar[er] Verwandtschaft", in der „von je zwei einander unend-

lich nahen Punkten der einen [Figur] auch die ihnen entsprechenden der andern einander unendlich nahe sind" (Möbius, 1863, 18). Klein, der nicht zufälligerweise bis 1887 an der Gesamtausgabe von Möbius' Werken (Möbius, 1885) mitarbeitete, gab schon 1872 in seinem ‚Erlanger Programm' den populärsten Ausdruck dieser Tendenz:[16] Klein definierte bekannterweise den Raum als eine *n*-dimensionale Mannigfaltigkeit, eine Verallgemeinerung des dreidimensionalen Koordinatenraumes der analytischen Geometrie, das was Sophus Lie (1893) später eine ‚Zahlenmannigfaltigkeit' nannte, die Gesamtheit der *n*-Tuple von reellen Werten. Dann betrachtete er nur diejenigen (in Koordinaten ausgedrückten) Beziehungen als ‚geometrisch', die beim Übergang von einem Koordinatensystem in ein anderes invariant bleiben. In späterer Sprache definiert Klein dann eine ‚Geometrie' als Punktmenge mit einer ‚Struktur'; wobei die strukturerhaltenden bijektiven Abbildungen des Raumes auf sich selbst eine Gruppe bilden, die man Automorphismengruppe nennen kann. Man kann an Automorphismengruppen denken, die progressiv weniger Struktur erhalten, und damit Räume betrachten, die mit einem progressiv höheren Grad von Homogenität oder Symmetrie ausgestattet sind.

Dieser Tendenz scheint sich der Ansatz Riemanns (1854/1868) in seinem Habilitationsvortrag von 1854 entgegen zu stellen (vgl. Norton, 1999). Riemann betrachtete den Raum als ein Beispiel für eine *n*-dimensionale stetige Mannigfaltigkeit, in der die verschiedenen Individuen durch die verschiedenen (reellen) Werte von n Variablen unterschieden werden (vgl. Scholz, 1982). Dann formulierte Riemann die Hypothese, dass der Raum sich von anderen *n*-dimensionalen stetigen Mannigfaltigkeiten (z.B. der der Farben) unterscheidet, weil die Entfernung zwischen zwei unendlich benachbarten Raumpunkten durch eine ‚quadratische Differentialform' bestimmt wird, eine verallgemeinerte Form des Pythagoreischen Lehrsatzes: $ds^2 = \Sigma g_{\mu\nu} dx_\mu dx_\nu$, wobei $g_{\mu\nu}(x_i)$ Konstanten oder Funktionen der Koordinaten sind. Der Unterschied zu dem vorherigen, am besten durch Klein repräsentierten Ansatz, tritt deutlich zu Tage, wenn man daran denkt, dass ein Automorphismus eines Riemannschen Raumes eine längentreue Abbildung dieses Raumes auf sich selbst ist und es vorkommen kann, dass die identische Abbildung der einzige Automorphismus ist (vgl. Laugwitz, 1996). Riemanns radikaler Ansatz, bei *welchem der Raum vollkommen inhomogen sein kann und keine Symmetrie zeigt*, spielte eigentlich eine Nebenrolle in der Debatte

des 19. Jahrhunderts über die Grundlagen der Geometrie (vgl. Hawkins, 1980; 2000). Riemanns Ansatz entwickelte sich vielmehr in einer nicht-geometrischen Tradition, die von den Arbeiten Erwin Bruno Christoffels (1869) bis zu Gregorio Ricci-Curbastros (1884; 1888; 1892; 1893) und Tullio Levi-Civitas ‚absolutem Differentialkalkül' (Levi-Civita und Ricci-Curbastro, 1900) reichte (vgl. Reich, 1994). Hier wurde das sogenannte, von Riemann (1854/1868; 1861/1876) gestellte und teilweise ausgearbeitete ‚Äquivalenzproblem' (vgl. Dell'Aglio, 1996) entwickelt: das Problem unter welchen Bedingungen es möglich ist, eine quadratische Differentialform in eine andere zu transformieren, wenn die Veränderlichen x_i durch andere Veränderliche x'_i ersetzt werden, die stetige und differenzierbare Funktionen der ersten sind. Es müssen also die Bedingungen gefunden werden, unter welchen die Funktionen $g_{\mu\nu}(x)$ durch neue Funktionen $g'_{\mu\nu}(x')$ ersetzt werden können, so dass ds^2 unverändert bleibt, d.h. $ds^2 = \Sigma g_{\mu\nu} dx_\mu dx_\nu = \Sigma\, g'_{\mu\nu} dx'_\mu dx'_\nu$. Geometrisch bedeutet das, dass verschiedene, ineinander transformierbare $g_{\mu\nu}$-Systeme *dieselbe* ‚metrische' Geometrie ausdrücken.[17]

3.1 Helmholtz, Hausdorff, Poincaré

Im Riemannschen Kontext sind Ununterscheidbarkeitsargumente im Sinne Leibniz' überhaupt nicht sinnvoll; denn im allgemeinen Fall gibt es in einem Riemannschen Raum nicht genug Automorphismen, um eine Anwendung von solchen Argumenten zu ermöglichen. Wenn man in einem inhomogenen Raum die Welt zehn Meter nach rechts verschieben würde, würde man den Unterschied geometrisch feststellen können.[18] Dass Riemanns Ansatz ‚philosophisch' nicht wirklich wahrgenommen wurde (vgl. Hawkins, 1980), sieht man gerade daran, dass die philosophische Debatte über die nicht-Euklidischen Geometrien im Lauf des 19. Jahrhunderts gerade von Ununterscheidbarkeitsargumenten im Sinne Leibniz' dominiert ist, die progressiv breitere Gruppen von Automorphismen betrachten.

Die Möglichkeit ‚Leibnizsche' Argumente in diesem Kontext zu formulieren scheint ihren Ursprung in ersten ‚Interpretationen' (etwa geographischen Karten) der nicht-Euklidischen Geometrien im Euklidischen Raum zu haben. Ein Beispiel hierfür ist etwa Eugenio Beltramis (1868/1902–1920) projektive Abbildung der pseudosphärischen

Fläche in der Euklidischen Ebene, die später auf den Raum verallgemeinert wurde. In solchen ‚Modellen', wie man sie später genannt hat, treten, wie bei allen Karten, Verzerrungen auf. Da aber Messungen im Modell mit ebenfalls verzerrten Maßstäben durchgeführt werden, wird es unmöglich, geometrisch den Unterschied festzulegen. Alle Theoreme der pseudosphärischen Geometrie gelten in Beltramis Euklidischem Abbild, gerade wie im Original.

Helmholtz übernahm 1870 in seinem Vortrag „Über Ursprung und Bedeutung der Geometrischen Axiome" Beltramis Idee um ein Ununterscheidbarkeitsargument im Sinne Leibniz' vorzuschlagen: Er bemerkt zuerst, „dass wenn die sämtlichen linearen Dimensionen" der Welt „in gleichem Verhältnisse, z.B. alle auf die Hälfte, verkleinert oder alle auf das Doppelte vergrössert würden, wir eine solche Aenderung durch unsere Mittel der Raumanschauung gar nicht würden bemerken können" (Helmholtz, 1870/1876, 44). Das gleiche würde passieren, „wenn die Dehnung oder Zusammenziehung nach verschiedenen Richtungen hin verschieden wäre" (ebd.). Letztlich betrachtete Helmholtz das Abbild der Welt in einem Convexspiegel. Die Welt würde damit völlig verzerrt erscheinen, wobei solche Verzerrungen von den Bewohnern der gespiegelten Welt nicht bemerkt werden könnten:

> Man denke an das Abbild der Welt in einem Convexspiegel. […] Das Bild eines Mannes, der mit einem Maßstab eine von dem Spiegel sich entfernende gerade Linie abmisst, würde immer mehr zusammenschrumpfen, je mehr das Original sich entfernt, aber mit seinem ebenfalls zusammenschrumpfenden Maßstab würde der Mann im Bilde genau dieselbe Zahl von Centimetern herauszählen, wie der Mann in der Wirklichkeit. Kurz, ich sehe nicht, wie die Männer im Spiegel herausbringen sollten, dass ihre Körper nicht feste Körper seien und ihre Erfahrungen gute Beispiele für die Richtigkeit der Axiome des Euklides. Könnten sie aber hinausschauen in unsere Welt, wie wir hineinschauen in die ihrige, ohne die Grenze überschreiten zu können, so würden sie unsere Welt für das Bild eines Convexspiegels erklären müssen und von uns gerade so reden, wie wir von ihnen, und wenn sich die Männer beider Welten mit einander besprechen könnten, so würde, soweit ich sehe, keiner den anderen überzeugen können, dass er die wahren Verhältnisse habe, der andere die verzerrten; ja ich kann nicht erkennen, dass eine solche Frage überhaupt einen Sinn hätte, so lange wir keine mechanischen Betrachtungen einmischen. (Helmholtz, 1870/1876, 45)

Zwei Körper, die im Urbild zur Koinzidenz gebracht werden können, würden auch im verzerrten Abbild zusammenfallen; die Bewoh-

ner der Kugelspiegel würden ihren Raum als Euklidisch betrachten und würden glauben, dass wir in einem verzerrten Universum leben. „Nun ist Beltrami's Abbildung des pseudosphärischen Raumes in einer Vollkugel des Euklid'schen Raumes von ganz ähnlicher Art" (Helmholtz, 1870/1876, 45). Es geht nämlich einfach um das entgegengesetzte Gedankenexperiment: in diesem Fall „würden Beobachter, deren Leiber selbst dieser Veränderung regelmässig unterworfen wären, bei geometrischen Messungen, wie sie sie ausführen könnten, Ergebnisse erhalten, als lebten sie selbst im pseudosphärischen Raume" (ebd.).[19]

Die Bedeutung von „Helmholtzens Convexspiegel" (NL FH, Kapsel 24; Fasz. 71; Bl. 33) als Ununterscheidbarkeitsargument wird deutlicher bei Felix Hausdorff.[20] Besonders in seinem philosophischen Jugendwerk *Das Chaos in kosmischer Auslese*, das unter dem Pseudonym Paul Mongré (1898) erschien. Nach Hausdorff habe nämlich Helmholtz in diesem „von der Gegenpartei zwar vielfach citirte[n] aber selten verstandene[n] Beispiel […] in populär einleuchtender Weise den Satz" umschrieben, „dass eine Raumtransformation sich der empirischen Wahrnehmung entzieht" (Mongré, 1898, 99f.). Obwohl Helmholtz' Absicht wahrscheinlich nur diejenige war, die nicht-Euklidischen Geometrien zu veranschaulichen, ordnet Hausdorff Helmholtz' Gedankenexperiment in eine Hierarchie von möglichen Ununterscheidbarkeitsargumenten ein. Hausdorff betrachtet zuerst Verschiebungen, Drehungen, Spiegelungen sowie Streckungen des gesamten Universums und legt dar, dass ein Bewusstsein notwendigerweise sich dessen nicht bewußt wäre (Mongré, 1898, 88ff.).

Hausdorff zu Folge würde man aber zwei Welten für ununterscheidbar halten, auch wenn die Objekte des Universums durch eine beliebige Deformation willkürlich in willkürliche Richtungen verzerrt würden. In einer ersten Annäherung kann man sich auf die Transformationen beschränken, die von „Unstetigkeiten und Singularitäten" (Mongré, 1898, 97) frei sind. Es wird nur vorausgesetzt, dass jedem Punkt des ursprünglichen Raumes genau ein Punkt seines verzerrten Abbildes entspricht, und ebenso umgekehrt, so dass die Koordinaten des einen Punktes stetige und eindeutige Funktionen – ganz gleichgültig welche – der Koordinaten des entsprechenden Punktes sind, d.h. zwei unendlich benachbarte Punkte zwei ebenfalls unendlich benachbarten entsprechen sollen: „Von gewissen Transformationen des Raumes würden wir nichts bemerken: das ist der Refrain meiner transzendentalen Dialektik" (NL

FH, Kapsel 49; Fasz. 1079; Bl. 26). In einem leider undatierten Fragment seines Nachlasses gesteht Hausdorff, dass er ähnliche Argumentationen „auch bei anderen (Poincaré)" fand (NL FH, Kapsel 49; Fasz. 1079; Bl. 4). Es ist schwer zu sagen, welche Schrift Poincarés Hausdorff genau meinte, aber die Ähnlichkeit von Hausdorffs und Poincarés Verfahren ist in vielen Hinsichten nicht zu leugnen (vgl. Epple, 2006). Poincaré verwendet vielleicht noch deutlicher als Hausdorff die Strategie, eine Klimax von Ununterscheidbarkeitsargumenten zu verwenden, denen Automorphismen entsprechen, die progressiv weniger Struktur erhalten.[21]

Nicht nur zwei Welten wären nicht zu unterscheiden, wenn sie einander kongruent oder ähnlich wären (vgl. Poincaré, 1905). Man kann noch kompliziertere Transformationen denken. Gemäß der Hypothese von Lorentz und Fitzgerald (vgl. Brown, 2005) erleiden alle Körper bei einer translatorischen Bewegung eine Kontraktion in Richtung dieser Translation. Geometrisch geht es um eine Zentralprojektion, bei welcher ein Körper, der in der Ruhe kugelförmig ist, die Form eines abgeplatteten Rotationsellipsoides annimmt, wenn er bewegt wird; der Beobachter aber wird ihn immer noch für kugelförmig halten, denn er selbst erleidet eine analoge Deformation. Wenn also alle Objekte der Welt ohne Ausnahmen, eine solche Deformation erleiden würden, gäbe es keine Möglichkeit den Unterschied zu bemerken.

> La déformation de Lorentz-Fitzgerald dont les lois sont particulièrement simples, on pourrait imaginer une déformation tout à fait quelconque. (Poincaré, 1907, 4)

Wenn wir die Welt in einem jener kompliziert gestalteten Spiegel betrachten, bleiben dabei doch die gegenseitigen Verhältnisse der einzelnen Teile dieser Welt unverändert. Alle relevante Struktur, die im Urbild zu finden ist, würde auch im Abbild auftauchen. Wenn sich zwei wirkliche Gegenstände berühren, so scheinen sich auch ihre Spiegelbilder zu berühren. Damit gäbe es geometrisch kein Mittel, um den Unterschied zu bemerken.[22] Für Poincaré hat also der Raum unabhängig von unseren Messinstrumenten, weder metrische noch projektive Eigenschaften; er hat nur topologische Eigenschaften, also solche, mit denen sich die *Analysis Situs* befasst.

Mathematisch heißt das, dass der Raum ein n-dimensionales Kontinuum ist, eine variété oder Mannigfaltigkeit, die Gesamtheit von n von-

einander unabhängig veränderlichen Größen, die fähig sind, alle reellen Werte anzunehmen. Wenn man eine Koordinatentransformation durchführt, ersetzt man die *n*-Koordinaten durch n beliebige Funktionen dieser *n*-Koordinaten. In seinem Aufsatz „Analysis Situs" hatte Poincaré (1895) präzisiert, dass solche Funktionen ‚continues' sind, und ‚elles ont des dérivées continues', d. h. dass sie stetig und differenzierbar sind. Poincaré nennt sie ‚Homomorphismen'. Es ist aber klar, dass sie in heutiger Sprechweise vielmehr den ‚Diffeomorphismen' entsprechen.[23]

Von diesem Standpunkt aus hat Poincaré ein Ununterscheidbarkeitsargument Leibnizscher Art formuliert, bei dem ‚Diffeomorphismen' genau die Rolle von Streckungen, Verschiebungen oder Spiegelungen in Leibniz' ursprünglichen Argumenten übernehmen. Poincaré (1905; 1907) bezieht sich hier auf die bahnbrechenden Arbeiten Riemanns (1892) im Bereich der *Analysis Situs*, die ein amorphes Kontinuum untersuchen, „[que] possède un certain nombre de propriétés, *exemptes de toute idée de mesure*" (Poincaré, 1905, 76; Hervorhebung von mir).[24]

Poincaré (1891, 773) scheint dagegen viel weniger fähig zu sein, die geometrische Bedeutung von Riemanns Ansatz in dessen Habilitationsvortrag wahrzunehmen. Riemanns Hauptanliegen waren dort selbstverständlich gerade die Maßverhältnisse des Raumes (vgl. Scholz, 1992). Wie wir gesehen haben, untersuchte Riemann, wie die Funktionen $g_{\mu\nu}$ der metrischen Fundamentalform $\Sigma\, g_{\mu\nu} dx_\mu dx_\nu$ sich transformieren, nachdem die ursprünglichen Koordinaten x_i durch neue Koordinaten x'_i ersetzt werden, die die stetigen und differenzierbaren Funktionen (Diffeomorphismen) der ersten sind, während ds^2 ungeändert bleibt. Es geht also nicht um einen Raum ohne Maßstruktur, sondern um Klassen von Räumen mit derselben Maßstruktur (isometrisch). Nicht ineinander transformierbare $g_{\mu\nu}$-Systeme beschreiben nicht nur Eigenschaften des Koordinatensystems, sondern auch ‚innere' Maßverhältnisse des Raumes. Es ist dieses Problem, das in der Tradition von Christoffel zu Ricci und Levi-Civita von einem rein analytischen und nicht-geometrischen Standpunkt ausgearbeitet wurde. Hier spielen eindeutig Diffeomorphismen nicht die Rolle, die Verschiebungen oder Streckungen in Leibniz Argumenten spielen. Es ist aber gerade diese Tradition, an die Einstein bekanntermaßen anknüpfte, um die Allgemeine Relativitätstheorie zu formulieren.

4. Einsteins Punkt-Koinzidenz-Argument und die Forderung der „allgemeinen Kovarianz"

Poincarés auf der Lorentz-Kontraktion basierendes Gedankenexperiment, das gerade erwähnt wurde, stammt aus seinen Untersuchungen über das Relativitätsprinzip. Am 5. Juni 1905 kündigte Poincaré (1906b) der *Académie des sciences eine Mémoire* „Sur la dynamique de l'électron" an, die im Januar 1906 auf der *Rendiconti del Circolo Matematico di Palermo* erschien. Hier wurde erstmals gezeigt, dass diejenigen Transformationen, die Poincaré selber als ‚Lorentz-Transformationen' bezeichnete, eine ‚Gruppe' bilden; noch dazu zeigte Poincaré, dass die Rechnungen vereinfacht werden können, wenn die Zeit betrachtet wird, als ob sie eine vierte Dimension des Raumes wäre: „les quatre coordonnées d'un point de notre nouvel espace ne seraient pas x, y, z et t, mais x, y, z et $t\sqrt{-1}$" (Poincaré, 1913a, 53).

Einstein publizierte 1905 seinen Aufsatz „Zur Elektrodynamik bewegter Körper", in dem die so genannte Lorentz-Kontraktion, die Poincaré immer noch als ein von Kräften verursachtes ‚dynamisches' Phänomen betrachtete, zu einer reinen ‚kinematischen' Änderung wurde. Einstein ging ‚algebraisch' vor; er betrachtete die Lorentz-Kovarianz der Naturgesetze gegenüber der Umrechnung von Raumkoordinaten und Zeiten. Bekanntlich ‚übersetzte' erst Hermann Minkowski (1909) in seiner Kölner Rede „Raum und Zeit" Einsteins Theorie in ‚geometrische' Sprache (ohne aber auf Poincaré zu verweisen): die Lorentz-Transformationen werden als Automorphismen einer einzigen geometrischen Struktur betrachtet, der Raumzeit. Minkowski bezeichnete sie bekanntlich als die ‚Welt', „[d]ie Mannigfaltigkeit aller denkbaren Wertsysteme x, y, z, t"; die Lorentz-Transformationen werden als „homogene lineare […] Transformationen von x, y, z, t in vier neue Variable x', y', z', t'" (Minkowski, 1909, 2) betrachtet, die die Entfernung zwischen zwei benachbarten Punkten $ds^2 = dx^2 + dy^2 + dz^2 - c^2t^2$ invariant lässt. Wie Felix Klein (1910, 540) schrieb: „Was die modernen Physiker Relativitätstheorie nennen, ist die Invariantentheorie des vierdimensionalen Raum-Zeit-Gebietes x, y, z, t (der Minkowskischen ‚Welt') gegenüber einer bestimmten Gruppe von Kollineationen, eben der ‚Lorentzgruppe'".

4.1 Die allgemeine Relativitätstheorie

In seinem Vortrag „L'espace et le temps", gehalten in London am 4. Mai 1912, betrachtete Poincaré (1913a, 54), ohne Einstein oder Minkowski zu erwähnen, den von ihm eingeführten vierdimensionalen Ansatz immer noch als „une convention [que] nous semblait commode". Gerade um dieselbe Zeit begann dagegen Einstein Minkowskis geometrische Darstellung der Speziellen Relativitätstheorie ernst zu nehmen. Um 1912 hatte er den ‚entscheidenden Gedanken' (vgl. Einstein, 1923), die Analogie zwischen Gaußscher Flächentheorie und dem Problem der Gravitation. Nach der Speziellen Relativitätstheorie gehen die, die allgemeinen Naturgesetze ausdrückenden Gleichungen in Gleichungen derselben Form über, wenn man statt der Raum-Zeit-Variablen x, y, z, t unter Benutzung der Lorentz-Transformation die Raum-Zeit-Variablen x', y', z', t' einführt (Lorentz Kovarianz). Nach der allgemeinen Relativitätstheorie dagegen müssen die Gleichungen bei Anwendung beliebiger Substitutionen der Variablen x_1, x_2, x_3, x_4 in Gleichungen derselben Form übergehen, denn jede Transformation entspricht dem Übergang eines Gaußschen Koordinatensystems in ein anderes (Allgemeine Kovarianz). Alle ineinander durch stetige und differenzierbare Transformationen überführbare Gaußschen Koordinatensysteme sind gleichwertig für die Formulierung der Naturgesetze. Da (nahe eines gegebenen Raumzeitpunktes) das Gravitationsfeld durch ein Beschleunigungsfeld imitiert werden kann (Äquivalenzprinzip), wurde das Prinzip der ‚allgemeinen Kovarianz' zur selben Zeit als Ausdehnung des Relativitätsprinzips zur beschleunigten Bewegungen und als Lösung des Gravitationsproblems betrachtet.

Als Einstein 1912 von Prag nach Zürich zurück kam, bat er seinen Studienfreund Marcel Grossman um Hilfe. Grossman führte ihn in die Werke von Riemann, Christoffel und besonders von Ricci und Levi-Civita ein (Levi-Civita und Ricci-Curbastro, 1900) und schrieb zusammen mit Einstein den ersten Entwurf zu einer Gravitationstheorie (Einstein und Grossmann, 1913). In einem kurze Zeit später publizierten Aufsatz „Mathematische Begriffsbildung zur Gravitationstheorie" fasst Grossman das Thema ihrer Zusammenarbeit zusammen: „Der mathematische Grundgedanke der Einstein'schen Gravitationstheorie" besteht in der Idee, „ein Gravitationsfeld zu charakterisieren durch eine quadratische Differentialform mit variablen Koeffzienten […]. Von grundle-

gender Bedeutung ist hierbei die berühmte Abhandlung von Christoffel […] und die auf dieser fussende Abhandlung von Ricci und Levi-Civita" (Grossmann, 1913, 291). Bezogen auf letztere Arbeit schreibt er dort dann weiter, dass die „Verfasser Methoden [entwickelten], die den Differentialgleichungen der mathematischen Physik eine invariante, d.h. vom Koordinatensystem unabhängige Form geben lassen" (ebd.).

4.2 Die Lochbetrachtung

Einsteins Problem war dann, die allgemeinen kovarianten Differentialgleichungen zu finden, die die Koeffizienten $g_{\mu\nu}$ ‚eindeutig' als Funktion der Materieverteilung bestimmen können. Bei der Suche nach solchen Feldgleichungen stieß Einstein auch auf ein in den letzten dreißig Jahren (seit Stachel, 1980) sehr berühmt gewordenes ‚philosophisches Problem'. Dieses Problem, das verschiedene Formulierungen erhielt,[25] wird am deutlichsten im § 12 von „Die formale Grundlage der allgemeinen Relativitätstheorie" (Einstein, 1914a), der ersten systematischen Darstellung der Entwurfstheorie. Es geht um die so genannte ‚Lochbetrachtung'. Einstein betrachtete Lösungen $g_{\mu\nu}$ seiner Feldgleichungen in Bezug auf das Koordinatensystem x_i in einem Loch, d.h einem endlichen Teil „des Kontinuum, in welchem ein materieller Vorgang nicht stattfindet" (CPAE, Doc. 9, 110). Er führte dann ein neues Koordinatensystem x'_i innerhalb des Loches ein. Nach der Regel der ‚absoluten Differentialrechnung' werden die ursprünglichen $g_{\mu\nu}(x)$ durch neue $g'_{\mu\nu}(x')$ ersetzt. Wenn man diese letzte Lösung bezüglich des ursprünglichen Koordinatensystems x_i, d.h. als $g'_{\mu\nu}(x)$ betrachtet, würden diese immer noch Lösungen der Feldgleichungen sein. Man hat also zwei Lösungen der Feldgleichungen, für die gleiche Materieverteilung im Bezug auf dasselbe Koordinatensystem. Die zwei Lösungen erscheinen physikalisch verschieden: vor der Transformation würden sich Teilchen, die das Loch durchqueren, z.B. auf geraden Linien bewegen, aber nicht nach der Koordinatentransformation; sie würden vor der Transformation bestimmte Punkte durchlaufen, aber durch andere Punkte nach der Transformation. Da aber die Materieverteilung außerhalb des Loches unverändert geblieben ist, würden die Feldgleichungen das ‚Kausalgesetz' verletzen; sie wären nicht in der Lage, das Geschehen im Gravitationsfeld ‚eindeutig' festzulegen.

Lange Zeit wurde unter Wissenschaftshistorikern dieses Argument gegen allgemein kovariante Feldgleichungen einfach als ein ‚Fehler' betrachtet (vgl. z.B. Pais, 1982). Einstein schien nicht berücksichtigt zu haben, dass die Lösungen $g_{\mu\nu}(x)$ und $g'_{\mu\nu}(x')$ einfach ‚isometrisch' sind. Es geht um eine triviale Anwendung der ‚absoluten Differentialrechnung'. Nach genauer Untersuchung (Stachel, 1980) wurde aber festgestellt, dass Einsteins Argument eigentlich die Lösungen $g'_{\mu\nu}(x')$ und $g'_{\mu\nu}(x)$ vergleicht. Anschaulich gesprochen, glaubte Einstein, dass es möglich sei, das ursprüngliche Koordinatensystem stetig innerhalb des Loches zu verformen, indem man aber das unverformte Koordinatensystem im Hintergrund unangetastet lässt. Durch eine bloße mathematische Umrechnung bekommt man damit zwei physikalisch verschiedene Systeme von Feldlinien innerhalb des Loches, obwohl die Materie außerhalb des Loches ungeändert geblieben ist.

4.3 Das Punkt-Koinzidenz-Argument bei Kretschmann und Einstein

1915 publizierte Erich Kretschmann in den *Annalen der Physik* den zweiteiligen Aufsatz: „Über die prinzipielle Bestimmbarkeit der berechtigten Bezugssysteme beliebiger Relativitätstheorien" (Kretschmann, 1915; vgl. Giovanelli, 2013). In dieser etwas weitschweifigen Untersuchung über die Relationen zwischen Transformationsgruppen und Relativitätspostulaten, zeigte Kretschmann „[m]it weitgehender Benutzung der von E. Mach und H. Poincaré gegebenen Analysen physikalischer Erfahrung"[26], dass eine solche Erfahrung „von bestimmten räumlich-zeitlichen Beziehungen nur solche topologischer Art liefern kann" (Kretschmann, 1915, 911). Unter ‚topologischen Beziehungen' zwischen räumlich-zeitlich ausgedehnten Gegenständen versteht Kretschmann das „räumlich-zeitliche Zusammenfallen oder Nichtzusammenfallen von Teilen des Meßinstrumentes mit Teilen des Meßgegenstandes" (ebd., 914). Die einzige geometrische Struktur, die der Erfahrung zugänglich ist, besteht also aus solchen ‚topologischen Beziehungen'. Wegen der „Invarianz der Beobachtungstatsachen gegen beliebige stetige Raum-Zeittransformationen" kann dann behauptet werden, „daß zwischen zwei quantitativ verschiedenen, aber topologisch gleichen Abbildungen der Erscheinungswelt […] [d]urch bloße Beobachtungen in keinem Falle eine sicher begründete Entscheidung getroffen werden kann" (ebd.).

Kretschmanns Aufsatz erschien am 21. Dezember. Nur einige Tage später, am 26. Dezember 1915 schrieb Einstein an Paul Ehrenfest:

> An die Stelle des § 12 [die Lochbetrachtung] hat folgende Darlegung zu treten. Das physikalisch Reale an dem Weltgeschehen (im Gegensatz zu dem von der Wahl des Bezugssystem Abhängigen) besteht in raumzeitlichen Koinzidenzen […]. Wenn zwei Systeme der $g_{\mu\nu}$ (bezw. allg. der zur Beschreibung der Welt verwandten Variabeln) so beschaffen sind, dass man das zweite aus dem ersten durch blosse Raum-Zeit-Transformation erhalten kann, so sind sie völlig gleichbedeutend. Denn sie haben alle zeiträumlichen Punktkoinzidenzen gemeinsam, d.h. alles Beobachtbare. (CPAE 8a, Doc. 173, 228; Brief an Paul Ehrenfest; 26.12.1915)

Es kann vermutet werden, dass dieses Argument Einsteins, das davor nie in seinen Schriften vorkommt, von Kretschmann inspiriert war (vgl. Howard und Norton, 1993).

Im Gegensatz zu Kretschmann, der sich auf Poincaré bezieht, fügt aber Einstein das Argument in eine ganz andere Tradition ein, die von den Arbeiten von Riemann, Christoffel, Ricci-Curbastro und Levi-Civita geprägt ist. Zwei metrische Felder, d.h. zwei Systeme der $g_{\mu\nu}$, die in den Punkt-Koinzidenzen übereinstimmen, d.h. die ineinander durch stetig und differenzierbare Koordinatentransformationen verformbar sind, drücken dasselbe Gravitationsfeld aus. Das Punkt-Koinzidenz-Argument erscheint damit als Antwort auf die Lochbetrachtung, wie Einstein auch in späteren Briefen, insbesondere an Michele Besso bestätigt: „Anstelle der Lochbetrachtung tritt folgende Überlegung. Real ist physikalisch nichts als die Gesamtheit der raumzeitlichen Punktkoinzidenzen" (CPAE 8a, Doc. 26, 235).

Dass Einsteins Paradox weniger trivial ist als man ursprünglich dachte, kann auch daran erkannt werden, dass auch Hendrik Lorentz sich genau mit dem gleichen Problem konfrontiert sah.[27] Die Natur des Punkt-Koinzidenz-Arguments als einem Ununterscheidbarkeitsargument ist wahrscheinlich nirgendwo offensichtlicher als in Einsteins Briefwechsel mit Ehrenfest.[28] In seinem verloren gegangenen Brief stellte sich Ehrenfest vermutlich Lichtstrahlen vor, die von einem Stern ausgehen, durch eines von Einsteins Löchern hindurch laufen und durch eine Blende eine Platte erreichen. In seiner Antwort schlägt Einstein vor, die von Ehrenfest beschriebene Situation auf ein ‚vollkommen dehnbares Pauspapier' zu zeichnen. Deformiere man dann das Pauspapier beliebig, so erscheinen die zwei Situationen verschieden. Die Bahnen der Licht-

strahlen, die in der ursprünglichen Situation geradlinig waren, sind jetzt gekrümmt und durchlaufen andere Punkte: Einsteins Feldgleichungen scheinen nicht fähig zu sein, die physikalische Situation eindeutig festzulegen.

> Wenn Du die Figur nun wieder auf orthogonale Briefpapierkoordinaten beziehst, so ist die Lösung *mathematisch* eine andere als vorher, natürlich auch bezüglich der $g_{\mu\nu}$. Aber *physikalisch ist es genau dasselbe*, weil eben das Briefpapier-Koordinatensystem nur etwas eingebildetes ist. Immer erhalten dieselben Punkte der Platte Licht. […] Wesentlich ist: Solange das Zeichenpapier, d.h. ›der Raum‹ keine Realität hat, unterscheiden sich beide Figuren überhaupt nicht. Es kommt nur auf ‚Koinzidenzen' an. z.B. darauf, ob Plattenpunkte vom Lichte getroffen werden oder nicht. So wird der Unterschied Deiner Lösungen *A* und *B zu einem blossen Unterschied der Darstellung bei physikalischer Übereinstimmung*. Dies wird Dir bei genauer Erwägung sicher einleuchten (CPAE, 8a, Doc. 180, S. 238; Hervorhebungen von mir).

An dieser berühmten Passage ist hervorzuheben, dass Einstein merkte, dass dies lediglich „ein mathematischer Unterschied" ist, während es „physikalisch genau dasselbe" ist. Das Hintergrundkoordinatensystem („das Briefpapier-Koordinatensystem"), unter dessen Berücksichtigung die Situation als verschieden erscheinen würde, ist „nur etwas Eingebildetes": „Es kommt nur auf Koinzidenzen an".

4.4 Die öffentliche Version des Punkt-Koinzidenz-Arguments

Aber die Verbindung zwischen Punkt-Koinzidenz-Argument und Lochbetrachtung wird nie in veröffentlichten Schriften erwähnt. Am 20. März 1916 schickte Einstein den *Annalen der Physik* die erste Gesamtdarstellung der Allgemeinen Relativitätstheorie, die dann am 21. Mai publiziert wurde. Im §3 findet sich die mehr als berühmte Passage: „Daß diese Forderung der allgemeinen Kovarianz […] eine natürliche Forderung ist, geht aus folgender Überlegung hervor. Alle unsere zeiträumlichen Konstatierungen laufen stets auf die Bestimmung zeiträumlicher Koinzidenzen hinaus" (Einstein, 1916, 776). Die Koordinatensysteme werden nur dafür gebraucht, um die „Beschreibung der Gesamtheit solcher Koinzidenzen" (ebd.) zu gestatten. Man ordnet der Welt vier raumzeitliche Variablen x_1, x_2, x_3, x_4 derart zu, dass jedem Punktereignis durch ein einziges Wertsysteme der Variablen x_1, x_2, x_3, x_4 bestimmt ist. Die Koinzidenz wird dann durch die Übereinstimmung der Koordina-

ten charakterisiert, während räumlich benachbarten Punkten sehr wenig verschiedene Zahlenwerte zugeordnet werden. Dem Grundgedanken des allgemeinen Relativitätsprinzips entspricht die Aussage: „alle Gaußschen Koordinatensysteme sind für die Formulierung der allgemeinen Naturgesetze prinzipiell gleichwertig" (ebd). Da alle Gaußschen Koordinatensysteme in Punkt-Koinzidenzen übereinstimmen, ist ein Koordinatensystem so gut wie jedes andere um die Gesetze der Physik auszudrücken. Das Punkt-Koinzidenz-Argument erscheint hier einfach als identisch mit der Forderung der allgemeinen Kovarianz.

Das Problem, das Einstein beschäftigte, erschien später in der Literatur in der Form, die David Hilbert (1917) ihm in seiner berühmten „Zweiten Mitteilung" von Februar 1916 gab. Hilbert formulierte die Lochbetrachtung (deutlicher als Einstein) als eine Verletzung des Kausalgesetzes: Die allgemeine Relativitätstheorie wäre unfähig aus der Kenntnis der physikalischen Größen $g_{\mu\nu}$ in Gegenwart und Vergangenheit eindeutig ihre Werte in der Zukunft zu bestimmen.[29] Besonders in seiner Vorlesung von Dezember 1916 „Das Kausalitätsprinzip in der Physik" erhält das Argument eine Form, die an die von Einstein deutlich erinnert: „Wir können leicht einen Vorgang konstruieren, der mit dem alten Kausalitätsprinzip unvereinbar ist" (Hilbert, 1916/1917, 4f.). Nehmen wir das Lösungssystem unserer 10 Gleichungen, das für alle Werte von $x_4 = t$ durch die bestimmten Funktionen $g_{\mu\nu}$ gegeben ist. Führen wir dann eine beliebige Transformation $x_i = x_i\,(x'_1, x'_2, x'_3, x'_4)$ aus, so sind die entsprechenden $g'_{\mu\nu}(x')$ „nur ein anderer mathematischer Ausdruck für dasselbe mathematische Geschehen" (ebd.).

Nun sind aber die 10 Differentialgleichungen gegenüber dieser Transformation invariant, also bleiben die neuen Funktionen $g'(x')$ „Lösungen der Differentialgleichungen", wenn man darin die x'_i durch irgendwelche Funktionen von x'_i, z.B. durch die x_i ersetzt: $g'_{\mu\nu}(x)$ „sind ebenfalls Lösungen der 10 Differentialgleichungen. Sie stellen natürlich einen ganz anderen individuellen physikalischen Vorgang dar". Insbesondere wählt man $x'_i = x_i$ für $t = 0$, aber $x'_i \neq x_i$ für $t > 0$, doch stetig inklusive aller Ableitungen, so stimmt das »physikalische Ereignis«, $g'_{\mu\nu}$ bis zur Zeit $t = 0$ mit dem Ereignis $g_{\mu\nu}$ überein, aber es „weicht dann vollkommen von ihm ab" (ebd.). Der Unterschied ist aber nur illusorisch, es ist nur ein durch den mathematischen Formalismus eingeführter Unterschied. Es geht um verschiedene mathematische Ausdrücke des gleichen physikalischen Geschehens. „Die Aufklärung des Paradoxons erhalten

wir", behauptet Hilbert, in einer Sprechweise, die an Einsteins Punkt-Koinzidenz-Argument erinnert, „wenn wir nur den Begriff der Relativität schärfer zu erfassen suchen". Nicht nur sind die Weltgesetze vom Bezugssystem unabhängig, vielmehr muss man sagen, dass „jede einzelne Behauptung über eine Begebenheit oder ein Zusammentreffen von Begebenheiten physikalisch nur dann einen Sinn hat, wenn sie von der Benennung unabhängig, d.h. wenn sie invariant ist" (ebd.).

5. Das Punkt-Koinzidenz-Argument als ein Ununterscheidbarkeitsargument Leibnizscher Art. Schlicks Interpretation der Allgemeinen Relativitätstheorie

Das Punkt-Koinzidenz-Argument, wie man es in Einsteins (1916) Aufsatz „Die Grundlage der allgemeinen Relativitätstheorie" lesen konnte, übte eine unwiderstehliche Faszination auf die Philosophen aus. Gerade weil es einfach war, es als eines der vielen Ununterscheidbarkeitsargumente zu interpretieren, die man bei Helmholtz, Poincaré oder – wie wir gesehen haben – Hausdorff finden kann.[30] Auf diese Weise konnte Einsteins neue Theorie im Kontext der zu diesem Zeitpunkt wohlbekannten Debatte über Geometrie gelesen werden. Das beste Beispiel dieser Strategie ist sicher Moritz Schlicks klassisches Werk „Raum und Zeit in der gegenwärtigen Physik".[31] Poincaré, Helmholtz und später sogar Hausdorff sind nicht zufälligerweise die Autoren, auf die sich Schlick bezieht. Er verwendete nämlich gerade das Verfahren, das bei solchen Autoren üblich war, nämlich Automorphismen zu berücksichtigen, die allmählich weniger Struktur erhalten.

Schlick fängt mit der „Fiktion einer durchgehenden Größenänderung der Welt" an, d.h. mit dem „Fall, daß die gedachte transformierte Welt der ursprünglichen geometrisch ähnlich ist" (Schlick, 1917b, 202). Dann betrachtet er eine kompliziertere Transformation, bei welcher die „Abmessungen aller Objekte sich nur nach einer Richtung hin beliebig verlängerten oder verkürzten" (ebd.). Letztlich gibt Schlick zu, dass es möglich ist, „die Gegenstände des Universums nach beliebigen Richtungen beliebig verzerrt vor[zu]stellen" (ebd.), wenn nur die Koordinaten von unendlich benachbarten Punkten in die Koordinaten von unendlich benachbarten Punkten übergehen:

In mathematischer Sprechweise können wir dies Resultat ausdrücken, indem wir sagen: zwei Welten, die durch eine völlig beliebige (aber stetige und eindeutige) Punkttransformation ineinander übergeführt werden können, sind hinsichtlich ihrer physikalischen Gegenständlichkeit miteinander identisch. Das heißt: wenn das Universum sich irgendwie deformierte, so daß die Punkte aller physischen Körper dadurch an neue Orte gerückt werden, so ist damit […] überhaupt gar keine feststellbare, keine ‚wirkliche' Änderung eingetreten, wenn die Koordinaten eines physischen Punktes am neuen Orte auch ganz beliebige Funktionen der Koordinaten seines alten Ortes sind; nur wird natürlich vorauszusetzen sein, daß die Körperpunkte ihren Zusammenhang bewahren, daß also solche, die vor der Deformation benachbart waren, es auch nachher bleiben (d.h. jene Funktionen müssen stetig sein), und ferner darf jedem Punkt der ursprünglichen Welt nur ein Punkt der neuen entsprechen, und umgekehrt (d.h. die Funktionen müssen eindeutig sein). (Schlick, 1917b, 164)

Das Prinzip, dass alle Gaußschen Koordinatensysteme für die Formulierung der allgemeinen Naturgesetze gleichwertig sind, bedeutet, dass beim Übergang von einem System Gaußscher Koordinaten zu einem beliebig deformierten, raumzeitliche Koinzidenzen, d.h. alle wirklich feststellbaren Tatsachen der Physik, unberührt bleiben:

Statt zu sagen: ich deformiere die Welt in bestimmter Weise, kann ich ebensogut sagen: ich beschreibe die unveränderte Welt durch neue Koordinaten […]. Beides ist einfach dasselbe, und jene gedachten Deformationen würden gar keine reale Änderung der Welt bedeuten, sondern nur eine Beziehung auf andere Koordinaten. (Schlick, 1917b, 165)

Das Punkt-Koinzidenz-Argument wird also als ein Ununterscheidbarkeitsargument Leibnizscher Art betrachtet, gerade wie diejenigen, die in der Debatte des 19. Jahrhunderts über die Grundlagen der Geometrie so verbreitet waren. Die Argumente, die Poincaré, Helmholtz oder Hausdorff in Bezug auf den Raum angewandt haben, hat Einstein einfach auf die vierdimensionale Raum-Zeit-Mannigfaltigkeit übertragen:

Denken wir uns eine derartige durchgehende Veränderung im Universum vorgenommen, welche jeden physischen Punkt so an einen anderen Raum-Zeit-Punkt bringt, daß seine neuen Koordinaten x'_1, x'_2, x'_3, x'_4 ganz beliebige (nur stetige und eindeutige) Funktionen seiner vorigen Koordinaten x_1, x_2, x_3, x_4 sind, so ist wiederum die neue Welt von der alten physikalisch überhaupt gar nicht verschieden, die ganze Änderung ist weiter nichts als eine Transformation auf andere Koordinaten. Denn das durch unsere Apparate allein Beobachtbare, die raum-zeitlichen Koinzidenzen, bleibt ja erhalten.

Zwei Punkte, die in dem einen Universum in dem Weltpunkt x_1, x_2, x_3, x_4 zusammenfielen, koinzidieren im andern in dem Weltpunkt x'_1, x'_2, x'_3, x'_4 ihr Zusammenfallen – und weiter läßt sich ja nichts beobachten – findet in der zweiten Welt genau so gut statt, wie in der ersten. (Schlick, 1917b, 235)

Das Punkt-Koinzidenz-Argument, so formuliert, ist also nochmals ein Ununterscheidbarkeitsargument à la Leibniz, in dem Diffeomorphismen einfach die Rolle der Verschiebungen übernommen haben. Beide Universen würden also nicht zu unterscheiden sein, da alle relevante geometrische Struktur ja in beiden Universen gleich aussähe: „das räumlich-zeitlich deformierte Universum ist mit dem ursprünglichen in jeder Hinsicht physikalisch identisch, sofern nur nach der Deformation alle räumlich-zeitlichen Koinzidenzen der Punktpaare dieselben sind wie vorher" (Schlick, 1917b, 201).

Im für die Buchversion hinzugefügten Kapitel „Beziehungen zur Philosophie" konnte Schlick dann Einsteins Argument mit seinem Ansatz in der *Allgemeine Erkenntnisslehre* (Schlick, 1918, die schon 1916 so gut wie fertig war) in Verbindung setzen; er leitete den objektiven Raum-Zeitbegriff aus den sinnlichen Empfindungskomplexen unterschiedlicher Vorstellungsräume ab: da „diese Koinzidenzen für alle anschaulichen Räume der verschiedenen Sinne und Individuen stets übereinstimmend auftreten", eben deshalb wird durch sie „ein objektiver, d.h. von den Einzelerlebnissen unabhängiger", für alle gültiger Punkt definiert (Schlick, 1917b, 274; vgl. Engler, 2006).

6. Das Punkt-Koinzidenz-Argument und Kretschmanns Argument gegen Allgemeine Kovarianz

Schlicks Strategie, das Punkt-Koinzidenz-Argument im Kontext der Debatte des 19. Jahrhunderts über die Grundlagen der Geometrie zu interpretieren, ist alles andere als unplausibel. Wie wir gesehen haben, hatte Kretschmann (1915) das Argument gerade aus dieser Tradition übernommen. Es war gerade Kretschmann der betonte, dass unsere physikalische Erkenntnis nur zu „den durch Beobachtungen verifizierbaren […] rein topologischen Inhalten der Gesetze" (Kretschmann, 1915, 938), d.h. nur zu Punkt-Koinzidenzen Zugang hat.

Indem aber Schlick das Punkt-Koinzidenz-Argument als eine Radikalisierung von Leibniz' Ununterscheidbarkeitsargumenten nach dem

Muster von Poincaré oder Hausdorff betrachtete, verallgemeinerte er die Automorphismengruppe der Raumzeit zur Diffeomorphismengruppe, zur Gruppe aller stetigen und eindeutigen Punkttransformationen, die nur die ‚topologischen' Punkt-Koinzidenzen erhalten. Die mathematische Struktur zu der die Erfahrung Zugang hat ist nicht reich genug um Welten zu unterscheiden, die, obwohl eine das verzerrte Abbild der anderen ist, in Punkt-Koinzidenzen übereinstimmen.

Es ist gerade Kretschmanns berühmter zweiter, 1918 in den *Annalen der Physik* erschienener Aufsatz „Über den physikalischen Sinn der Relativitätspostulate" (Kretschmann, 1918), der zeigte, dass diese Interpretation schwer zu verteidigen ist. Hier wendete Kretschmann das Punkt-Koinzidenz-Argument gegen Einstein. Dabei verweist er selbstverständlich auf seinen Aufsatz von 1915. Der Kontext des Punkt-Koinzidenz-Arguments ist aber jetzt auch für Kretschmann nicht mehr Poincarés Philosophie der Geometrie, sondern die ‚absolute Differentialrechnung' von Ricci und Levi-Civita:

> Denn nach den Untersuchungen von Ricci und Levi-Cività [sic] dürfte es kaum zweifelhaft sein, daß man jedes physikalische Gleichungssystem ohne Änderung seines durch Beobachtungen prüfbaren Inhaltes auf eine allgemein kovariante Form bringen kann. Das leuchtet von vornherein ein, wenn man sich wieder vergegenwärtigt, daß in Strenge nur rein topologische Tatsachen des Naturgeschehens oder nach Einstein Koinzidenzen beobachtbar sind. (Kretschmann, 1918, 579)

Wenn Einstein seine „Forderung der Kovarianz der physikalischen Gleichungen" bei beliebigen stetigen Koordinatentransformationen auf die Tatsache stützt, dass „alle physikalische Erfahrung letzten Endes in der Beobachtung rein topologischer Beziehungen oder ‚Koinzidenzen' zwischen den räumlich zeitlichen Beobachtungsgegenständen besteht", dann ist diese Forderung trivial: alle Theorien können „durch Einführung der unbestimmten Koeffzienten $g_{\mu\nu}$ in den Ausdruck des Linienelementes auf allgemeine kovariante Form gebracht werden" (Kretschmann, 1918, 578).

Die Forderung der allgemeinen Kovarianz hat nichts mit der Verallgemeinerung des Relativitätsprinzips zu tun. Kretschmann in seiner etwas schwerfälligen Prosa zeigt, dass ein Relativitätsprinzip einer Raum-Zeit-Lehre modern gesprochen durch Automorphismen bestimmt ist, die lichtartige und zeitartige Weltlinien in sich selbst übertragen. In allgemein-relativistischer Raumzeit (im allgemeinsten Fall) ist der ein-

zige Automorphismus, der lichtartige und zeitartige Weltlinien in sich selbst überträgt, die Identität (vgl. Rynasiewicz, 1999): „Die Einsteinsche Theorie genügt demnach physikalisch […] überhaupt keinem Relativitäts-postulate; sie ist ihrem Inhalte nach eine vollkommene Absolutheorie" (Kretschmann, 1918, 610).

Das Prinzip der allgemeinen Kovarianz kann nicht als ein Relativitätspostulat interpretiert werden (vgl. Norton, 1995). Eine Theorie, in welcher die lichtartigen und zeitartigen Weltlinien nach „beliebige[r] stetige[r] Verzerrung wieder in sich selbst übergehen" (Kretschmann, 1918, 612), d.h. eine Theorie, die invariant wäre gegenüber der breitesten Automorphismengruppe, der Diffeomorphimsengruppe, wäre sinnlos. Ganz im Gegenteil schrumpft in der allgemein-relativistischen Raumzeit im allgemeinen Fall die Automorphismengruppe zur Identität zusammen. Im Kontext der allgemeinen Relativitätstheorie ergeben also Ununterscheidbarkeitsargumente im Sinne Leibniz' überhaupt keinen Sinn. Wenn Kretschmann 1915 die logisch-empiristische Interpretation alles in allem antizipiert hatte, hatte er sie schon 1917 widerlegt.

6.1 Einsteins und Weyls Antwort auf Kretschmann

Einsteins Antwort auf Kretschmann, dass eine allgemein kovariante Formulierung von vor-allgemein-relativistischen Theorien unnötig kompliziert wäre (Einstein, 1918), gilt gewöhnlich als nicht besonders überzeugend. Wie Weyl (1918b) in *Raum, Zeit, Materie* betont, ohne explizit auf Kretchmann zu verweisen, ist es unverkennbar, dass das Postulat der allgemeinen Invarianz nur „eine rein mathematische Angelegenheit" (Weyl, 1918b, 205) ist, die keine physikalische Bedeutung hat: „die Naturgesetze so formulieren, dass sie invariant sind gegenüber beliebigen Transformationen; das ist eine mathematische Wesensmöglichkeit, es liegt darin gar keine besondere Eigentümlichkeit dieser Gesetze" (ebd.). Die physikalische Neuigkeit der allgemeinen Relativitätstheorie muss vielmehr darin gefunden werden, dass während es in der speziellen Relativitätstheorie immer möglich ist, die Koordinaten so zu wählen, dass die $g_{\mu\nu}$' die Minkowski-Werte annehmen, in der allgemeinen die $g_{\mu\nu}$', wie die elektro-magnetische Potentiale, „physikalische Zustandsgrößen [sind], denen etwas Reales entspricht" (Weyl, 1918b, 198) und die partiellen Differentialgleichungen unterworfen sind:

„Weniger in der Forderung der allgemeinen Invarianz, sondern in dieser Annahme erblicke ich daher den eigentlichen Kern der allgemeinen Relativitätstheorie" (Weyl, 1918b, 181).

Es gibt aber einen wesentlichen Unterschied zwischen Einsteins Gravitationstheorie und den klassischen Feldtheorien. Hat man in der Allgemeinen Relativitätstheorie einmal eine bestimmte Lösung der allgemeinen kovarianten Feldgleichungen gefunden, kann man daraus durch eine bloße Koordinatenänderung eine beliebige Anzahl von anderen Lösungen ableiten. Wenn die Anfangswerte für die $g_{\mu\nu}$ gegeben sind, bleibt nämlich die Evolution von vier der zehn Komponenten von den $g_{\mu\nu}$ durch Einsteins Feldgleichungen unbestimmt:

> In dem System der Feld- und Gravitationsgesetze sind daher vier *überschüssige* Gleichungen enthalten. In der Tat muß die allgemeine Lösung vier willkürliche Funktionen enthalten, da die Gleichungen ja, zufolge ihrer invarianten Natur das Koordinatensystem der x_i vollständig unbestimmt lassen und mithin durch willkürliche stetige Transformation dieser Koordinaten aus einer Lösung dieser Gleichungen immer wiederum Lösungen hervorgehen (die aber *objektiv denselben Weltverlauf* darstellen). (Weyl, 1918b, 215; Hervorhebungen von mir).

Eben diese überschüssigen mathematischen Freiheitsgrade, die keinen entsprechenden Anteil in der Realität haben, bilden das Problem, das Einstein beschäftigt hatte. Das Punkt-Koinzidenz-Argument zeigte ihm gerade, dass die bis zu einer bloßen Koordinatentransformation verschiedenen Lösungen der Feldgleichungen denselben Weltverlauf darstellen: zwei Systeme von Feldlinien, die sich auf gleiche Weise schneiden, d.h. die in raumzeitlichen Koinzidenzen übereinstimmen, definieren die gleiche physikalische Situation. Es geht nur um mathematisch verschiedene Darstellungen desselben physikalischen Feldes.

Im Nachhinein kann man sagen, dass es gerade Weyl war, der die Mittel zur Verfügung stellte, die Natur einer solchen ‚Redundanz' begrifflich einzuordnen. 1918 unternahm Weyl (Weyl, 1918a) den Versuch, die die Gravitation beschreibende Allgemeine Relativitätstheorie mit dem Elektromagnetismus in einem einzigen geometrischen Schema zu vereinigen.[32] Wenn Einsteins Gravitationstheorie die „Willkürlichkeit des Koordinatensystems" voraussetzt, verlangt Weyls Theorie außerdem die „Willkürlichkeit der Maßeinheiten" (Weyl, 1919, 101); nicht nur die Invarianz der Naturgesetze gegenüber einer beliebigen Koordinatentransformation (‚Koordinaten-Invarianz', Weyl, 1919, 101), d.h. gegen-

über der Ersetzung $g_{\mu\nu}(x)$ durch $g'_{\mu\nu}(x')$; sondern auch gegenüber einer beliebigen Änderung der Maßeinheit, d.h. gegenüber der Ersetzung von $g_{\mu\nu}$ durch $\lambda g_{\mu\nu}$ (‚Maßstab-Invarianz'), wobei das Koordinatensystem das gleiche bleibt. In beiden Fällen gibt es mathematisch verschiedene $g_{\mu\nu}$-Systeme, die dieselbe einzelne physische Situation beschreiben. „Und bekamen wir damals die Gravitation, so bekommen wir jetzt den Elektromagnetismus geschenkt" (Weyl, 1919, 112).

7. Weyl über Einsteins Ununterscheidbarkeitsargument

Das Punkt-Koinzidenz-Argument drückt also nicht einen Mangel an mathematischer Struktur aus; sondern es ist die Antwort auf die Entdeckung eines Überschusses an mathematischer Struktur, die man in Anspielung an Weyls Theorie Eichfreiheit nennt. Schlick hat die Bedeutung des Punkt-Koinzidenz-Arguments auf zweierlei Art verkannt: Das ‚öffentliche' Punkt-Koinzidenz-Argument ist kein Ununterscheidbarkeitsargument. Es ist nur der triviale Ausdruck der Forderung der allgemeinen Kovarianz, die alle Theorien erfüllen können. Das ‚private' Punkt-Koinzidenz-Argument ist ein Ununterscheidbarkeitsbargument, aber offensichtlich nicht ein Ununterscheidbarkeitsbargument á la Leibniz, d.h. nicht eines von solchen Argumenten wie diejenigen von Helmholtz, Poincaré oder Hausdorff.

In den folgenden Jahren hat Schlick das Vertrauen in seinen Interpretationsansatz nie verloren. 1921 gab Schlick Helmholtz' *Erkenntnistheoretische Schriften* heraus (Helmholtz, 1921). In Schlicks Kommentaren erscheint Einsteins Argument schon implizit in Helmholtz' Konvexspiegel Gedankenexperiment: Die beiden Welten sind ununterscheidbar, weil „alle Punkt-Koinzidenzen erhalten bleiben" (Helmholtz, 1921, 34; Schlicks Kommentar). In der 3. Auflage von *Raum und Zeit in der gegenwärtigen Physik* erwähnt Schlick außer Poincaré und Helmholtz auch Hausdorff. „[E]rst nach Erscheinen der zweiten Auflage" (Schlick, 1920b, 198, Anm. 1), so liest man in einer Fußnote, lernte Schlick „das höchst scharfsinnige und faszinierende Buch", „Das Chaos in kosmischer Auslese. Ein erkenntniskritischer Versuch von Paul Mongré, Leipzig 1898" (Schlick, 1922, 28) kennen. „Das fünfte Kapitel dieses Werkes", schreibt Schlick weiter, „gibt eine sehr vollkommene Darstellung der oben im Text folgenden Erörterungen. Nicht nur die Gedanken

Poincarés, sondern auch einige der oben hinzugefügten Ergänzungen sind dort bereits vorweggenommen" (ebd.).

7.1 Weyls Anspielung auf die Lochbetrachtung

Die Logischen Empiristen, Schlick folgend, werden dann überzeugt, dass „die allgemeine Theorie […] [i]n den Koinzidenzen das einzig Invariante [sieht] und […] nur die Maßbeziehungen zwischen den Koinzidenzen [relativiert]" (Reichenbach, 1922, 332). In „Massenträgheit und Kosmos" (Weyl, 1924) macht Weyl klar, was der Fehler der Logischen Empiristen war.

Die Logischen Empiristen, um ein von Weyl oft verwendetes Bild zu benutzen, verglichen die allgemein-relativistische Raumzeit mit einer ‚Plastelinenmasse' (Weyl, 1924, 198), die geometrisch ununterscheidbar von jeder anderen ist, die von ihr durch irgendeine stetige Deformation hervorgeht.[33] Auf diese Weise würde aber die Raumzeit zu einem baren Kontinuum in dem nur das Zusammenfallen und die unmittelbare Nachbarschaft von Ereignissen relevant wären: „Darum", bemerkt Weyl, „ist aber eine Lösung des Problems, die […] eine Weltstruktur überhaupt ausschalten will, unmöglich" (Weyl, 1927, 73). Die Raumzeit ist nicht „amorph, sondern trägt eine Struktur" (Weyl, 1931, 49), insbesondere eine affingeometrische Struktur, die als Standard für nicht-Beschleunigung gilt. Einsteins Theorie hat nicht eine solche Struktur abgeschafft, sondern gezeigt, dass sie ein dynamisches Feld ist: „*An dem Dualismus von Führung und Kraft wird also festgehalten; aber die Führung ist ein physikalisches Zustandsfeld*" (wie das elektromagnetische), ein „Führungsfeld", wie es Weyl nennt (Weyl, 1924, 198; Weyls Hervorhebung).

Wenn dieses Resultat in der Literatur erwähnt wird (Coleman und Korté, 1982), wird jedoch übergangen, dass Weyl in „Massenträgheit und Kosmos", in dialogischer Form, den Unterschied eines solchen fehlerhaften Arguments zu Einsteins Ununterscheidbarkeitsargument gezeigt hat. Soweit ich sehen kann, wurde Weyls Anspielung auf Einsteins Ununterscheidbarkeitsargument in der Literatur nicht berücksichtigt, obwohl Weyl, der unter anderem 1912–1913 Einsteins Kollege in Zürich war, sicher eine relevante Quelle ist. Einer der Protagonisten des Dialogs sagt dem anderen:

Begehst du da nicht den gleichen Fehler, den Einstein 1914 machte, als er aus dem Kausalitätsprinzip auf die Unmöglichkeit der allgemeinen Relativitätstheorie schloß? Denn, so sagte er, worin die Naturgesetze invariant sind gegenüber beliebigen Koordinatentransformationen, so erhalte ich aus einer Lösung durch Transformation unendlich viele neue. Teile ich der Welt durch einen dreidimensionalen Querschnitt, welcher ihre beiden Säume voneinander trennt, in zwei Teile und verwende nur solche Transformationen, welche die untere Hälfte unberührt lassen, so stimmen alle diese Lösungen gleichwohl in der unteren Welthälfte mit der ursprünglichen überein (Weyl, 1924, 202).

Man definiert den Zustand der Welt in einem Augenblick durch den dreidimensionalen Querschnitt $t = 0$, (eine sog. Cauchy Oberfläche), der Vergangenheit und Zukunft trennt; es sollte dann möglich sein nach streng gültigen bekannten mathematischen Gesetzen den künftigen Geschehensverlauf ableiten zu können. Man führe aber in der ‚Zukunft' $t > 0$ (in der oberen Welthälfte), eine Koordinatentransformation durch. Dann würden sich bewegende Körper durch andere Punkte laufen, obwohl die Vergangenheit $t \leq 0$ (die untere Welthälfte) unverändert geblieben ist. ‚Einsteins Fehler' besteht nach Weyl im Folgenden:

Er übersah, daß alle diese Lösungen auch in der oberen Welthälfte objektiv den gleichen Zustandsverlauf wiedergeben, daß ein Unterschied nur bestünde, wenn die vierdimensionale Welt ein stehendes Medium wäre, in das sich die Spuren der materiellen Vorgänge so oder so einzeichnen. Und nur dann kann man auch die Möglichkeiten der Realisierung, von denen du sprichst, als verschieden anerkennen. Ein solches stehendes Medium wird aber, ohne Zweifel mit deinem Beifall, von der Relativitätstheorie durchaus geleugnet. (Weyl, 1924, 202).

Wenn die vierdimensionale leere Welt, die stetige vierdimensionale Mannigfaltigkeit aller möglichen Raumzeitpunkte, ein stehendes Medium wäre, dann hätte es einen Sinn zu sagen, dass die Bahnen, die vor der Koordinatentransformation durch bestimmte Raumzeitpunkte laufen, nach der Koordinatentransformation durch andere Punkte laufen würden. Wie wir gesehen haben, ist es nicht sinnvoll, ein solches Kontinuum von individuierten Raumzeitpunkten als Hintergrund zu postulieren auf dem das Gravitationsfeld ‚liegen' sollte. Da alle Körper ohne Ausnahme durch die Gravitationskraft auf die gleiche Weise ‚abgelenkt' werden, hätte man kein Mittel um ein solches Feld von dem angeblichen Hintergrund abzusondern. Das ‚bare Kontinuum', auf dem das Gravitationsfeld und die anderen Felder liegen, ist physikalisch irrelevant.

Unterschiede die bloß im Bezug auf diesen Hintergrund erscheinen, sind also keine Unterschiede.

Man unterscheidet also auch hier „zwischen dem amorphen Kontinuum und seiner metrischen Struktur" (Weyl, 1931, 51). Die Logischen Empiristen hatten aber irrigerweise behauptet, dass in der Allgemeinen Relativitätstheorie ein solcher strukturloser ‚topologischer' Raum die einzige physikalisch relevante geometrischer Struktur der Theorie sei; der Raum, könnte man sagen, wird dadurch von der metrischen Struktur abgelöst, die bloß willkürlich ist. In Einsteins Theorie ist aber gerade die metrische Struktur die einzige relevante Struktur, während, wie Weyls Darstellung von Einsteins Unterscheidbarkeitsargument zeigt, der strukturlose Raum eine bloße mathematische Redundanz ist: „Die metrische Struktur", schreibt Weyl, „wird dadurch gleichsam vom Raume abgelöst, sie wird zu einem in dem zurückbleibenden strukturlosen Raume existierenden Feld" (Weyl, 1925/1988, 5).

7.2 Weyl und der Begriff von ‚Eichinvarianz'

Nach der Entdeckung, dass die absolute Längeneinheit atomistisch durch die Wellenlänge h/mc des Elektrons geliefert wird, verlor Weyls Theorie von 1918 definitiv ihre Überzeugungskraft. Schon 1927 hatte aber Fritz London vorgeschlagen, dass „die komplexe Amplitude der de Broglie'schen Welle", die gleiche Rolle wie das Eichmaß in Weyls ursprünglicher Theorie übernehmen könnte (London, 1927, 380). Weyl selber wandte später seinen ursprünglichen Ansatz auf die Phasenverschiebung der Dirac-Gleichung an, die relativistische Wellengleichung, die Verhalten von Elektronen beschreibt.[34] Die Dirac-Gleichung, bleibt unverändert, wenn man die vierkomponentige Wellenfunktion ψ durch $\psi^{\lambda ei}$ ersetzt (man bekommt nämlich das gleiche Interferenzmuster in einem Doppelschlitz-Experiment). Wenn man eine stetig differenzierbare, von Zeit und Ort abhängige Änderung des Faktors λ durchführt, muss ein Kraftfeld eingeführt werden, das die lokale Veränderung der Phasenverschiebung kompensiert (vgl. C.-N. Yang und Mills, 1954). Das leistet gerade das elektromagnetische Feld: „das Elektrische Feld [ist] ein notwendiges Begleitphänomen nicht des Gravitationsfeldes, sondern des materiellen, durch ψ dargestellten Wellenfeldes" (Weyl, 1929a, 348).

Obwohl man „lieber von Phasen- statt von Eichinvarianz sprechen" sollte (Weyl, 1951, 81), erlaubt sich Weyl aber einen solchen nicht mehr reellen, sondern rein imaginären Faktor λ ‚Eichfaktor' zu nennen. Es geht nämlich, wie Weyl betont, um eine „Invarianzeigenschaft, die in formaler Hinsicht derjenigen gleicht, die ich in meiner Theorie von Gravitation und Elektrizität vom Jahre 1918 ab Eichinvarianz bezeichnet hatte" (Weyl, 1929a, 331). Die Eichinvarianz verbindet aber jetzt die elektromagnetischen Potentiale nicht mit den $g_{\mu\nu}$ der Gravitation, sondern mit den ψ des Materiefeldes. Auch in diesem Fall wird dasselbe physikalische Feld mathematisch durch eine Klasse äquivalenter Dirac-Gleichungen beschrieben. Dabei betonte Weyl nämlich immer noch die Analogie zur Allgemeinen Relativitätstheorie: „Da die Eichinvarianz eine willkürliche Funktion λ einschließt, hat sie den Charakter ‚allgemeiner' Relativität und kann natürlich nur in ihrem Rahmen verstanden werden" (Weyl, 1929a, 331).

Die Analogie zwischen Eichinvarianz und Koordinateninvarianz bleibt dann begrifflich fundamental. Es geht um mathematisch verschiedene Darstellungen, die aber die gleiche physikalische Situation beschreiben. Philosophisch relevant ist, dass Einsteins Lochbetrachtung eben diese Form von Ununterscheidbarkeit voraus zu setzen scheint.[35] In Einsteins Ununterscheidbarkeitsargument spielen also Diffeomorphismen nicht die Rolle von Symmetrietransformationen (wie Verschiebungen oder Spiegelungen in Leibniz' Argument), sondern von Eichtransformationen. Die Lochbetrachtung kann überwunden werden, wenn man feststellt, dass physikalisch dasselbe Gravitationsfeld durch eine Äquivalenzklasse von mathematisch verschiedenen metrischen Feldern ausgedrückt wird.

8. Fazit: Leibniz-Äquivalenz und Einstein-Äquivalenz

Wie erwähnt, hat Weyl seit den dreißiger Jahren hervorgehoben, dass Leibniz' Ununterscheidsbarkeitsargumente den Symmetriebegriff in die Geschichte der Naturwissenschaft eingeführt haben.[36] Gerade Weyl scheint aber auch die Mittel zu Verfügung gestellt zu haben, um Einsteins Ununterscheidbarkeitsargument als Ausdruck davon zu verstehen, was wir heute, immer noch in Anspielung auf Weyls Theorie, ‚Eichfreiheit' nennen (vgl. Rovelli, 2004). Dies ist meines Erachtens

aufschlussreich für die angeregte moderne Debatte über Raumzeittheorien, die der wichtige Aufsatz von Earman und Norton (1987) entfacht und dominiert hat. Das Punkt-Koinzidenz-Argument kann nicht einfach als „a stronger version of a famous argument due to Leibniz himself against Newton" (Janssen, 2005, 74), als Ausdruck von Leibniz-Äquivalenz, verstanden werden. Die Analogie zwischen Leibniz' und Einsteins Argumenten könnte prima facie zwar plausibel erscheinen: in beiden Fällen gibt es ja durch die Theorie erlaubte, verschiedene mögliche Welten, die sich letztlich als dieselbe Welt erweisen. Jedoch ist die Ähnlichkeit beider Argumente nur eine scheinbare. Wenn man in beiden Fällen von Ununterscheidbarkeit sprechen darf, so handelt es sich doch um zwei verschiedene Arten von Ununterscheidbarkeit:

– eine Ununterscheidbarkeit, die aus einem *Mangel an mathematischer Struktur* entsteht, in der angenommene physikalische Unterschiede keinen Ausdruck finden können, die deshalb für *mathematisch irrelevant* erklärt werden. Da alle relevante mathematische Struktur im Urbild auch im Abbild gefunden werden kann, kann der Unterschied zwischen Abbild und Urbild in der mathematischen Struktur der Theorie nicht ausgedrückt werden; Urbild und Abbild sind individuell verschieden aber *begrifflich identisch*.
– eine Ununterscheidbarkeit, die aus einem *Überschuss an mathematischer Struktur* folgt. In Bezug auf eine solche Struktur entstehen mathematische Unterschiede, die jedoch für *physikalisch irrelevant* erklärt werden: Urbild und Abbild sind *begrifflich verschieden*, aber drücken individuell dieselbe physikalische Situation aus.

Das tritt am deutlichsten zutage, wenn man auf die heute[37] auch in der philosophischen Debatte übliche, auf raum-zeitlichen ‚Modellen' beruhende Darstellung, zurückgreift. In pre-relativistischen Theorien, wie der Speziellen Relativitätstheorie, sind nur diejenigen Transformationen h (Lorentz-Transformationen) erlaubt, die die relevante Struktur der Theorie $\langle M, \eta_{\mu\nu}\rangle$ unverändert lassen ($h * \eta_{\mu\nu} = \eta_{\mu\nu}$): Ununterscheidbarkeit entsteht hier weil bestimmte physikalische Unterschiede (Position, Geschwindigkeit, Orientierung) in der mathematischen Struktur nicht vorkommen dürfen, vor und nach der Transformation erscheinen nur die *gleichen* Funktionen $\eta_{\mu\nu}$ (wobei $\eta_{\mu\nu} = \pm\, \delta_{\mu\nu}$). In der Allgemeinen Relativitätstheorie sind dagegen Transformationen h (Diffeomorphismen) erlaubt, die die relevante Struktur der Theorie $\langle M, g_{\mu\nu}\rangle$ nicht erhalten ($h * g_{\mu\nu} \neq g_{\mu\nu}$): Ununterscheidbarkeit entsteht, weil Unterschiede zwi-

schen mathematisch verschiedenen Strukturen, $\langle M, g_{\mu\nu}\rangle$, $\langle M', g'_{\mu\nu}\rangle$, $\langle M'', g''_{\mu\nu}\rangle$, … als physikalisch irrelevant gelten (es geht um *das gleiche* Gravitationsfeld, das durch verschiedene $g_{\mu\nu}$-Systeme ausgedrückt wird).

Weyl hat uns die begrifflichen Mittel zur Verfügung gestellt, um diesen Unterschied einzuordnen: Leibniz' Ununterscheidbarkeitsargumente im Briefwechsel mit Clarke führten den Begriff der ‚Symmetrie' in die Wissenschaftsgeschichte ein, Einsteins Ununterscheidbarkeitsargument, die Lochbetrachtung, dagegen brachte die Idee der heute so genannten ‚Eichfreiheit' ein. Die Logischen Empiristen, so kann man dann modern ausgedrückt argumentieren, betrachteten irrigerweise Diffeomorphismen als die *Symmetriegruppe* der allgemeinen-relativistischen Raumzeit. Einstein, modern gesprochen, betrachtete Diffeomorphismen als die *Eichgruppe* der Theorie.[38]

In der heutigen Debatte ist man sich dieses Unterschieds selbstverständlich bewusst. Indem man aber den Ausdruck ‚Leibniz-Äquivalenz' verwendet, um die Leistung von Einsteins Argument zu benennen, bringt man meines Erachtens irrtümlich zwei ganz verschieden Formen von ‚Ununterscheidbarkeit' unter den gleichen Oberbegriff. Es wäre daher ratsam, eine Unterscheidung zwischen Leibniz-Äquivalenz (als Konsequenz eines mathematischen ‚Mangels') und Einstein-Äquivalenz (als Konsequenz mathematischer ‚Redundanz') einzuführen. Dies könnte ein nützlicher Ansatz sein, um einige der bedeutendsten konzeptuellen Neuigkeiten der Physik im 20. Jahrhundert philosophisch einzuordnen.[39] Die Existenz zweier Formen von Ununterscheidbarkeitsargumenten zeigt, dass Invarianzen nach Symmetrien und Redundanzen unterschieden werden sollten, auch wenn dieser Unterschied manchmal durch den Begriff ‚Eichsymmetrien' verwischt wird (vgl. Giulini und Straumann, 2006).

Anmerkungen

1 Diffeomorphismen sind Abbildungen eines Raumes auf sich selbst, die, grob gesagt, das unmittelbare Benachbartsein von Punkten unangetastet lassen.
2 Verschiebungen sind Abbildungen eines Raumes auf sich selbst, bei denen die Abstände nicht geändert werden.
3 Vgl. insbes. Maudlin (1988), Butterfield (1989), Stachel (1993), Rynasiewicz (1994), Hoefer (1996), Saunders (2002).

4 Für einen Überblick siehe Rickles (2008).
5 Für eine Darstellung des received view siehe Jammer (1993, 231 ff.), Carrier (2009, § 4.3).
6 Vgl. Friedman (1983), Ryckman (1992), Howard (1999), Ryckman (2005).
7 Vgl. Schneider (1988), De Risi (2007).
8 Vgl. GM, V, 179–80; VII, 30; VII, 281–82.
9 Vgl. Leibniz zu Gallois in (GM, I, 180).
10 Vgl. z. B. (GM, V, 155); für eine Liste von Passagen: Schneider (1988).
11 Vgl. Newton (1962), McGuire (1978), DiSalle (2006b).
12 Vgl. etwa Stein (1967/1970), DiSalle (2002; 2006a).
13 Vgl. Lariviere (1987), Arthur (1994), Roberts (2003), Jauernig (2008), Huggett et al. (2009).
14 Vgl. Weyl (1921; 1922a; 1923a; 1927)
15 Vgl. Sklar (1974), Rynasiewicz (1996), Dorato (2000).
16 Vgl. Hawkins (1984), Rowe (1989; 1997).
17 Vgl. Farwell (1990), Farwell und Knee (1990).
18 Vgl. Nerlich (1994), DiSalle (2006b).
19 Vgl. auch Helmholtz (1879, Beilage III, 60); siehe dazu DiSalle (2006a).
20 Hausdorffs Reflexionen über die Geometrie haben erst kürzlich das Interesse der Fachwelt auf sich gezogen. Vgl. hierzu Epple (2006; 2007), Giovanelli (2010).
21 Siehe insbes. Poincaré (1903; 1907; 1912).
22 Vgl. Poincaré (1907, 4): „[…] ces deux univers seront indiscernables l'un et l'autre"; Poincaré (1913a, 62): „Il est amorphe, c'est-à-dire qu'il ne diffère pas de celui qu'on en déduirait par une déformation continue quelconque".
23 Vgl. Poincaré (1895, 10): „Si toutes ces conditions sont remplies, nous dirons que les deux variétés *V* et *V'* sont équivalentes au point de vue de *Analysis Situs*, ou, pour abréger le langage, qu'elles sont homéomorphes". Oder besser: „diffeomorph"; vgl. dazu: Moore (2007); Scholz (1979, 288).
24 Vgl. dazu: Bollinger (1972); Scholz (1979).
25 Vgl. Einstein (1914b, 178); Einstein und Grossmann (1914, 218).
26 Kretschmann bezieht sich hier auf Mach („Wenn wir nun fragen, was denn eigentlich der physiologische Raum mit dem geometrischen Raum gemein hat, so finden wir nur wenige Übereinstimmungen. […] Höchstens könnte man auf Grund desselben eine Topologie aufbauen" (Mach, 1905, 337 ff.; 423 ff.; hier: 337f.)) und die deutschen Übersetzungen der Klassiker von Poincaré (1906a; 1906c; 1913b).
27 Wie aus einem Brief von Lorentz an Ehrenfest, 9. Januar 1916, hervorgeht (vgl. Kox, 1987).
28 CPAE, 8a, Doc. 173, 26. Dezember 1915; Doc. 180, 5. Januar 1915.
29 Vgl. Stachel (1993), Renn und Stachel (2007).
30 Vgl. Friedman (1983, 47), Howard (1999).
31 Erstmals im März 1917 als Aufsatz in *Die Naturwissenschaften* veröffentlicht (Schlick, 1917b), ist es zwei Monate später (ebenfalls bei Julius Springer) als Buch erschienen (Schlick, 1917a) und hat bis 1922 vier Auflagen erfahren (Schlick, 1919; 1920a; 1922).

32 Vgl. Straumann (1987), Vizgin (1994), O'Raifeartaigh und Straumann (2000).
33 Vgl. z. B. Weyl (1927, 74); siehe auch Eddington (1920), 87–88.
34 Vgl. Weyl (1928, 89; 1929b; 1929c; 1929a); Scholz (2004; 2008) und Dirac (1928a; 1928b).
35 Vgl. Norton (2003), (Maudlin (2002), Belot (1998).
36 Siehe Weyl (1934, 56 f.; 1938, 268; 1939).
37 Seit Hawking und Ellis (1973); vgl. Wald (1984).
38 Vgl. Norton (2003), Rovelli (2004); dagegen Weinstein (1999).
39 Vgl. Chen-Ning Yang (1980; 1986).

Siglen

CPAE: Albert Einstein, 1987–: The Collected Papers of Albert Einstein. Hrsg. von Diana Kormos Buchwald. 13 Bde. Princeton University Press.
GM: Gottfried Wilhelm Leibniz, 1850: Leibnizens mathematische Schriften. Hrsg. von Carl Immanuel Gerhardt. 7 Bde. Halle: Schmidt.
GP: Gottfried Wilhelm Leibniz, 1875: Die philosophischen Schriften von Gottfried Wilhelm Leibniz. Hrsg. von Carl Immanuel Gerhardt. 7 Bde. Berlin: Weidmann.
LH: Gottfried Wilhelm Leibniz, 1895: Die Leibniz-Handschriften der Königlichen Öffentlichen Bibliothek zu Hannover. Hannover und Leipzig.
NL FH: Felix Hausdorff, –: ‚Nachlass'. Der Nachlass befindet sich in der Universitätsbliothek Bonn. (Online-Katalog: http.//www.aic.uni-wuppertal.de/fb7/hausdorff/findbuch.asp)

Literatur

Arthur, Richard, 1994: Space and Relativity in Newton and Leibniz. In: The British Journal for the Philosophy of Science 45, S. 219–240.
Bartels, Andreas, 1994: What is Spacetime, if not a Substance? Conclusions from the New Leibnizian Argument. In: Majer, Ulrich; Schmidt, Heinz-Jürgen (Hg.): Semantical Aspects of Spacetime Theories. Mannheim: BI Wissenschaftsverlag, S. 41–51.
Bartels, Andreas, 1996: Modern Essentialism and the Problem of Individuation of Spacetime Points. In: Erkenntnis 45, S. 25–43.

Belot, Gordon, 1998: Understanding Electromagnetism. In: The British Journal for the Philosophy of Science 49, S. 531–555.

Beltrami, Eugenio, 1868: Saggio di interpretazione della geometria non-euclidea. In: Tonelli, Alberto; Cremona, Luigi (Hg.): Opere matematiche. Bd. 1. Milano: Hoepli, S. 374–405.

Beltrami, Eugenio, 1868: Teoria fondamentale degli spazii di curvatura constante. In: Tonelli, Alberto; Cremona, Luigi (Hg.): Opere matematiche. Bd. 1. Milano: Hoepli, S. 406–429.

Beltrami, Eugenio, 1902–1920: Opere matematiche di Eugenio Beltrami. Pubblicate per cura della Facoltà di scienze della R. Università di Roma. Hrsg. von Alberto Tonelli, Guido Castelnuovo und Luigi Cremona. Milano: U. Hoepli.

Bollinger, Maja, 1972: Geschichtliche Entwicklung des Homologiebegriffs. In: Archive for History of Exact Sciences 9, S. 94–170.

Brown, Harvey R., 2005: Physical relativity. Space-Time Structure from a Dynamical Perspective. Oxford: Clarendon.

Butterfield, Jeremy, 1989: The Hole Truth. In: The British Journal for the Philosophy of Science 40, S. 1–28.

Carrier, Martin, 2009: Raum-Zeit. Grundthemen Philosophie. Berlin: de Gruyter.

Christoffel, Elwin Bruno, 1869: Ueber die Transformation der homogenen Differentialausdrücke zweiten Grades. In: Journal für die reine und angewandte Mathematik 70, S. 46–70. [Neudr. in Christoffel, 1910, I, 352–377, 378–382].

Christoffel, Elwin Bruno, 1910: Gesammelte mathematische Abhandlungen. Hrsg. von Ludwig Maurer. Leipzig: Teubner.

Coleman, Robert Alan und Herbert Korté, 1982: The Status and Meaning of the Laws of Inertia. In: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1982. Volume One. Contributed Papers, S. 257–274.

Couturat, Louis, 1902: La logique de Leibniz d'après des documents inédits. Hildesheim: Olms 1961.

De Risi, Vincenzo, 2005: Leibniz on Geometry. Two Unpublished Texts with Translation and Commentary. In: The Leibniz Review 15, S. 127–132.

De Risi, Vincenzo, 2007: Geometry and Monadology. Leibniz's analysis situs and Philosophy of Space. Basel/Boston: Birkhäuser.

Dell'Aglio, Luca, 1996: On the Genesis of the Concept of Covariant

Differentiation. In: Revue d'histoire des Mathematiques 2, S. 215–264.
Dirac, Paul Adrien Maurice, 1928a: The Quantum Theory of the Electron. In: Proceedings of the Royal Society of London. Series A 117.778, S. 610–624.
Dirac, Paul Adrien Maurice, 1928b: The Quantum Theory of the Electron II. In: Proceedings of the Royal Society of London. Series A 118, S. 351–361.
DiSalle, Robert, 2002: Newton's Philosophical Analysis of Space and Time. Cohen, I. Bernard (Hg.): The Cambridge Companion to Newton. Cambridge: Cambridge University Press, S. 33–54.
DiSalle, Robert, 2006a: Kant, Helmholtz, and the Meaning of Empiricism. In: Friedman, Michael; Nordmann, Alfred (Hg.): The Kantian Legacy in Nineteenth-Century Science. Cambridge (Ma.): The MIT Press.
DiSalle, Robert, 2006b: Understanding Space-Time. The Philosophical Development of Physics from Newton to Einstein. Cambridge: Cambridge University Press.
Dorato, Mauro, 2000: Substantivalism, Relationism, and Structural Spacetime Realism. In: Foundations of Physics 30, S. 1605–1628.
Earman, John und John D. Norton, 1987: What Price Substantivalism. The Hole Story. In: British Journal for the Philosophy of Science 38, S. 515–525.
Eddington, Arthur Stanley, 1920: Report on the Relativity Theory of Gravitation. London: Fleetway Press.
Einstein, Albert, 1914a: Die formale Grundlage der allgemeinen Relativitätstheorie. In: Sitzungsberichte der Preussischen Akademie der Wissenschaften 1914, S. 1030–1085. [Neudr. in CPAE 6, Doc. 9].
Einstein, Albert, 1914b: Prinzipielles zur verallgemeinerten Relativitätstheorie. In: Physikali-sche Zeitschrift 15, S. 176–180. [Neudr. in CPAE 4, Doc. 25].
Einstein, Albert, 1916: Die Grundlage der allgemeinen Relativitätstheorie. In: Annalen der Physik 49, S. 769–822. [Neudr. in CPAE 6, Doc 30].
Einstein, Albert, 1917: Über die spezielle und die allgemeine Relativitätstheorie (gemeinverständlich). Braunschweig: Vieweg. [Neudr. in CPAE 6, Doc. 42].
Einstein, Albert, 1918: Prinzipielles zur allgemeinen Relativitätstheorie. In: Annalen der Physik 55, S. 241–244. [Neudr. in CPAE 7, Doc. 4].

Einstein, Albert, 1923: Vorwort des Autors zur Tschechischen Ausgabe. In: Theorie relativity speciálni i obecna. Lehce srozumitelný výklad. Praha: Borový. [geschrieben 1922 für die tschechische Ausgabe (1923) von Einstein, 1917; zitiert in CPAE 6, 53, Note 4].

Einstein, Albert und Marcel Grossmann, 1913: Entwurf einer verallgemeinerten Relativitätstheorie und eine Theorie der Gravitation. I. Physikalischer Teil von A. Einstein II. Mathematischer Teil von M. Grossmann. Leipzig: Teubner. [Neudr. in CPAE 4, Doc. 13].

Einstein, Albert und Marcel Grossmann, 1914: Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie. In: Zeitschrift für Mathematik und Physik 63, S. 215–225. [Neudr. in CPAE 6, Doc. 2].

Engler, Fynn Ole, 2006: Moritz Schlick und Albert Einstein. Url: http://www.mpiwg-berlin.mpg.de/Preprints/P309.PDF.

Epple, Moritz, 2006: Felix Hausdorff's Considered Empiricism. In: Ferreirós, José; Gray, Jeremy (Hg.): The Architecture of Modern Mathematics. Essays in History and Philosophy. Oxford: Oxford University Press.

Epple, Moritz, 2007: An Unusual Career between Cultural and Mathematical Modernism. Felix Hausdorff, 1868–1942. In: Charpa, Ulrich; Deichmann, Ute: Jews and Sciences in German Contexts. Case Studies from the 19th and 20th centuries. Tübingen: Mohr Siebeck, S. 77–100.

Farwell, Ruth, 1990: The Missing Link. Riemann's 'Commentatio', Differential Geometry and Tensor Analysis. In: Historia Mathematica 17, S. 223–255.

Farwell, Ruth und Christopher Knee, 1990: The End of the Absolute. A Nineteenth-Century Contribution to General Relativity. In: Studies in History and Philosophy of Modern Physics 21, S. 91–121.

Farwell, Ruth und Christopher Knee, 1992: The Geometric Challenge of Riemann and Clifford. In: Boi, Luciano; Flament, Dominique; Salanskis, Jean-Michel (Hg.): 1830–1930. A Century of Geometry. Epistemology, History and Mathematics. New York/Berlin: Springer.

Friedman, Michael, 1983: Foundations of Space-Time Theories. Relativistic Physics and Philosophy of Science. Princeton: Princeton University Press.

Galilei, Galileo, 1632: Dialogo di Galileo Galilei Linceo matematico sopraordinario dello studio di Pisa . e filosofo e matematico primario

del serenissimo gr. duca di Toscana . doue ne i congressi di quattro giornate si discorre sopra i due massimi sistemi del mondo tolemaico e copernicano . proponendo indeterminatamente le ragioni filosofiche e naturali tanto per l'una quanto per l'altra parte. Florenz: Per Gio.Batista Landini. [Neudr. in Galilei, 1890–1909, vol. 7].

Galilei, Galileo, 1890–1909: Le Opere di Galileo Galilei, Edizione Nazionale. Hrsg. von Antonio Favaro. 20 Bde. Florenz: Barbera.

Giovanelli, Marco, 2010: Leibniz, Kant und Hausdorff über das Raumproblem. In: Zeitschrift für allgemeine Wissenschaftstheorie 41, S. 283–313.

Giovanelli, Marco, 2013: Erich Kretschmann as a Proto-Logical-Empiricists: Adventures and Disadventures of the Point-Coincidence Argument. In: Studies in the History of Modern Phyiscs 44, S. 115–134.

Giulini, Domenico und Norbert Straumann, 2006: Einstein's Impact on the Physics of the Twentieth Century. In: Studies in History and Philosophy of Science. Part B: Studies in History and Philosophy of Modern Physics 37, S. 115–173.

Grossmann, Marcel, 1913: Mathematische Begriffsbildungen zur Gravitationstheorie. In: Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich 58.

Hawking, Stephen W. und George Francis Rayner Ellis, 1973: The Large Scale Structure of Space-Time. Cambridge: Cambridge Univ. Press.

Hawkins, Thomas, 1980: Non-Euclidean Geometry and Weierstrassian Mathematics. The Background to Killing's Work on Lie Algebras. In: Historia Mathematica 7, S. 289–342.

Hawkins, Thomas, 1984: The Erlanger Programm of Felix Klein. Reflections on its Place in the History of Mathematics. In: Historia Mathematica 11, S. 442–470.

Hawkins, Thomas, 2000: Emergence of the Theory of Lie Groups. An Essay in the History of Mathematics, 1869–1926. Sources and Studies in the History of Mathematics and Physical Sciences. New York: Springer.

Helmholtz, Hermann von, 1870/1876: Über Ursprung und Bedeutung der Geometrischen Axiome. In: Vorträge und Reden. Braunschweig/Wiesbaden: Vieweg, S. 21–51. [mit dem Appendix ‚Mathematische Erläuterungen', S. 51–54; Neudr. in Helmholtz, 2003, vol. I.2.2, S. 640–659].

Helmholtz, Hermann von, 1879: Die Thatsachen in der Wahrnehmung. Rede gehalten zur Stiftungsfeier der Friedrich-Wilhelms-Universität zu Berlin am 3. August 1878. Berlin: A. Hirschwald.

Helmholtz, Hermann von, 1921: Schriften zur Erkenntnistheorie. Hrsg. von Moritz Schlick und Paul Hertz. Berlin: Springer.

Helmholtz, Hermann von, 2003: Gesammelte Schriften. Hrsg. von Fabian Bernhard. Hildesheim: Olms.

Hilbert, David, 1916/1917: Das Kausalitätsprinzip in der Physik. In: Sauer, Tillman und Majer, Ulrich (Hg.): David Hilbert's Lectures on the Foundations of Physics, 1915–1927. Berlin: Springer, S. 335–346.

Hilbert, David, 1917: Grundlagen der Physik. Zweite Mitteilung, vorgelegt in der Sitzung vom 23. Dezember 1916. In: Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Universität zu Göttingen. Math.-physik. Klasse, 1915, S. 53–72. [Neudr. in Hilbert, 2009, Kp. I, S. 47–72].

Hilbert, David, 2009: David Hilbert's Lectures on the Foundations of Physics 1915–1927. Hrsg. von Tilman Sauer und Ulrich Majer. Berlin/Heidelberg: Springer.

Hoefer, Carl, 1996: The Metaphysics of Space-Time Substantivalism. In: The Journal of Philosophy 93, S. 5–27.

Howard, Don, 1999: Point Coincidences and Pointer Coincidences. Einstein on the Invariant Content of Space-Time Theories. In: Goenner, Hubert (Hg.): The Expanding Worlds of General Relativity. Basel: Birkhäuser, S. 463–500.

Howard, Don und John D. Norton, 1993: Out of the Labyrinth. Einstein, Hertz, and the Gottingen Answer to the Hole Argument. In: The Attraction of Gravitation. New Studies in the History of General Relativity. Basel: Birkhäuser, S. 30–62.

Huggett, Nick und Carl Hoefer, 2009: Absolute and Relational Theories of Space and Motion. In: Zalta, Edward N. (Hg.): The Stanford Encyclopedia of Philosophy, Fall 2009.

Jammer, Max, 1993: Concepts of Space. The History of Theories of Space in Physics. New York: Dover Publications.

Janssen, Michel, 2005: Of Pots and Holes. Einstein's Bumpy Road to General Relativity. In: Annalen der Physik 14 (Supplement), S. 58–85.

Jauernig, Anja, 2008: Leibniz on Motion and the Equivalence of Hypotheses. In: The Leibniz Review 18, S. 1–40.

Klein, Felix, 1872: Vergleichende Betrachtungen über neuere geometrische Forschungen. Erlangen: Verlag von Andreas Deichert. [Neudr. in Klein, 1921, I, S. 460–497; erneut veröffentlicht als: Vergleichende Betrachtungen über neuere geometrische Forschungen. In: Mathematische Annalen 43, S. 63–100].

Klein, Felix, 1910: Über die geometrischen Grundlagen der Lorentzgruppe. In: Jahresbericht der Deutschen Mathematiker-Vereinigung 19, S. 281–300. [Neudr. in Klein, 1921, S. 533–552].

Klein, Felix, 1921: Gesammelte mathematische Abhandlungen. Springer, Berlin.

Kox, A. J., 1987: Hendrik Antoon Lorentz, the Ether, and the General Theory of Relativity. In: Archive for History of Exact Sciences 38, S. 67–78.

Kretschmann, Erich, 1915: Über die prinzipielle Bestimmbarkeit der berechtigten Bezugssysteme beliebiger Relativitätstheorien. In: Annalen der Physik 48, S. 907–982.

Kretschmann, Erich, 1918: Über den physikalischen Sinn der Relativitätspostulate. A. Einsteins neue und seine ursprüngliche Relativitätstheorie. In: Annalen der Physik 53, S. 575–614.

Lariviere, Barbara, 1987: Leibnizian Relationalism and the Problem of Inertia. In: Canadian Journal of Philosophy 17, S. 437–448.

Laugwitz, Detlef, 1996: Bernhard Riemann, 1826–1866. Wendepunkte in der Auffassung der Mathematik. Basel/Boston: Birkhäuser.

Leibniz, Gottfried Wilhelm, 1850: Leibnizens mathematische Schriften. Hrsg. von Carl Immanuel Gerhardt. 7 Bde. Halle: Schmidt.

Levi-Civita, Tullio und Gregorio Ricci-Curbastro, 1900: Méthodes de calcul différentiel absolu et leurs applications. In: Mathematische Annalen 54, S. 125–201. [Neudr. in Ricci-Curbastro, 1956/1957, II, S. 185–271].

Lie, Sophus, 1893: Theorie der Transformationsgruppen. Bd. 3. Leipzig: Teubner.

London, Fritz, 1927: Die Theorie von Weyl und die Quantenmechanik. In: Naturwissenschaften 15, S. 187.

Mach, Ernst, 1905: Erkenntnis und Irrtum. Skizzen zur Psychologie der Forschung. Leibzig: Barth.

Maudlin, Tim, 1988: The Essence of Space-Time. In: Proceedings of the Biennial Meeting of the Philosophy of Science Association 2, S. 2–91.

Maudlin, Tim, 2002: Thoroughly Muddled McTaggart. Or, How to Abuse Gauge Freedom to Create Metaphysical Monostrosities. In: Philosophers' Imprint 2, S. 1–19.

McGuire, J. E., 1978: Newton on Place, Time, and God. An Unpublished Source. In: British Journal for the History of Science 11, S. 113–129.

Minkowski, Hermann, 1909: Raum und Zeit. In: Jahresberichte der Deutschen Mathematiker-Vereinigung 1908/1909. Leibzig: Teubner, S. 75–88.

Möbius, Ferdinand August, 1827: Der barycentrische Calcul, ein neues Hülfsmittel zur analytischen Behandlung der Geometrie, dargestellt und insbesondere auf die Bildung neuer Classen von Aufgaben und die Entwickelung mehrerer Eigenschaften der Kegelschnitte angewendet. Leipzig: Barth. [Neudr. in Möbius, 1885, vol. 1].

Möbius, Ferdinand August, 1863: Theorie der elementaren Verwandtschaft. In: Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-physik. Klasse 17, S. 31–68.

Möbius, Ferdinand August, 1885: Gesammelte Werke. Hrsg. von Felix Klein, Richard Baltzer und Wilhelm Scheibner. Leipzig: Hirzel.

Mongré, Paul, 1898: Das Chaos in kosmischer Auslese. Ein erkenntniskritischer Versuch. Leipzig: Naumann.

Moore, Gregory H., 2007: The Evolution of the Concept of Homeomorphism. In: Historia Mathematica 34, S. 333–343.

Nerlich, Graham, 1994: What Spacetime Explains. Metaphysical Essays on Space and Time. Cambridge: Cambridge Univ. Press.

Newton, Isaac, 1687: Philosophiae naturalis principia mathematica. 1. Aufl. London: jussi Societatus Regiae ac typis Josephi Streater; prostat apud plures bibliopolas.

Newton, Isaac, 1962: Unpublished Scientific Papers of Isaac Newton. A Selection from the Portsmouth Collection in the University Library, Cambridge. Hrsg. von Alfred Rupert Hall. Cambridge: Cambridge Univ. Press.

Norton, John D., 1984: How Einstein found his Field Equations. 1912–1915. In: Historical Studies in the Physical Sciences 14, S. 253–316.

Norton, John D., 1995: Did Einstein Stumble? The Debate over General Covariance. In: Erkenntnis 42, S. 223–245.

Norton, John D., 1999: Geometries in Collision. Einstein, Klein and

Riemann. In: Gray, Jeremy (Hg.): The Symbolic Universe. Oxford/New York: Oxford University Press, S. 128–144.
Norton, John D., 2003: General Covariance, Gauge Theories, and the Kretschmann Objection. In: Brading, Katherine; Castellani, Elena (Hg.): Symmetries in Physics. Philosophical Reflections. Cambridge: Cambridge University Press, S. 110–123.
O'Raifeartaigh, Lochlainn und Norbert Straumann, 2000: Gauge theory. Historical Origins and Some Modern Developments. In: Reviews of Modern Physics 72, S. 1–23.
Pais, Abraham, 1982: Subtle is the Lord. The Science and the Life of Albert Einstein. New York: Oxford University Press.
Poincaré, Henri, 1891: Les géométries non euclidiennes. In: Revue générale des sciences pures et appliquées 2, S. 769–774. [Neudr. in Poincaré, 1902, ch. 3].
Poincaré, Henri, 1895: Analysis Situs. In: Journal de l'École Polytechnique 2, S. 1–123.
Poincaré, Henri, 1902: La science et l'hypothèse. Paris: Flammarion.
Poincaré, Henri, 1903: L'espace et ses trois dimensions. In: Revue de métaphysique et de morale 11, S. 281–301.
Poincaré, Henri, 1905: La valeur de la science. Paris: Flammarion.
Poincaré, Henri, 1906a: Der Wert der Wissenschaft. Hrsg. von Emilie Weber und Heinrich Weber. Leipzig: B.G. Teubner.
Poincaré, Henri, 1906b: Sur la dynamique de l'électron. In: Rendiconti del circolo matematico di Palermo 21, S. 129–176.
Poincaré, Henri, 1906c: Wissenschaft und Hypothese. Leipzig: Teubner.
Poincaré, Henri, 1907: La relativité de l'espace. In: Année psychologique 13, S. 1–17.
Poincaré, Henri, 1908: Science et méthode. Paris: Flammarion.
Poincaré, Henri, 1912: L'espace et le temps. In: Scientia (Rivista di Scienza) 12, S. 159–170.
Poincaré, Henri, 1913a: Dernières pensées. Paris: E. Flammarion.
Poincaré, Henri, 1913b: Letzte Gedanken. Leipzig: Akademische Verlagsgesellschaft.
Poncelet, Jean-Victor, 1822: Traité des propriétés projectives des figures. Paris: Gauthier-Villars.
Reich, Karin, 1994: Die Entwicklung des Tensorkalküls. Vom absoluten Differentialkalkül zur Relativitätstheorie. Berlin: Birkhäuser.

Reichenbach, Hans, 1922: Der gegenwärtige Stand der Relativitätsdiskussion. Eine kritische Untersuchung. In: Logos 22, S. 316–378. [Neudr. in Reichenbach, 1977, III].

Reichenbach, Hans, 1977: Gesammelte Werke in 9 Bänden. Hrsg. von Andreas Kamlah und Maria Reichenbach. Braunschweig: Vieweg.

Renn, Jürgen und John Stachel, 2007: Hilbert's Foundation of Physics. From a Theory of Everything to a Constituent of General Relativity. In: Renn, Jürgen; Janssen, Michel (Hg.): The Genesis of General Relativity. Dordrecht: Springer, S. 857–972.

Ricci-Curbastro, Gregorio, 1884: Principii di una teoria delle forme differenziali quadratiche. In: Annali di Matematica Pura ed Applicata 12, S. 135–167. [Neudr. in Ricci-Curbastro, 1956/1957, I, 138–171].

Ricci-Curbastro, Gregorio, 1888: Delle derivazioni covarianti e controvarianti e del loro uso nella analisi applicata. In: Studi editi dalla Università di Padova a commemorare l'ottavo centenario della Università di Bologna. Bd. 3. Padova: Tip. del Seminario, S. 3–23. [Neudr. in Ricci-Curbastro, 1956/1957, I, S. 245–267].

Ricci-Curbastro, Gregorio, 1892: Résumé de quelques travaux sur les systèmes variables de fonctions. In: Bulletin des sciences mathématiques 16, S. 167–189. [Neudr. in Ricci-Curbastro, 1956/1957, I, S. 288–310].

Ricci-Curbastro, Gregorio, 1893: Di alcune applicazioni del Calcolo differenziale assoluto alla teoria delle forme differenziali quadratiche binarie e dei sistemi a due variabili. In: Atti dell'Istituto Veneto di scienze, lettere ed arti 7.4, S. 1336–1364. [Neudr. in Ricci-Curbastro, 1956/1957 I, S. 311–335].

Ricci-Curbastro, Gregorio, 1956/1957: Opere. Hrsg. von Unione matematica italiana e col contributo del Consiglio nazionale delle ricerche. Roma: Cremonese.

Rickles, Dean, 2008: Symmetry, Structure and Spacetime. Amsterdam: Elsevier.

Riemann, Bernhard, 1854/1868: Ueber die Hypothesen, welche der Geometrie zu Grunde liegen (aus dem Nachlass des Verfassers mitgetheilt durch R. Dedekind). In: Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen 13, S. 132–152. [Neudr. in Riemann, 1876].

Riemann, Bernhard, 1861/1876: Commentatio mathematica, qua respondere tentatur quaestioni ab Illma Academia Parisiensi propositae.

In: Weber, Heinrich; Dedekind Richard (Hg.): Bernhard Riemann's Gesammelte Mathematische Werke. Leipzig: Teubner.

Riemann, Bernhard, 1876: Bernhard Riemann's gesammelte mathematische Werke und wissenschaftlicher Nachlass. Hrsg. von Heinrich Weber und Richard Dedekind. Leipzig: Teubner.

Riemann, Bernhard, 1892: Gesammelte mathematische Werke und wissenschaftlicher Nachlass. Hrsg. von Richard Dedekind. 2. Aufl. Leipzig: Teubner.

Roberts, John T., 2003: Leibniz on Force and Absolute Motion. In: Philosophy of Science 70, S. 553–573.

Rovelli, Carlo, 2004: Quantum gravity. Cambridge: Cambridge University Press.

Rowe, David E., 1989: Klein, Lie, and the Geometric Background of the Erlangen Program. In: Rowe, David (Hg.): The History of Modern Mathematics. Proceedings of the Symposium on the History of Modern Mathematics. Volume 1. Boston: Academic Press, S. 209–273.

Rowe, David E., 1997: German Mathematics and the Early Mathematical Career of Felix Klein. In: Hunger Parshall, Karen; Rowe, David E. (Hg.): The Emergence of the American Mathematical Research Community, 1876–1900. J. J. Sylvester, Felix Klein, and E. H. Moore. Providence, R. I.: American Mathematical Society.

Ryckman, Thomas, 1992: (P)oint-(C)oincidence Thinking. The Ironical Attachment of Logical Empiricism to General Relativity. In: Studies in History and Philosophy of Modern Physics 23, S. 471–497.

Ryckman, Thomas, 2005: The Reign of Relativity. Philosophy in Physics 1915–1925. Oxford/New York: Oxford University Press.

Rynasiewicz, Robert, 1994: The Lessons of the Hole Argument. In: The British Journal for the Philosophy of Science 45, S. 407–436.

Rynasiewicz, Robert, 1996: Absolute Versus Relational Space-Time. An Outmoded Debate? In: The Journal of Philosophy 93, S. 279–306.

Rynasiewicz, Robert, 1999: Kretschmann's Analysis of Covariance and Relativity Principles. In: Goenner, Hubert (Hg.): The Expanding Worlds of General Relativity. Basel: Birkhäuser, S. 431–462.

Saunders, Simon, 2002: Indiscernibles, General Covariance, and other Symmetrie. In: Renn, Jürgen; Howard, Don (Hg.): Revisiting the Foundations of Relativistic Physics. Festschrift in Honour of John Stachel. Boston/London: Kluwer, S. 151–173.

Schlick, Moritz, 1917a: Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der allgemeinen Relativitätstheorie. Berlin: Springer. [Neudr. in Schlick, 2006, vol. II].

Schlick, Moritz, 1917b: Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der allgemeinen Relativitätstheorie. In: Die Naturwissenschaften 5, S. 162–186. [Neudr. in Schlick, 2006, vol. II].

Schlick, Moritz, 1918: Allgemeine Erkenntnislehre. Naturwissenschaftliche Monographien und Lehrbücher. Berlin: J. Springer. [Neudr. in Schlick, 2006, vol. I].

Schlick, Moritz, 1919: Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der Relativitäts- und Gravitationstheorie. 2. Aufl. Berlin: Springer. [Neudr. in Schlick, 2006, vol. II].

Schlick, Moritz, 1920a: Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der allgemeinen Relativitätstheorie. 3. Aufl. Berlin: Springer. [Neudr. in Schlick, 2006, vol. II].

Schlick, Moritz, 1920b: Space and Time in Contemporary Physics, an Introduction to the Theory of Relativity and Gravitation. Hrsg. von Henry L. Brose. Oxford/New York: Oxford University Press.

Schlick, Moritz, 1922: Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der allgemeinen Relativitätstheorie. 3. Aufl. Berlin: Springer. [Neudr. in Schlick, 2006, vol. II].

Schlick, Moritz, 2006: Gesamtausgabe. Hrsg. von Friedrich Stadler und Hans Jürgen Wendel. Berlin: Springer.

Schneider, Martin, 1988: Funktion und Grundlegung der Mathesis Universalis. In: Studia Leibnitiana (Sonderheft) 15, S. 162–182.

Scholz, Erhard, 1979: Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré. Boston, Basel/Stuttgart: Birkhäuser.

Scholz, Erhard, 1982: Riemanns frühe Notizen zum Mannigfaltigkeitsbegriff und zu den Grundlagen der Geometrie. In: Archive for History of Exact Sciences 27, S. 213–232.

Scholz, Erhard, 1992: Riemann's Vision of a New Approach to Geometry. In: Flament, Dominique; Salanskis, Jean-Michel (Hg.): 1830–1930. A Century of Geometry. Epistemology, History and Mathematics. Berlin/New Berlin/New York: Springer, S. 22–34.

Scholz, Erhard, 2004: Hermann Weyl's Analysis of the ‚Problem of Space' and the Origin of Gauge Structures. In: Science in Context 17, S. 165–197.

Scholz, Erhard, 2008: Weyl Geometry in Late 20th Century Physics. In: Rowe, David E. (Hg.): Beyond Einstein. Proceedings Mainz Conference September 2008. Birkhäuser: Basel. [im Erscheinen].

Sklar, Lawrence, 1974: Space, Time, and Spacetime. Berkeley: University of California.

Stachel, John, 1980: Einstein's Search for General Covariance, 1912–1915. Gelesen auf der neunten internationalen Konferenz zu ‚General Relativity and Gravitation', Jena 1980. [Neudr. in Stachel, 2002, 301–337].

Stachel, John, 1993: The Meaning of General Covariance. The Hole Story. In: Erman, John (Hg.): Philosophical Problems of the Internal and External Worlds. Essays on the Philosophy of Adolf Grünbaum. Pittsburgh-Konstanz series in the philosophy and history of science. Pittsburgh: Univ. of Pittsburgh Pr., S. 129–160.

Stachel, John, 2002: Einstein from ‚B' to ‚Z'. Einstein Studies 9. Boston: Birkhäuser.

Stein, Howard, 1967/1970: Newtonian Space-Time. In: Palter, Robert (Hg.): The Annus mirabilis of Sir Isaac Newton, 1666–1966. Cambridge (Ma.)/London: The MIT Press, S. 254–284.

Steiner, Jakob, 1832: Systematische Entwicklung der Abhängigkeit geometrischer Gestalten von einander. Mit Berücksichtigung der Arbeiten alter und neuer Geometer über Porismen, Projections-Methoden, Geometrie der Lage, Transversalen, Dualität und Reciprocität, etc. Berlin: Fincke.

Straumann, Norbert, 1987: Zum Ursprung der Eichtheorien bei Hermann Weyl. In: Physikalische Blätter 43, S. 414–421.

Vizgin, Vladimir Pavlovich, 1994: Unified Field Theories in the First Third of the 20th Century. Boston, Basel/Stuttgart: Birkhäuser.

Wald, Robert M., 1984: General relativity. Chicago: Univ. of Chicago Press.

Weinstein, Steven, 1999: Gravity and Gauge Theory. In: Philosophy of Science 66, S. 146–155.

Weyl, Hermann, 1918a: Gravitation und Elektrizität. In: Sitzungsberichte der Preussischen Akademie der Wissenschaften, S. 465–480. [Neudr. in Weyl, 1968, II, Doc. 31].

Weyl, Hermann, 1918b: Raum, Zeit, Materie. Vorlesungen über allgemeine Relativitätstheorie. Berlin: Springer.

Weyl, Hermann, 1919: Eine neue Erweiterung der Relativitätstheorie.

In: Annalen der Physik 59, S. 101–133. [Neudr. in Weyl, 1968, II, Doc. 34].
Weyl, Hermann, 1920/21: Die Einsteinsche Relativitätstheorie. In: Schweizerland/Schweizerische Bauzeitung. [Neudr. in Weyl, 1968, II, Doc. 39].
Weyl, Hermann, 1921: Das Raumproblem. In: Jahresbericht der Deutschen Mathematikervereinigung 30, S. 92–93.
Weyl, Hermann, 1922a: Das Raumproblem. In: Jahresbericht der Deutschen Mathematikervereinigung 31, S. 205–221.
Weyl, Hermann, 1922b: Die Relativitätstheorie auf der Naturforscherversammlung. In: Jahresbericht der Deutschen Mathematikervereinigung 31, S. 51–63. [Neudr. in Weyl, 1968, II, Doc. 52].
Weyl, Hermann, 1923: Mathematische Analyse des Raumproblems. Vorlesungen gehalten in Barcelona und Madrid. Berlin: Springer.
Weyl, Hermann, 1924: Massenträgheit und Kosmos. Ein Dialog. In: Naturwissenschaften 12, S. 197–204. [Neudr. in Weyl, 1968, II, Doc. 65].
Weyl, Hermann, 1925/1988: Riemanns geometrische Ideen, ihre Auswirkung und ihre Verknüpfung mit der Gruppentheorie. New York, Berlin/Heidelberg: Springer.
Weyl, Hermann, 1927: Philosophie der Mathematik und Naturwissenschaft. München/Berlin: Oldenbourg.
Weyl, Hermann, 1928: Gruppentheorie und Quantenmechanik. Leipzig: Hirzel.
Weyl, Hermann, 1929a: Elektron und Gravitation. In: Zeitschrift für Physik 56, S. 330–352. [Neudr. in Weyl, 1968, III, Doc. 85].
Weyl, Hermann, 1929b: Gravitation and the Electron. In: Proceedings of the National Academy of Sciences of the United States of America 15, S. 323–334. [Neudr. in Weyl, 1968, III, Doc. 84].
Weyl, Hermann, 1929c: Gravitation and the Electron. In: The Rice Institute Pamphlet 16, S. 280–295.
Weyl, Hermann, 1930: Felix Kleins Stellung in der mathematischen Gegenwart. In: Die Naturwissenschaften 18, S. 4–20.
Weyl, Hermann, 1931: Geometrie und Physik. In: Die Naturwissenschaften 19, S. 49–58. [Neudr. in Weyl, 1968, III, Doc. 93].
Weyl, Hermann, 1934: Mind and Nature. Philadelphia: University of Pennsylvania Press. [Neudr. in Weyl, 2009, Ch. 5].
Weyl, Hermann, 1938: Symmetry. In: Journal of the Washington Academy of Sciences 28, S. 253–271. [Neudr. in Weyl, 1968].

Weyl, Hermann, 1939: The Classical Groups. Their Invariants and Representations. Princeton: Princeton University Press.
Weyl, Hermann, 1951: 50 Jahre Relativitätstheorie. In: Die Naturwissenschaften 38, S. 73–83. [Neudr. in Weyl, 1968, III, Doc. 149].
Weyl, Hermann, 1952: Symmetry. Princeton: Princeton Univ. Press.
Weyl, Hermann, 1968: Gesammelte Abhandlungen. Hrsg. von Komaravolu Chandrasekharan. 4 Bde. Berlin: Springer.
Weyl, Hermann, 1990: Philosophie der Mathematik und Naturwissenschaft. München: Oldenbourg.
Weyl, Hermann, 2009: Mind and Nature. Selected Writings on Philosophy, Mathematics, and Physics. Hrsg. von Peter Pesic. Princeton: Princeton Univ. Press.
Yang, Chen-Ning, 1980: Einstein's Impact on Theoretical Physics. In: Physics Today 33, S. 42.
Yang, Chen-Ning, 1986: Hermann Weyl's Contribution to Physics. In: Yang, Chen-Ning u. a. (Hg.): Hermann Weyl. 1885–1985 Centenary Lectures. Berlin: Springer.
Yang, Chen-Ning und R. L. Mills, Okt. 1954: Conservation of Isotopic Spin and Isotopic Gauge Invariance. In: Physical Review Letters 96, S. 191–195.

Verzeichnis der Autoren

Dr. Simon Friederich
Universität Göttingen
Philosophisches Seminar
Humboldtallee 19
37073 Göttingen
email@simonfriederich.eu

Marco Giovanelli
Forum Scientiarum
Doblerstraße 33
72074 Tübingen
marco.giovanelli@uni-tuebingen.de

Prof. Dr. Frank Hofmann
Universität Luxemburg
Department of philosophy
FSLHASE, IPSE
Campus Walferdange
Route de Diekirch, B.P. 2
7220 Walferdange
Luxemburg
frank.hofmann@uni.lu

Prof. Dr. Olaf L. Müller
Humboldt-Universität zu Berlin
Institut für Philosophie
Unter den Linden 6
10099 Berlin
muelleol@philosophie.hu-berlin.de

Ferdinand Pöhlmann, M.A.
Philosophy of Neuroscience
Werner Reichardt Centre for Integrative Neuroscience
Ottfried-Müller-Str. 25
72076 Tübingen
ferdinand.poehlmann@uni-tuebingen.de

Prof. Dr. Oliver R. Scholz
Westfälische Wilhelms-Universität Münster
Philosophisches Seminar
Domplatz 23
48143 Münster
oscholz@uni-muenster.de

PHILOSOPHIA NATURALIS

Eingereichte Beiträge dürfen weder schon veröffentlicht worden sein noch gleichzeitig einem anderen Organ angeboten werden. Mit der Annahme des Manuskriptes zur Veröffentlichung in der *Philosophia naturalis* räumt der Autor dem Verlag Vittorio Klostermann das zeitlich und inhaltlich unbeschränkte Nutzungsrecht im Rahmen der Print- und Online-Ausgabe der Zeitschrift ein. Dieses beinhaltet das Recht der Nutzung und Wiedergabe im In- und Ausland in körperlicher und unkörperlicher Form sowie die Befugnis, Dritten die Wiedergabe und Speicherung des Werkes zu gestatten. Im Übrigen räumt der Autor dem Verlag alle sonstigen durch Verwertungsgesellschaften (z.B. VG Wort) wahrgenommenen Rechte nach deren Satzung, Wahrnehmung und Verteilungsplan zur gemeinsamen Einbringung ein. Der Autor behält jedoch das Recht, nach Ablauf eines Jahres anderen Verlagen eine einfache Abdruckgenehmigung zu erteilen.

Richtlinien zur Manuskriptgestaltung

Bitte jeden Beitrag mit *Titelblatt* abgeben, das folgende Angaben enthält: Name und Vorname des Autors / der Autorin (mit akad. Titel), Titel des Beitrags, vollständige Adresse (inkl. Telefon-Nummer), nähere Bezeichnung der Arbeitsstätte.

Die *Manuskripte* sollten 3-fach ausgedruckt und als Word- oder rtf-File eingereicht werden und ein deutsch- und englischsprachiges Abstract enthalten. Das Manuskript sollte einen breiten Rand haben.

Der *Umfang* (einschließlich Anmerkungen und Bibliografie) soll bei den Aufsätzen nicht mehr als 30 maschinengeschriebene Seiten (ca. 2.000 Anschläge, 2-zeilig) betragen.

Für *Abbildungen* im Text bitte die Originalvorlage einreichen. Abbildungen müssen numeriert und mit Autorennamen versehen sein.

Zitate im Text sollten vom Haupttext durch eine Leerzeile abgehoben werden. Nach dem zitierten Text stehen Name des zitierten Verfassers, Erscheinungsjahr und Seitenangaben in Klammern, z.B.: (Elkana 1974, S. 34). Bei mehreren Autoren werden die jeweiligen Namen durch Schrägstriche getrennt, z.B.: Krantz/Luce/Suppes/Tversky 1971, S. 8). Wird auf mehrere Publikationen desselben Autors im selben Erscheinungsjahr verwiesen, so sollen sie numeriert werden: (Ludwig 1970 a) bzw. (Ludwig 1970 b).

Die *Anmerkungen* sind im Manuskript fortlaufend zu numerieren; sie stehen am Schluss des Beitrags in numerischer Reihenfolge.

Für das anschließende *Literaturverzeichnis* in alphabetischer und chronologischer Reihenfolge gilt folgendes Muster:

Elkana, Y., 1974: *The Discovery of the Conservation of Energy*. London: Huchinson.

Clausius, R., 1850: Über die bewegende Kraft der Wärme. In: *Annalen der Physik und Chemie*, 79, S. 500–524.

Klein, M.J., 1978: The Early Papers of J. Willard Gibbs: A Transformation of Thermodynamics. In: E.G. Forbes (Hg.): *Human Implications of Scientific Advance*. Edinburgh: University Press, S. 330–341.

Korrekturen: Die Autoren erhalten vom Verlag die Fahnen ihres Beitrags mit der Bitte, die korrigierten Fahnen *innerhalb von zwei Wochen* an den Herausgeber zu schicken. In den Fahnen sollen nur noch Satzfehler berichtigt werden.

Nach Erscheinen des Heftes erhalten die Autoren elektronische Belege.

philosophia naturalis

Located at the crossroads between natural philosophy, the theory and history of science, and the philosophy of technology, Journal for the Philosophy of Nature has represented for many decades – not only in the German speaking countries but internationally – a broad range of topics not addressed by any other periodical.
The journal has a highly interdisciplinary focus. Articles with systematic as well as historical approaches are published in German and English. Their quality is assured by a strict peer review policy.

philosophia naturalis

Inhaltlich an der Schnittstelle zwischen Naturphilosophie, Wissenschaftstheorie, Wissenschaftsgeschichte und Technik-Philosophie angesiedelt, vertritt die Zeitschrift Journal for the Philosophy of Nature seit mehreren Jahrzehnten nicht nur im deutschen Sprachraum, sondern auch im internationalen Vergleich, einen weiten Themenbereich, der von keinem anderen Publikationsorgan vertreten wird.
Die Zeitschrift ist ausgesprochen interdisziplinär ausgerichtet. Sie veröffentlicht Aufsätze in deutscher und englischer Sprache, die sowohl systematisch als auch historisch orientiert sind. Deren Qualität wird durch ein besonders strenges Begutachtungsverfahren gesichert.

ISSN 0031-8027 www.klostermann.de